Notre Dame Journal of Formal Logic
Volume 56, Number 1, 2015

The Arithmetics of a Theory

Albert Visser

Abstract In this paper we study the interpretations of a weak arithmetic, like Buss's theory S^1_2, in a given theory U. We call these interpretations the *arithmetics of* U. We develop the basics of the structure of the arithmetics of U. We study the provability logic(s) of U from the standpoint of the framework of the arithmetics of U. Finally, we provide a deeper study of the arithmetics of a finitely axiomatized sequential theory.

1 Introduction

In this paper, we propose and expose a particular way of viewing theories. We look at theories as a class of interpretations of a given weak arithmetical theory. Consider a theory U. We view the interpretations of the given weak arithmetical theory in U as "occurrences" of that given theory in U. Thus, U appears as a class of copies of the given weak theory. If we consider a model $\mathcal{M}$ of U, the versions of the weak theory sitting inside U take the form of the set of internal models of the weak theory in $\mathcal{M}$.

We will call an interpretation N of the given weak theory in U an *arithmetic of* U. The arithmetics of U have a natural ordering, the (definable) initial embedding ordering $\preceq$. We study basic facts concerning the arithmetics of U and the ordering $\preceq$ in Section 3.

From the perspective of theories as containers of (possibly) lots of arithmetics, we study the provability logics of theories. We fully characterize the propositional modal principles for provability that hold in all arithmetics in any theory U, the only assumption being a constraint on the complexity of the set of axioms of U. The comparatively easy success of this characterization contrasts with the remaining great open questions of provability logic concerning the provability logics of theories like S^1_2 or $\mathrm{I}\Delta_0 + \Omega_1$.

Received April 26, 2012; accepted December 9, 2013
2010 Mathematics Subject Classification: Primary 03F45; Secondary 03F30, 03F40, 03F25
Keywords: interpretation, weak arithmetic, provability logic, Σ_1-sentence
 10.1215/00294527-2835029

Section 4 briefly reviews some basic ideas concerning provability logic.

In Section 5, we study Solovay's theorem in various settings. In Sections 5.1–5.4, we present a proof of Solovay's completeness theorem for Löb's logic via a wonderful version of the proof given by Dick de Jongh, Marc Jumelet, and Franco Montagna. The main part of the proof is itself formulated in a richer modal logic which was formulated and studied by David Guaspari and Robert Solovay. The advantage of the de Jongh–Jumelet–Montagna proof is that it allows us to see clearly what arithmetical principles are involved in Solovay's proof. In Section 5.5, we prove our characterization of the provability logic of all arithmetics of a given theory. In Section 5.6, we give a sufficient condition for when the provability logic of all theories is assumed at a single arithmetic N in U.

In Section 7, we provide an example of a theory U where the provability logic of U is not assumed at any arithmetic N in U.

In Section 6, we study the wondrous world of the arithmetics of a finitely axiomatized sequential theory U. In the sequential case we have many extra properties of our structure of arithmetics to work with. In this section we strengthen certain results due to Harvey Friedman and, independently, to Jan Krajíček. We use the methods of Section 6 to construct the example of Section 7.

The reader interested in provability logic could very well choose to read Sections 2–5. The reader who is interested in the fine structure of the arithmetics of a theory could study Sections 2, 3, 6, and 7. More details on the basics are provided in Appendix A.

About this paper The present paper is, in a sense, a remake of my paper [33]. It is the result of my reflection on what the earlier paper is saying. We strengthen the results of that paper by presenting them in a better framework, and we add new results relevant to the framework.

Prerequisites We will presuppose some knowledge of weak arithmetics. See, for example, Buss [6] or Hájek and Pudlák [15, Chapter V]. Some basic knowledge of provability logic will help to understand the paper. At present there are many expositions: Smoryński [28], Boolos and Sambin [5], Boolos [4], Lindström [21], Japaridze and de Jongh [17], Švejdar [29], and Artemov and Beklemishev [1]. The most comprehensive source concerning the provability logic of weak theories is Verbrugge [30].

2 Basics

In this section we sketch the framework in which our discussion will take place. One problem of sections with basic notions and facts is that they are so long and so boring that the reader gets stuck in them and never arrives at the real stuff. So what I did is to make the present section rather sketchy. At the end of the paper, in Appendix A more details are provided. Regrettably, even without the details this section is rather long, so the reader is advised to go over it lightly and come back to it or Appendix A when needed.

2.1 Theories Theories are, in this paper, one-sorted theories of first-order predicate logic that have a finite signature and that are axiomatized by an axiom set that is represented by a Δ^{b}_1-formula.[1]

Remark 2.1 The demand for Δ^{b}_1-axiomatization seems to be rather restrictive. However, it seems to me that every real-life theory is given by an axiomatization that is p-time decidable.

Because in S^1_2 we have the Σ^{b}_1-replacement axiom, we can relax our demand to the consideration of theories which are Σ^{b}_1-axiomatized. In this case, the witnesses of $\Box_U A$ would not be really a code of a proof but a somewhat modified object.

Note that, by a version of Craig's trick, every recursively enumerable theory in extension can be given a Δ^{b}_1-axiomatization. Of course, a weak theory will not be able to see that both axiomatizations prove the same theorems, so for the eyes of the weak theory the *Craigified* theory will be a different theory. We need a theory like EA, also known as $\mathrm{I}\Delta_0 + \mathsf{exp}$, plus Σ_1-replacement to make this construction work in a verifiable way.

The formula specifying the axiom set is part of the data for the theory. Thus, we treat theories *intensionally* and not as mere sets of theorems. We will explain why this is important for our purposes in Section 4.

We say that a theory is *finitely axiomatized* if its axiomatization has the form $\bigvee_{i<n} x = \underline{\ulcorner A_i \urcorner}$.[2] Note that S^1_2 may prove that a theory has an axiom set of, say, less than two axioms, without being able to prove the equivalence of the formula defining the axiom set with any formula of the prescribed form.

We take identity to be a logical constant. Our official signatures are relational; however, via the term-unwinding algorithm, we can also accommodate signatures with functions.

2.2 Translations and interpretations The notion of interpretation that we will employ in this paper will be *m-dimensional interpretation without parameters*. There are two extensions of this notion: we can consider piecewise interpretations, and we can add parameters. We refrain from considering piecewise interpretations. We explain why in Section A.3 of Appendix A. We sketch a few basic ingredients of adding parameters in Section A.4 of Appendix A. We explain why, in the sequential case, addition of parameters makes no difference for the provability logic of all arithmetics of a given theory in Remark 3.10.

Consider two signatures Σ and Θ. An m-dimensional translation $\tau : \Sigma \to \Theta$ is a quadruple $\langle \Sigma, \delta, \mathcal{F}, \Theta \rangle$, where $\delta(v_0, \ldots, v_{m-1})$ is a Θ-formula and where for any n-ary predicate P of Σ, $\mathcal{F}(P)$ is a formula $A(\vec{v}_0, \ldots, \vec{v}_{n-1})$ in the language of signature Θ, where $\vec{v}_i = v_{i0}, \ldots, v_{i(m-1)}$. In the case of both δ and A all free variables are among the variables shown. Moreover, if $i \neq j$ or $k \neq \ell$, then v_{ik} is syntactically different from $v_{j\ell}$.

We demand that we have $\vdash \mathcal{F}(P)(\vec{v}_0, \ldots, \vec{v}_{n-1}) \to \bigwedge_{i<n} \delta(\vec{v}_i)$. Here $\vdash$ is provability in predicate logic. This demand is inessential, but it is convenient to have.

We define B^τ as follows:

- $(P(x_0, \ldots, x_{n-1}))^\tau := \mathcal{F}(P)(\vec{x}_0, \ldots, \vec{x}_{n-1})$;
- $(\cdot)^\tau$ commutes with the propositional connectives;
- $(\forall x\, A)^\tau := \forall \vec{x}\, (\delta(\vec{x}) \to A^\tau)$;
- $(\exists x\, A)^\tau := \exists \vec{x}\, (\delta(\vec{x}) \wedge A^\tau)$.

We allow identity to be translated to a formula that is not identity. We can define the identity translation id_Σ on Σ, the composition $\rho \circ \tau$ of translations τ and ρ, and the disjunctive translation $\tau\langle A\rangle\rho$, which is τ if A and ρ if $\neg A$. We refer the reader to Appendix A for details.

A translation relates signatures; an interpretation relates theories. An interpretation $K : U \to V$ is a triple $\langle U, \tau, V\rangle$, where U and V are theories and $\tau : \Sigma_U \to \Sigma_V$. We demand that for all axioms A of U, we have $V \vdash A^\tau$.

In the context of the formalization of interpretability, we have to distinguish between *axioms-interpretability*, which is the notion we just introduced, and *theorems-interpretability*, where we demand that: for all theorems A of U, we have $V \vdash A^\tau$. In the real world these notions are equivalent, but we need a principle like Σ_1-collection to prove that, so, for example, Buss's theory S^1_2 does not "know" this equivalence. See Visser [32] for more information about this matter.

Here are some further definitions and conventions.

- Suppose $K : U \to V$. We often write A^K for A^{τ_K}, in the context of a theory W that extends V.
- We write $\overline{U}$ for the set of theorems of U. Suppose $K : U \to V$. We write $\overline{K} := \{A \mid V \vdash A^K\}$. We note that $\overline{U} \subseteq \overline{K}$. If $\overline{K} = \overline{U}$, we will say that K is *faithful*.
- $\mathsf{ID}_U : U \to U$ is the interpretation $\langle U, \mathsf{id}_{\Sigma_U}, U\rangle$.
- Suppose $\overline{U} \subseteq \overline{V}$. Then, $\mathcal{E}_{UV} : U \to V$ is $\langle U, \mathsf{id}_{\Sigma_U}, V\rangle$.
- Suppose $K : U \to V$ and $M : V \to W$. Then, $KM := M \circ K : U \to W$ is $\langle U, \tau_M \circ \tau_K, W\rangle$.
- Suppose $K : W \to U$ and $U \subseteq V$. We write $K \uparrow V$ for $\mathcal{E}_{UV} \circ K$.
- Suppose $M : V \to Z$ and $U \subseteq V$. We write $U \downarrow M$ for $M \circ \mathcal{E}_{UV}$.
- Suppose $K : U \to (V + A)$ and $M : U \to (V + \neg A)$. Then $K\langle A\rangle M : U \to V$ is the interpretation $\langle U, \tau_K\langle A\rangle\tau_M, V\rangle$. In an appropriate category $K\langle A\rangle M$ is a special case of a product.

The notation $K : U \to V$ is inspired by the idea of interpretations as arrows in a category. There is also an intuition of interpretability as a generalization of provability. The traditional notation and notions associated to this intuition are the following:

- $K : U \lhd V$ stands for $K : U \to V$.
- $K : V \rhd U$ stands for $K : U \to V$.
- $U \lhd V$ stands for $\exists K\ K : U \lhd V$. We say that U is *interpretable* in V.
- $V \rhd U$ stands for $\exists K\ K : V \rhd U$. We say that V *interprets* U.
- $U \lhd_{\mathsf{loc}} V$ means that all finitely axiomatized subtheories U_0 of U are interpretable in V. We say that U is *locally interpretable* in V.
- $U \lhd_{\mathsf{mod}} V$ means that, for every $\mathcal{M} \models V$, there is a translation τ such that $\tau(\mathcal{M}) \models U$. We say that U is *model-interpretable* in V.

2.3 i-morphisms Consider an interpretation $K : U \to V$. We can view this interpretation as a uniform way of constructing internal models $\tau_K(\mathcal{M})$ of U from models $\mathcal{M}$ of V. This construction gives us the contravariant model functor as soon as we have defined an appropriate category of interpretations.

Now consider two interpretations $K, M : U \to V$. Between the inner models $\tau_K(\mathcal{M})$ and $\tau_M(\mathcal{M})$ we have the usual structural morphisms of models. We are interested in the case where these morphisms are V-definable and uniform over models. This idea leads to the notion of i-morphism. An i-morphism $F : K \to M$ is a triple $\langle K, F(\vec{u}, \vec{v}), M\rangle$, where $F(\vec{u}, \vec{v})$ is a V-formula and where in all models of V, F represents a morphism of models from $\tau_K(\mathcal{M})$ to $\tau_M(\mathcal{M})$.

Two i-morphisms $F, G : K \to M$ are *i-equal*, when

$$V \vdash \forall \vec{u}, \vec{v} \left(F(\vec{u}, \vec{v}) \leftrightarrow G(\vec{u}, \vec{v})\right).$$

We will think about i-morphisms modulo i-equality without dividing this equivalence relation out.

In the obvious way, we can define the identity i-morphism $\mathsf{Id}_K : K \to K$, composition of i-morphisms, i-isomorphisms, and so forth. All these operations preserve i-equality. We easily see that i-isomorphisms really are isomorphisms in the category given by these operations.

We will say that two interpretations K, M are *i-equivalent* when there is an i-isomorphism between them, that is, they are i-isomorphic. The notion of i-equivalence is our intended notion of sameness of interpretations. We will, however, *not* divide out i-equivalence. This enables us to use the notation τ_M meaningfully, to speak about the dimension of an interpretation, and so forth. Of course, we demand that operations on interpretations preserve i-equivalence. One may show that operations like identity, composition, $(\cdot)\langle\cdot\rangle(\cdot)$ do indeed preserve i-equivalence. Moreover, if K is i-equivalent to M, then $\overline{K} = \overline{M}$.

The category INT_1 is the category of theories (as objects) and interpretations modulo i-equivalence (as arrows). One may show that we have indeed defined a category. Two theories U and V are *bi-interpretable* if they are isomorphic in INT_1. Wilfrid Hodges calls this notion *homotopy*. See Hodges [16, p. 222].

Thus, U and V are bi-interpretable if there are interpretations $K : U \to V$ and $M : V \to U$, so that $M \circ K$ is i-isomorphic to ID_U and $K \circ M$ is i-isomorphic to ID_V. We call the pair K, M a *bi-interpretation* between U and V. One can show that the components of a bi-interpretation are faithful interpretations. Many good properties of theories like finite axiomatizability, decidability, κ-categoricity are preserved by bi-interpretations.

2.4 Sequential theories The sequential theories form an important class of theories in this paper. A sequential theory provides an interpretation N of a weak number theory, say, S^1_2, and sequences of all objects of the domain of the theories with projections in N. We can use these sequences to develop partial satisfaction predicates. Using these we can prove restricted consistency statements of U in U. See Section 2.5 for more about this.

The notion of sequential theory has a very simple definition discovered by Pavel Pudlák. We first need the definition of a very weak set theory. The theory adjunctive set theory or AS is a one-sorted theory with a binary relation $\in$:

($\mathsf{AS}1$) $\vdash \exists x\, \forall y\; y \notin x$,
($\mathsf{AS}2$) $\vdash \forall x, y\, \exists z\, \forall u\, (u \in z \leftrightarrow (u \in x \vee u = y))$.

We note that we do not demand extensionality. For example, in AS we could have lots of "empty sets."

An interpretation is *direct* if and only if it is one-dimensional, unrelativized (i.e., it has the trivial domain), and identity preserving (i.e., it translates identity to identity).

A theory U is sequential if and only if it directly interprets AS. By a substantial bootstrap, we can define, in a sequential theory U, an interpretation N of a weak number theory, sequences of all objects, and so forth.

For details see, for example, Pudlák [24], [25], Mycielski, Pudlák, and Stern [22], Hájek and Pudlák [15], and Visser [37], [40].

We can generalize the notion of sequentiality a bit to *polysequentiality* by replacing *direct interpretation* in the definition by its obvious generalization to the m-dimensional case. All results in this paper that we prove for sequential theories also hold for polysequential theories.

2.5 Complexity and satisfaction *Restricted provability* plays an important role in this paper. An n-proof is a proof from axioms with Gödel number smaller than or equal to n only involving formulas of complexity smaller than or equal to n. To work conveniently with this notion, a good complexity measure ρ is needed. This should satisfy three conditions. (i) Eliminating terms in favor of a relational formulation should raise the complexity only by a fixed standard number. (ii) Translation of a formula via the translation corresponding to an interpretation K should raise the complexity of the formula by a fixed standard number depending only on K. (iii) The tower of exponents involved in cut elimination should be of height linear in the complexity of the formulas involved in the proof.

Such a good measure of complexity together with a verification of desideratum (iii)—a form of nesting degree of quantifier alternations—is supplied in the work of Philipp Gerhardy [12], [13]. It is also provided by Samuel Buss in his preliminary draft [7]. Buss also proves that (iii) is fulfilled. We give some details about these measures in Appendix A.

We will use $\mathsf{proof}_{U,n}$ for the proof predicate where only U-axioms with Gödel numbers $\leq n$ are allowed and where the formulas occurring in the proof are in the complexity class Γ_n of all formulas of complexity $\leq n$. Similarly we use $U \vdash_n A$, $\mathsf{con}_n(U)$, $\Box_{U,m} A$, and so forth.

In sequential theories we can define partial satisfaction predicates for formulas with complexity below n, for any n. The presence of these predicates has as a consequence that for any sequential theory U and for any n, we can find an interpretation N of a weak arithmetic like Buss's S^1_2 in U such that $U \vdash \mathsf{con}^N_n(U)$. See, for example, Visser [35] for more details.

3 The Arithmetics of a Theory

There are many heuristic ways to look at interpretations. For example, an interpretation is a uniform internal model construction. In the case of definitional extensions an interpretation is an enrichment of the interpreting or target theory. In this paper, we opt for a third heuristic: we view an interpretation as *the interpreted theory in the context of the interpreting theory*.

We will say that an interpretation $N : \mathsf{S}^1_2 \to U$ is an *arithmetic in* U or an *arithmetic of* U. The theory S^1_2 is Buss's theory of p-time computability (see [6]). We stipulate that we work with a version of S^1_2 that is formulated in the language of arithmetic with (the relational versions of) 0, S, $+$, and $\times$.

Remark 3.1 In our definition of arithmetic we are both rewarded and punished for having a strict typing regime of interpretations. The reward is that the target theory or interpreting theory or context is part of the data for an arithmetic. So we can speak about an arithmetic N without mentioning the context. The punishment is that, for example, an interpretation $K : \mathsf{EA} \to U$ is *not* an arithmetic. The associated arithmetic is $\mathsf{S}^1_2 \downarrow K$. See also Remark A.1.

There are several reasons for choosing S^1_2. First, it is a sequential theory. Second, the usual metamathematics leading up to Gödel's incompleteness theorems can be formalized in S^1_2 without the use of any extraneous tricks. Moreover, it is a reasonably weak choice among theories in which this can be done. Third, S^1_2 is finitely axiomatizable. On the other hand, the results of the present paper are rather robust with respect to different choices of the basic arithmetic. For example, T^1_2 or $\mathrm{I}\Delta_0 + \Omega_1$ would have worked as well. (But often some extra care is needed for $\mathrm{I}\Delta_0 + \Omega_1$, since it is not known whether this theory is finitely axiomatizable.)

The main structure between arithmetics that we will consider is the initial embedding ordering $\preceq$. Consider two arithmetics N and N' in U. An initial embedding $F : N \to N'$ is an i-morphism satisfying the following additional property:

- $U \vdash (F(\vec{x}_0, \vec{y}_0) \wedge \vec{y}_1 <_{N'} \vec{y}_0) \to \exists \vec{x}_1 <_N \vec{x}_0 \; F(\vec{x}_1, \vec{y}_1)$.

We write $N \preceq N'$ for: there is an initial embedding $\langle N, F, N' \rangle$ of N in N'. We note that $\preceq$ is preserved by i-equivalence. In other words, i-equivalence is a congruence relation for the arithmetics of U with $\preceq$. So, i-equivalence is a subrelation of the induced equivalence relation of $\preceq$.

We call N a *cut* of N' if and only if $\mathsf{emb} : N \preceq N'$, where emb is the identical embedding.

The most salient fact about $\preceq$ is upward preservation of Σ_1-sentences and downward preservation of Π_1-sentences. We formulate this as a theorem.

Theorem 3.2 *Suppose that N and N' are arithmetics in U and $N \preceq N'$. Let S be any Σ_1-statement, and let P be any Π_1-statement. We have $U \vdash S^N \to S^{N'}$ and $U \vdash P^{N'} \to P^N$.*

We leave the trivial proof to the reader. Arithmetics commute in all the right ways with bi-interpretations, as is shown in the next theorem.

Theorem 3.3 *Suppose that $K : U \to V$ and $M : V \to U$ are a bi-interpretation between U and V. Then the mapping $\Phi : N \mapsto NK$ is a bijection between the arithmetics of U and the arithmetics of V modulo i-equivalence. Moreover, Φ is an isomorphism with respect to $\preceq$ and $\overline{N} = \overline{\Phi(N)}$.*

Proof Let $\Psi : N' \mapsto N'M$ be a mapping between the arithmetics of V and the arithmetics of U. It is easy to see that Ψ is the inverse of Φ, modulo i-equivalence. Clearly Φ and Ψ preserve $\preceq$, so it easily follows that Φ is an isomorphism with respect to $\preceq$. Since the interpretations of a bi-interpretation are faithful, we find that $\overline{N} = \overline{\Phi(N)}$. □

Arithmetics of a sequential theory can always be assumed to be one-dimensional, as is formulated in the following theorem.

Theorem 3.4 *Suppose that U is sequential and that N is an arithmetic in U. Then there is a one-dimensional arithmetic M in U that is i-equivalent to N.*[3]

We leave the trivial proof to the reader. A fundamental fact about the arithmetics of a sequential theory follows by Pavel Pudlák's adaptation of Dedekind's proof of his categoricity theorem for second-order arithmetic.

Theorem 3.5 (Pudlák) *Consider a sequential theory U. Let N_0 and N_1 be arithmetics in U. Then, there is an arithmetic M in U such that $M \preceq N_0$ and $M \preceq N_1$.*

For a proof see, for example, [25]. In a sequential theory we have a convenient reflection principle. We write Γ_n for the class of all formulas of complexity $\leq n$.

Theorem 3.6 *Consider any sequential theory U, and let N be an arithmetic in U. For any n, we can find an arithmetic $M \preceq N$ such that, verifiably in S^1_2, we have, for all sentences A in Γ_n, that $U \vdash \Box^M_{U,n} A \to A$.*

Sketch of the proof The idea of the proof is that, in U, we can define a satisfaction predicate for Γ_n, using the N-numbers, and prove Γ_n-reflection by replacing induction over proof length by the use of a definable cut M of N. For details see the proof of [35, Fact 2.4.5(ii)]. In [35] only verifiability of this fact in EA was claimed. However, the big disjunctions and conjunctions of exponential size used there are not needed, since for each proof p we only need the truth of the axioms occurring in p. So the disjunctions we really need are polynomial in p. ☐

As far as we can ascertain, this theorem was known (or versions of it were known), at an early stage, to, independently, Pavel Pudlák, Robert Solovay, Alex Wilkie and Jeff Paris, and Harvey Friedman. The paper [25] contains a version.

The previous theorem shows that, for any n, we can "improve" a given N to obtain n-reflection. In contrast, if U is finitely axiomatized, for any N, we can find an n such that, for any $m \geq n$, we have anti-m-reflection, in other words, a version of Löb's theorem for N and m. We first need [35, Lemma 4.1]. For the convenience of the reader we reproduce it here.

Lemma 3.7 *The following fact is verifiable in S^2_1. Suppose that A is any finitely axiomatized theory, $\rho(A) \leq m$, $\rho(B) \leq m$, and $A \vdash B$. Then $\mathsf{S}^1_2 \vdash \Box_{A,m} B$.*

We note that without the verifiability clause we could conclude $A \vdash_m B$ from $A \vdash B$. Since this step uses superexponentiation, it is not available in the context of S^1_2.

Proof We can prove the lemma in two ways.

The first uses the insight of Pudlák [26, Lemma 2.2] that, in S^1_2, we have, for all x and y, that $\mathsf{S}^1_2 \vdash \mathsf{itexp}(x, |y|)$ exists. Here we define

- $\mathsf{itexp}(x, 0) := x$,
- $\mathsf{itexp}(x, z + 1) := 2^{\mathsf{itexp}(x,z)}$.

Suppose $A \vdash B$. By $\exists\Sigma^{\mathsf{b}}_1$-completeness, we find, for some p, $\mathsf{S}^1_2 \vdash \mathsf{proof}_A(p, B)$. Since we have that $\mathsf{S}^1_2 \vdash \mathsf{itexp}(kp, k'|p|)$ exists, for standard k, k', we can apply cut elimination to p inside S^1_2.

The second way is to note that, in $\mathsf{S}^1_2 + \mathsf{con}_m(A + \neg B)$, we can build a Henkin interpretation H of $A + \neg B$. It follows that $\mathsf{con}(\mathsf{S}^1_2 + \mathsf{con}_m(A + \neg B))$ implies $\mathsf{con}(A + \neg B)$. We find the desired result by contraposition. ☐

Theorem 3.8 *Suppose that A is a finitely axiomatized sequential theory and that N is an arithmetic in A. We can find an n such that for any $m \geq n$, we have, for all $B \in \Gamma_m$, if $A \vdash \Box^N_{A,m} B \to B$, then $A \vdash B$. This fact is verifiable in S^1_2.*

This theorem is a weaker version of [35, Theorem 4.1]. We sketch the proof since it is easier to read without the ballast of the stronger version of [35].

Proof The proof is just the usual proof of Löb's theorem with some checks that all the complexities are correct and one step involving Lemma 3.7 added. We choose

$$n := \mathsf{max}\big(\rho\big(\mathsf{prov}^N_{A,y}(z)\big) + 1, \rho\big(\mathsf{sub}^N(x,y)\big) + 1, \rho(A)\big).$$

Here sub is the formula defining the Gödel substitution function. It follows that $n \geq \Box^N_{A,m} B$, for any B and m, since both B and n appear as numerals and, thus, only add a nonalternating block of quantifiers.

Let C be a Gödel fixed point with $A \vdash C \leftrightarrow (\Box^N_{A,m} C \to B)$. The complexity of C is again m as can be seen by inspecting the construction.

Note that, for example, $\rho(\mathsf{prov}^N_{A,y}(z))$ is polynomial in the data for N. Suppose $A \vdash \Box^N_{A,m} B \to B$. We have

$$\begin{aligned} A \ \ \vdash \Box^N_{A,m} C &\to \big(\Box^N_{A,m}\Box^N_{A,m} C \wedge \Box^N_{A,m}(\Box^N_{A,m} C \to B)\big) \\ &\to \Box^N_{A,m} B \\ &\to B. \end{aligned}$$

So, (a) $A \vdash \Box^N_{A,m} C \to B$, and, hence, $A \vdash C$. By Lemma 3.7, we find that (b) $A \vdash \Box^N_{A,m} C$. Combining (a) and (b), we may conclude that $A \vdash B$. □

Theorem 3.9 *Consider a theory U and an arithmetic N in U. Then, there is an arithmetic $N' \preceq N$ and a U-formula* TRUE *such that, for Σ_1-sentences S, we have $U \vdash \mathsf{TRUE}(S) \leftrightarrow S^{N'}$. (Here S is coded in N'.)*

Proof We first work S^1_2. Let $\mathsf{sat}(v)$ be a Δ_0-satisfaction predicate for formulas with just one designated variable v free. The main ingredients for the construction of such a predicate can be found in [15, Chapter V(5)].[4] We will use the following two properties of sat, for $D(v)$ in Δ_0:

(S1) $\mathsf{S}^1_2 \vdash \forall x\, (\mathsf{sat}(x, D(v)) \to D(x))$,
(S2) $\mathsf{S}^1_2 \vdash \forall x\, ((2^{2^x} \text{ exists } \wedge D(x)) \to \mathsf{sat}(x, D(v)))$.

Let J be an S^1_2-cut such that $\mathsf{S}^1_2 \vdash x \in J \to 2^{2^x}$ exists. We suppose that Σ_1-sentences are written in the form $\exists x\, S_0(x)$ where S_0 is Δ_0. (If not, we add an algorithm that rewrites the Σ_1-sentence to this normal form.) We define

- $\mathsf{true}(\exists x\, S_0(x)) := \exists x \in J\ \mathsf{sat}(x, S_0(v))$,
- $N' := N \circ J$,
- $\mathsf{TRUE}(x) := \mathsf{true}^N(x)$.

We easily verify that TRUE has the desired property. □

Inspection of the proof of Theorem 3.9 shows that we can obtain reasonable commutation properties for TRUE in addition to mere Tarskian disquotation.

Remark 3.10 Suppose that U is sequential. Let N be an arithmetic with parameters in U. In a model $\mathcal{M}$ of U we can view N as a definable class of internal models parameterized by models of U. Theorem 3.5 tells us how to construct a parameter-free arithmetic below two parameter-free arithmetics. With some care we can generalize the construction to produce one parameter-free arithmetic below N viewed as a class of internal models. For details on such a construction, see [40], the second proof of Theorem 5.2. As a result of this observation, the provability logic of all parameter-free arithmetics of U is the same as the provability logic of all arithmetics of U with parameters.

We end this section with a tentative discussion of what it means that $\preceq$ has a minimal element.

Theorem 3.11 *Consider any theory U. Suppose that N is a $\preceq$-minimal arithmetic in U, that is, for any arithmetic M in U with $M \preceq N$, we have $N \preceq M$. Then, we have the following:*

(i) For any Σ_1-sentence S, and any $M \preceq N$, we have $U \vdash S^N \to S^M$.
(ii) U proves parameter-free Π_1-induction for the N-numbers. In other words, we have $U \vdash (\mathrm{I}\Pi_1^-)^N$. As a consequence, we have sentential Σ_1-completeness in N.
(iii) We have a Σ_1-truth predicate TRUE satisfying Tarskian disquotation for Σ_1-sentences on N.

Proof Ad (i). We have (i) simply because if $M \preceq N$, then $M \preceq N$, and Σ_1-sentences are upwards preserved.

Ad (ii). As is easily seen, parameter-free Π_1-induction is equivalent to the parameter-free Σ_1-minimum principle (over PA^-).[5] We prove the parameter-free Σ_1-minimum principle for N. We reason in U. Suppose $\exists x \in N\, S(x)$, where S is Σ_1. Consider the virtual class $X := \{x \in N \mid \forall y < x\, \neg S(x)\}$. Clearly $0 \in X$. If X is not closed under successor, there is a $z \in N$ such that $z \in X$ and $\mathsf{S}z \notin X$. By elementary reasoning we find that z is the minimal N-number such that Sz. If X is not closed under successor, we can shorten X to a cut J that satisfies S^1_2. Thus J is an arithmetic below N. It follows that on J we have both $\neg\exists x\, Sx$ and $\exists x\, Sx$, a contradiction.

The theory $\mathrm{I}\Pi_1^-$ proves sentential Σ_1-completeness since EA is conservative over $\mathrm{I}\Pi_1^-$ with respect to Σ_2-sentences as was proved in Kaye, Paris, and Dimitracopoulos [18].

Ad (iii). The existence of the desired truth predicate is immediate from Theorem 3.9. □

Theorem 3.12 *Consider any sequential theory U. Suppose that N is a $\preceq$-minimal arithmetic in U. It follows that*

(i) N is a $\preceq$-minimum in U; in other words, for all arithmetics M in U, $N \preceq M$;
(ii) U is parameter-free essentially reflexive for N; that is, for any n, and any sentence $B \in \Gamma_n$, we have $U \vdash \Box^N_{U,n} B \to B$;
(iii) U is not finitely axiomatizable.

Proof Claim (i) is immediate by Theorem 3.5, and (ii) follows from Theorem 3.6 in combination with Theorem 3.11(i). Claim (iii) is immediate from (ii) in combination with Theorem 3.8. □

Remark 3.13 Consider a sequential theory U, and suppose that N is $\preceq$-minimal in U. It follows that the interpretability logic of U (for sentential substitutions), with respect to arithmetization in N, is ILM. See Beklemishev and Visser [2] for most ingredients of the proof.

Open Question 3.14 Suppose that U is sequential and has a $\preceq$-minimal arithmetic N. Can we get a precise estimate of what this implies? For example, can one show that we do not get full induction for N? Is any $M \preceq N$ i-equivalent to N? Such questions are both interesting in general and in the sequential case.

4 Introduction to Provability Logic

We start with the basics concerning Löb's logic GL. We define the language $\mathcal{L}_{\mathsf{mod}}$ of propositional modal logic by

- $\alpha ::= p_0 \mid p_1 \mid \ldots,$
- $\varphi ::= \alpha \mid \bot \mid \top \mid \neg\varphi \mid \Box\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \to \varphi).$

The logic GL is axiomatized by the following axioms and rules:

(GL1) we have all substitution instances of propositional tautologies;
(GL2) $\vdash \Box(\varphi \to \psi) \to (\Box\varphi \to \Box\psi)$;
(GL3) $\vdash \Box\varphi \to \Box\Box\varphi$;
(GL4) $\vdash \Box(\Box\varphi \to \varphi) \to \Box\varphi$;
(GL5) if $\vdash \varphi \to \psi$ and $\vdash \varphi$, then $\vdash \psi$;
(GL6) if $\vdash \varphi$, then $\vdash \Box\varphi$.

We have a completeness theorem for GL in finite, transitive, irreflexive Kripke models. We define arithmetical interpretations of the modal language as follows. Let U be a theory, and let N be an arithmetic in U. We define an N-translation σ as a mapping of the formulas of $\mathcal{L}_{\mathsf{mod}}$ to the sentences of the language of U, where σ commutes with the propositional connectives and where

$$\sigma(\Box\varphi) := \Box_U^N \sigma(\varphi) := \mathsf{prov}_U^N(\underline{\ulcorner\sigma(\varphi)\urcorner}).$$

The variable a will range over $0, 1, \ldots, \infty$. We define

- $\Box^0\varphi := \varphi$, $\Box^{n+1}\varphi := \Box\Box^n\varphi$, $\Box^\infty\varphi := \top$;
- $\varphi \in \mathsf{prl}(N)$ if and only if, for all N-translations σ, $U \vdash \sigma(\varphi)$;
- $\varphi \in \mathsf{prl}_{\mathsf{all}}(U)$ if and only if, for all arithmetics N in U, $\varphi \in \mathsf{prl}(N)$;
- $\mathsf{deg}(N) := \mathsf{min}(\{a \mid \Box^a\bot \in \mathsf{prl}(N)\})$;
- $\mathsf{deg}_{\mathsf{all}}(U) := \mathsf{min}(\{a \mid \Box^a\bot \in \mathsf{prl}_{\mathsf{all}}(U)\})$;
- in the case when U is an extension of S^1_2 in the language of arithmetic, we write $\mathsf{prl}(U)$ for $\mathsf{prl}(\mathcal{E}_{\mathsf{S}^1_2\, U})$ and $\mathsf{deg}(U)$ for $\mathsf{deg}(\mathcal{E}_{\mathsf{S}^1_2\, U})$.

We note that $\mathsf{deg}_{\mathsf{all}}(U) := \mathsf{sup}(\{\mathsf{deg}(N) \mid N \text{ is an arithmetic in } U\})$. We have the following two major insights. Let exp be the axiom stating that exponentiation is total.

Theorem 4.1 *Consider any theory U. Let N be an arithmetic in U. We have*

(I) $\mathsf{prl}(N)$ *extends* GL *(and is closed under the rules of* GL*);*
(II) if $U \vdash \mathsf{exp}^N$*, then* $\mathsf{prl}(N) = \mathsf{GL} + \Box^{\mathsf{deg}(N)}\bot$.

In essence the proof of (I) is given in [6]. Most of the ideas are also in Wilkie and Paris [42]. Robert Solovay proved (II) for theories like PA which are reasonably strong and Σ_1-sound. The extension to the case of Σ_1-unsound theories extending PA was proved in Visser [39]. The fact that EA was needed on the designated interpretation of arithmetic slowly emerged (see [30]). In Section 5 we give a sharper formulation of Theorem 4.1(II).

The gap between (I) and (II) provides the great open problem of provability logic. What happens in the gap? For an extensive discussion of this problem, see Beklemishev and Visser [3].

Is the provability logic of an arithmetic a good property of arithmetics? It should at least be preserved under our chosen notion of sameness of arithmetics. We note that, if N is i-equivalent to N', then S^1_2 verifies this i-equivalence. The next theorem follows.

Theorem 4.2 *Consider any theory U, and suppose that N and N' are arithmetics in U and that N is i-equivalent to N'. Then, for any N-translation σ and N'-translation σ', if $\sigma(p) = \sigma'(p)$ for all atoms p, then, for all φ, we have $U \vdash \sigma(\varphi) \leftrightarrow \sigma'(\varphi)$. It follows that $\mathsf{prl}(N) = \mathsf{prl}(N')$ and $\mathsf{deg}(N) = \mathsf{deg}(N')$. We have the same result on the weaker assumption that $N \preceq N'$ and $N' \preceq N$.*

So, in this sense, the provability logic of an arithmetic is a good property. On the other hand, the provability logic of a theory-in-extension is dependent on the specification of the axiom set. The provability logic of a theory is *intensional*.[6]

Example 4.3 Consider the theory $U := \mathsf{PA} + \Box_{\mathsf{PA}}\Box_{\mathsf{PA}}\bot$ with the obvious axiomatization. Clearly $U \vdash \Box_U\Box_U\bot$. On the other hand, suppose $U \vdash \Box_U\bot$. Then,

$$\mathsf{PA} + \Box_{\mathsf{PA}}\Box_{\mathsf{PA}}\bot \vdash \Box_{\mathsf{PA}+\Box_{\mathsf{PA}}\Box_{\mathsf{PA}}\bot}\bot.$$

So,

$$\mathsf{PA} + \Box_{\mathsf{PA}}\Box_{\mathsf{PA}}\bot \vdash \Box_{\mathsf{PA}}\neg\Box_{\mathsf{PA}}\Box_{\mathsf{PA}}\bot.$$

And, hence,

$$\mathsf{PA} + \Box_{\mathsf{PA}}\Box_{\mathsf{PA}}\bot \vdash \Box_{\mathsf{PA}}\neg\Box_{\mathsf{PA}}\bot.$$

By applying the formalized second incompleteness theorem to the conclusion, we get: $\mathsf{PA} + \Box_{\mathsf{PA}}\Box_{\mathsf{PA}}\bot \vdash \Box_{\mathsf{PA}}\bot$. By Löb's theorem, we obtain $\mathsf{PA} \vdash \Box_{\mathsf{PA}}\bot$. Quod non. So $U \nvdash \Box_U\bot$.

Now we modify the formula defining U, thus obtaining the theory $\tilde{U}$, by taking something to be an axiom if it is an axiom of U or it is of the form $\underline{p} \neq \underline{p}$, where p is a PA-proof of $\Box_{\mathsf{PA}}\bot$. Clearly, U and $\tilde{U}$ are extensionally equal. We easily see that $\mathsf{PA} + \Box_{\mathsf{PA}}\Box_{\mathsf{PA}}\bot \vdash \Box_{\tilde{U}}\bot$, and, hence, $\tilde{U} \vdash \Box_{\tilde{U}}\bot$. We conclude that $\mathsf{deg}_{\mathsf{all}}(U) = \mathsf{deg}(U) = 2$ and $\mathsf{deg}_{\mathsf{all}}(\tilde{U}) = \mathsf{deg}(\tilde{U}) = 1$.

We note that the arithmetics in our example are Σ_1-unsound. It is unknown whether we can find two extensionally equal theories V and V' and two arithmetics $N := \langle \mathsf{S}^1_2, \tau, V\rangle$ and $N' := \langle \mathsf{S}^1_2, \tau, V'\rangle$ such that N and, a fortiori, N' are Σ_1-sound that give rise to different provability logics. In the case $V \vdash \mathsf{exp}^N$, where exp is the axiom stating that exponentiation is total, we will see that $\mathsf{prl}_{\mathsf{all}}(V) = \mathsf{prl}(N) = \mathsf{GL} = \mathsf{prl}(N') = \mathsf{prl}_{\mathsf{all}}(V')$. So any counterexamples for $\mathsf{prl}(N)$ and $\mathsf{prl}(N')$ should fail to prove the totality of exponentiation for N, and any counterexamples for $\mathsf{prl}_{\mathsf{all}}(V)$ and $\mathsf{prl}_{\mathsf{all}}(V')$ should not contain any Σ_1-sound arithmetic M that proves the principle exp^M.

Because of intensionality, provability logics and degrees need not be preserved by bi-interpretation. To get the appropriate notion of sameness that preserves provability logics and degrees we consider bi-interpretability for S^1_2-verifiable interpretations:

- $K : U \rhd V$ is S^1_2-verifiably an interpretation, if $\mathsf{S}^1_2 \vdash \forall A\,(\Box_V A \to \Box_U A^K)$.

We have chosen the formalized version of the *theorems* formulation of interpretability. This is convenient but not really necessary. As Emil Jeřábek pointed out to me, Buss's witnessing theorem implies that S^1_2-verifiable axioms-interpretability implies S^1_2-verifiable theorems-interpretability.

Theorem 4.4 *Suppose that $K : U \to V$ and $M : V \to U$ form a bi-interpretation. Suppose further that both K and M are S^1_2-verifiably interpretations. Let N be an arithmetic in U. Then, $K \circ N$ is an arithmetic in V. We have* $\mathsf{prl}(K \circ N) = \mathsf{prl}(N)$. *It follows that* $\mathsf{prl}(U) = \mathsf{prl}(V)$. *Similarly for the* deg.

Proof We first note that (‡) U proves, for some F that "F is an isomorphism between ID_U and $M \circ K$." It follows that S^1_2 verifies the formalization of (‡). Similarly for the isomorphism between ID_V and $K \circ M$. Thus, we may conclude that (†) $\mathsf{S}^1_2 \vdash \Box_U A \leftrightarrow \Box_V A^K$, and so forth.

Suppose that σ is an N-translation. We prove by induction on φ that, for all φ, we have $V \vdash (\sigma(\varphi))^K \leftrightarrow (K \circ \sigma)(\varphi)$. The only interesting case is when $\varphi = \Box\psi$. We have

$$
\begin{aligned}
V \vdash \quad \big(\sigma(\Box\psi)\big)^K &\leftrightarrow \big(\Box^N_U \sigma(\psi)\big)^K && (1)\\
&\leftrightarrow \Box^{K\circ N}_U \sigma(\psi) && (2)\\
&\leftrightarrow \Box^{K\circ N}_V \big(\sigma(\psi)\big)^K && (3)\\
&\leftrightarrow \Box^{K\circ N}_V (K \circ \sigma)(\psi). && (4)
\end{aligned}
$$

We note that step (3) uses (†) and that step (4) uses the induction hypothesis.

Suppose that φ is in $\mathsf{prl}(K \circ N)$; then, for any N-translation σ, we have $V \vdash (K \circ \sigma)(\varphi)$. Ergo, $V \vdash \sigma(\varphi)^K$. Hence, $U \vdash \sigma(\varphi)$. It follows that $\varphi \in \mathsf{prl}(N)$.

Conversely, suppose that φ is in $\mathsf{prl}(N)$. Then, by Theorem 4.2, it follows that φ is in $\mathsf{prl}(M \circ K \circ N)$. By the above argument applied with V and U and K and M interchanged and with $K \circ N$ in the role of N, we find the following. If φ is in $\mathsf{prl}(K \circ N)$, then φ is in $\mathsf{prl}(M \circ K \circ N)$. Hence, if φ is in $\mathsf{prl}(K \circ N)$, then φ is in $\mathsf{prl}(N)$. □

Thus, if U and V are bi-interpretable via S^1_2-verifiable interpretations, then the interpretations provide an isomorphism between their arithmetics that preserves $\preceq$ and deg and prl.

5 Solovay's Theorem

In this section, we study the forms that Solovay's theorem takes in various settings.

5.1 The Guaspari–Solovay system R^- In this subsection we give a careful analysis of the proof of Solovay's theorem. We follow the modal presentation of the proof due to Dick de Jongh, Marc Jumelet, and Franco Montagna in their paper [9].

We introduce the logic R^- of Guaspari and Solovay [14] and some subsystems of this logic. The language of R^- is given by

- $\alpha ::= p_0 \mid p_1 \mid \cdots$,
- $\varphi ::= \alpha \mid \bot \mid \top \mid \neg\varphi \mid \Box\varphi \mid (\varphi \wedge \varphi) \mid (\varphi \vee \varphi) \mid (\varphi \to \varphi) \mid \Box\varphi < \Box\varphi \mid \Box\varphi \leq \Box\varphi$.

The logic R^- is axiomatized by the axioms and rules of GL (for the extended language) plus the following axioms:

(R^-1) $\vdash \Box\varphi \leq \Box\psi \to \Box\varphi$,
(R^-2) $\vdash (\Box\varphi \leq \Box\psi \wedge \Box\psi \leq \Box\chi) \to \Box\varphi \leq \Box\chi$,
(R^-3) $\vdash \Box\varphi < \Box\psi \leftrightarrow (\Box\varphi \leq \Box\psi \wedge \neg\,\Box\psi \leq \Box\varphi)$,
(R^-4) $\vdash \Box\varphi \to (\Box\varphi \leq \Box\psi \vee \Box\psi \leq \Box\varphi)$,

(R^-5) $\vdash \Box\varphi \leq \Box\psi \to \Box(\Box\varphi \leq \Box\psi)$,
(R^-6) $\vdash \Box\varphi < \Box\psi \to \Box(\Box\varphi < \Box\psi)$.

We can split Axiom R^-4 into two parts that are jointly equivalent to R^-4:

(R^-4a) $\vdash \Box\varphi \leq \Box\varphi \to (\Box\varphi \leq \Box\psi \vee \Box\psi \leq \Box\varphi)$,
(R^-4b) $\vdash \Box\varphi \to \Box\varphi \leq \Box\varphi$.

We will consider two subsystems of R^-, to wit, R^-_0 and R^-_1. R^-_0 is given by the axioms and rules of GL (for the extended language) plus R^-1, R^-2, R^-3, R^-4a, and R^-_1 is R^-_0 plus R^-4b; in other words, R^-_1 is given by R^-1, R^-2, R^-3, R^-4.

5.2 Arithmetical interpretations of R^- Consider any arithmetical theory U and any arithmetic N in U. We specify what it is to be an interpretation of the language of R^- for U, N.

We remind the reader of the witness comparison notation. We define, for any $C = \exists x\, C_0(x)$ and $D = \exists y\, D_0(y)$:

- $C \leq D := \exists x\, (C_0(x) \wedge \forall y < x\, \neg D_0(y))$,
- $C < D := \exists x\, (C_0(x) \wedge \forall y \leq x\, \neg D_0(y))$,
- $(C \leq D)^\perp := (D < C)$.

We interpret the language of R^- as follows. An N-translation σ sends the propositional variables to U-sentences, commutes with the propositional connectives, and satisfies

- $\sigma(\Box\varphi) = \mathsf{prov}_U^N(\ulcorner \underline{\sigma(\varphi)} \urcorner)$,
- $\sigma(\Box\varphi \leq \Box\psi) = \mathsf{prov}_U^N(\ulcorner \underline{\sigma(\varphi)} \urcorner) \leq \mathsf{prov}_U^N(\ulcorner \underline{\sigma(\psi)} \urcorner)$,
- $\sigma(\Box\varphi < \Box\psi) = \mathsf{prov}_U^N(\ulcorner \underline{\sigma(\varphi)} \urcorner) < \mathsf{prov}_U^N(\ulcorner \underline{\sigma(\psi)} \urcorner)$.

We assume that we are employing an ordinary single conclusion proof predicate. A modal formula φ is N-valid if for all N-translations σ, we have $U \vdash \sigma(\varphi)$.

It is easily seen that the theory R^-_0 is N-valid, for any N. The principles R^-4b, on the one hand, and R^-5 and R^-6, on the other, are not known to be N-valid, for example, in the case $N = \mathsf{ID}_{\mathsf{S}^1_2}$.

The principle R^-4b is a modal articulation of a special case of the minimum principle. It tells us that if a certain sentence has a proof, then it has a minimal proof. Since the proof predicate is $\Delta^{\mathsf{b}}_1(\mathsf{S}^1_2)$, a reasonable principle to ask for is the Σ^{b}_1-minimum principle, that is, Buss's minimization axiom Σ^{b}_1-MIN. By the results of [6, Section 2.9], this principle is equivalent over S^1_2 with Σ^{b}_1-induction, that is, Buss's principle Σ^{b}_1-IND. This means that a salient not-too-strong theory in which we have R^-4b is the theory T^1_2. Thus, if $U \vdash (\mathsf{T}^1_2)^N$, then R^-_1 is N-valid.

The principles R^-5 and R^-6 are instances of sentential $\exists\Pi^{\mathsf{b}}_1$-completeness. Is there a natural theory extending T^1_2 that is as weak as possible that delivers this principle? Of course, $\mathsf{B}_0 := \mathsf{T}^1_2 + \{S \to \Box_{\mathsf{S}^1_2} S \mid S \in \exists\Pi^{\mathsf{b}}_1\text{-sent}\}$ does the trick, but this principle involves coding. Let $\mathcal{J} := \{x \mid 2^x \downarrow\}$. We define

$$\mathsf{B} := \mathsf{T}^1_2 + \{S \in \exists\Pi^{\mathsf{b}}_1\text{-sent} \mid S \to S^{\mathcal{J}}\}.$$

Since the proof of completeness for $\exists\Pi^{\mathsf{b}}_1$-sentences only needs one exponent, we find $\mathsf{B} \vdash S \to \Box_{\mathsf{S}^1_2} S$, for S a $\exists\Pi^{\mathsf{b}}_1$-sentence. So B extends B_0.[7] Thus, if $U \vdash \mathsf{B}^N$, then R^- is N-valid.

The two main theorems concerning provability logic of this paper, to wit Theorems 5.11 and 5.15, will employ T^1_2 to ensure the principle R^-4b. In contrast, we will not use B to obtain R^-5 and R^-6. This last theory is, in a sense, still too strong. The theory T^1_2 is interpretable in S^1_2 on a cut, but B is locally but not globally interpretable in S^1_2. See Remark 5.1.

Remark 5.1 The two minimal *salient* theories in the literature in which we have B_0 are EA and $\mathrm{I}\Pi^-_1$. Since $\mathrm{I}\Pi^-_1$ does not fit our framework, we will consider $\mathrm{I}\Pi^-_1 + \Omega_1$ instead. The theory B is a subtheory of both EA and of $\mathrm{I}\Pi^-_1 + \Omega_1$. The theories EA and $\mathrm{I}\Pi^-_1$ have the same $\mathrm{B}\Sigma_1$-consequences (see [18] or [8]).

In [8, Theorem 1.3(2)] it is shown that the $\mathrm{B}\Sigma_1$-consequences of $\mathrm{I}\Pi^-_1$ and, hence, of EA can be axiomatized by the theory

$$\mathsf{CFL} := \mathrm{I}\Delta_0 + \{S \in \Sigma_1\text{-}\mathsf{sent} \mid S \to S^{\mathcal{J}}\}.^{8}$$

Clearly, $\mathsf{CFL} + \Omega_1$ extends B. The theories $\mathrm{I}\Pi^-_1 + \Omega_1$ and, a fortiori, $\mathsf{CFL} + \Omega_1$ and B are locally interpretable in S^1_2. The proof that $\mathrm{I}\Pi^-_1 + \Omega_1$ is locally interpretable in S^1_2 can be found in Visser [41]. Thus, they are *locally weak*. One can show that S^1_2 does not interpret B, so B is not a weak theory and, a fortiori, neither are $\mathsf{CFL} + \Omega_1$ and $\mathrm{I}\Pi^-_1 + \Omega_1$. To prove this one shows that $\mathsf{B} \vdash \mathsf{con}_n(\mathsf{S}^1_2)$, for every n. By the results of [25], S^1_2 cannot interpret $\mathsf{S}^1_2 + \{\mathsf{con}_n(\mathsf{S}^1_2) \mid n \in \omega\}$. See also, for example, Visser [38].

5.3 The basic proof In this subsection we present the version of Solovay's proof that is due to de Jongh, Jumelet, and Montagna.

Our first aim is to embed a finite Kripke frame for ordinary modal logic in the logic R^- extended with a finite set of constants and a finite set of axioms concerning these constants. Via the arithmetical validity of our modal theory this embedding subsequently induces an embedding in an arithmetic.

Let $\mathcal{F}$ be a finite, irreflexive, transitive Kripke frame on worlds $\{0, \ldots, n-1\}$. Our frame need not be rooted.

We write $i \parallel j$ for *i and j are incomparable*, that is, $i \not\preceq j$ and $j \not\preceq i$.

For $i = 0, \ldots, n-1$, we add constants ℓ_i to the language of R^-. Consider the following axioms:

($\mathcal{F}1$) $\vdash\ \ell_i \leftrightarrow (\Box\neg\,\ell_i \wedge \bigwedge_{j \succ i} \Diamond\ell_j \wedge \bigwedge_{j \parallel i}\ \bigvee_{k \preceq i,\, k \parallel j} \Box\neg\,\ell_k < \Box\neg\,\ell_j)$;
($\mathcal{F}2$) for $i \neq j$, we have $\vdash\ \Box\neg\,\ell_i \leq \Box\neg\,\ell_j\ \to\ \Box\neg\,\ell_i < \Box\neg\,\ell_j$.

We add these axioms to R^-_i and to R^-, thus obtaining $\mathsf{R}^-_{i,\mathcal{F}}$ and $\mathsf{R}^-_{\mathcal{F}}$. (Here we let the axioms and rules of $\mathsf{R}^-_{(i)}$ apply to the extended language with the new axioms.) We adhere to the usual convention that the empty conjunction is $\top$ and the empty disjunction is $\bot$.

The N-interpretation of these principles is given as follows. By the multiple fixed-point lemma we find sentences L_i such that

$$\mathsf{S}^1_2 \vdash\ L_i \leftrightarrow \Big(\Box_U\neg\, L^N_i \wedge \bigwedge_{j \succ i} \Diamond L^N_j \wedge \bigwedge_{j \parallel i}\ \bigvee_{k \preceq i,\, k \parallel j} \Box_U\neg\, L^N_k < \Box_U\neg\, L^N_j\Big).$$

We will assume that, for $i \neq j$, we have $L_i \neq L_j$.[9] We demand that $\sigma(\ell_i) := L^N_i$. Thus we treat the ℓ_i as constants.[10] For an arbitrary arithmetic N this gives us the validity of $\mathsf{R}^-_{0,\mathcal{F}}$.

Below we want to reason in an informal way in the theory $\mathsf{R}^-_{(1),\mathcal{F}}$. We want to reason as if we have predicate logic available, so that we can talk about statements like $i \preceq j$ and so that we can quantify over the nodes of F. These problems can be solved as follows. A statement like $i \preceq j$ in the context of $\mathsf{R}^-_{(1),\mathcal{F}}$ stands for $\top$ when it is true and for $\bot$ when it is false. Quantification over our finite domain is handled by translating it to iterated conjunctions and disjunctions.

We define

- $h_i := (\Box\neg\,\ell_i \wedge \bigwedge_{j \| i} \;\; \bigvee_{k \preceq i,\, k \| j} \Box\neg\,\ell_k < \Box\neg\,\ell_j)$.

Lemma 5.2 *In $\mathsf{R}^-_{1,\mathcal{F}}$ we have the following. Suppose $i \parallel j$; then $\neg\,(h_i \wedge h_j)$. In other words, h_i and h_j imply $i \preceq j$ or $j \preceq i$.*

Proof Reason in $\mathsf{R}^-_{1,\mathcal{F}}$. Suppose $i \parallel j$ and h_i and h_j. Consider the i', such that $i' \preceq i$, $i' \parallel j$, and $\Box\neg\ell_{i'}$. We note that there is such an i', to wit i, because $i \preceq i$, $i \parallel j$ and $\Box\neg\,\ell_i$. The $\Box\neg\ell_{i'}$ are linearly ordered in the witness comparison ordering $<$. Suppose that $\Box\neg\ell_{i^\star}$ is the $<$-smallest such element. Consider the j', such that $j' \preceq j$, $j' \parallel i$, and $\Box\neg\ell_{j'}$. The node j is an example of such a j'. The $\Box\neg\ell_{j'}$ are linearly ordered in the witness comparison ordering $<$. Suppose that $\Box\neg\ell_{j^\star}$ is the $\leq$-smallest such element.

By the second conjunct of h_i applied to $j^\star$, we find $\Box\neg\,\ell_{i^\star} < \Box\neg\,\ell_{j^\star}$. By the second conjunct of h_j applied to $i^\star$, we find $\Box\neg\,\ell_{j^\star} < \Box\neg\,\ell_{i^\star}$, a contradiction. $\Box$

Lemma 5.3 *In $\mathsf{R}^-_{1,\mathcal{F}}$ we have the following. Suppose $i \neq j$; then $\neg\,(\ell_i \wedge \ell_j)$.*

Proof Reason in $\mathsf{R}^-_{1,\mathcal{F}}$. In the case when i and j are incomparable, this is immediate by Lemma 5.2. Suppose, for example, $i \prec j$. Suppose ℓ_i and ℓ_j. From ℓ_i, we have $\Diamond\ell_j$, and, from ℓ_j, we have $\Box\neg\,\ell_j$, a contradiction. $\Box$

Lemma 5.4 *In $\mathsf{R}^-_{1,\mathcal{F}}$ we have the following result. Suppose h_i and $\neg\,\ell_i$. Then, for some $j \succ i$, we have h_j.*

Proof Reason in $\mathsf{R}^-_{1,\mathcal{F}}$. Suppose h_i and $\neg\,\ell_i$. Then for some $j' \succ i$, we have $\Box\neg\,\ell_{j'}$. The $\Box\neg\,\ell_{j'}$ with $j' \succ i$ can be linearly ordered by the witness comparison ordering $<$. Let $\Box\neg\,\ell_{j^\star}$ be the $<$-minimal element among these j'.

Consider any $m \parallel j^\star$. If $m \parallel i$, by h_i, we can find a k such that $k \preceq i \prec j^\star$, and $k \parallel m$ and $\Box\neg\,\ell_k < \Box\neg\,\ell_m$. If not $m \parallel i$, we must have $i \prec m$. In the case $\Box\neg\ell_m$, by the choice of $j^\star$, we find $\Box\neg\,\ell_{j^\star} < \Box\neg\,\ell_m$. In the case $\neg\,\Box\neg\,\ell_m$, the axioms of R^- imply $\Box\neg\,\ell_{j^\star} < \Box\neg\,\ell_m$. So in all cases we can find a k such that $k \preceq j^\star$ and $k \parallel m$ and $\Box\neg\,\ell_k < \Box\neg\,\ell_m$. We may conclude $h_{j^\star}$. $\Box$

Lemma 5.5 *In $\mathsf{R}^-_{1,\mathcal{F}}$ we have the following result. Suppose h_i; then, for some $j \succeq i$, we have ℓ_j.*

Proof Reason in $\mathsf{R}^-_{1,\mathcal{F}}$. Suppose h_i. If ℓ_i, we are done. If not, by Lemma 5.4, there is a $i' \succ i$ such that $h_{i'}$. If $\ell_{i'}$, we are done. By repeating this procedure, we eventually find a $j \succeq i$, such that ℓ_j. $\Box$

Lemma 5.6 *In $\mathsf{R}^-_{1,\mathcal{F}}$ we have the following result. Suppose $\Box\neg\,\ell_i$. Then, for some j, we have ℓ_j.*

Proof Reason in $\mathsf{R}^-_{1,\mathcal{F}}$. Suppose $\Box\neg\,\ell_i$. Consider all j' such that $\Box\neg\,\ell_{j'}$. There is one such j', to wit i. The $\Box\neg\,\ell_{j'}$ are linearly ordered by the witness comparison ordering $<$. Let $j^\star$ be the minimal such. It is easy to see that $h_{j^\star}$. By Lemma 5.5, we find a $j \succeq j^\star$ such that ℓ_j. $\Box$

We define the theory $\mathsf{R}^-_{2,\mathcal{F}}$ as $\mathsf{R}^-_{1,\mathcal{F}}$ plus the following axioms:

($\mathcal{F}3$) $\vdash \bigwedge_{i<n}(h_i \to \Box h_i)$.

Let $\mathcal{K}$ be any Kripke model on the frame $\mathcal{F}$. We define an interpretation $\sigma^\star$ from $\mathcal{L}_{\mathsf{mod}}$ to the closed formulas of the language of $\mathsf{R}^-_{\mathcal{F}}$, by setting $\sigma^\star(p) := \bigvee_{j \Vdash p} \ell_j$, where $\sigma^\star$ commutes with the propositional connectives and $\Box$. We have the following.

Theorem 5.7 *We have, for every formula φ of the modal language, the following:*

- *if $i \Vdash \varphi$, then $\mathsf{R}^-_{2,\mathcal{F}} \vdash \ell_i \to \sigma^\star(\varphi)$;*
- *if $i \nVdash \varphi$, then $\mathsf{R}^-_{2,\mathcal{F}} \vdash \ell_i \to \neg\,\sigma^\star(\varphi)$.*

Proof The proof is by induction on φ. The cases of the atoms and of the propositional connectives are trivial using Lemma 5.3.

Suppose $\varphi = \Box\psi$.

Suppose $i \Vdash \Box\psi$. We reason in $\mathsf{R}^-_{2,\mathcal{F}}$. Suppose ℓ_i. Then, h_i and, hence, $\Box h_i$. By Lemma 5.5, in combination with $\Box\neg\,\ell_i$, we find $\Box \bigvee_{j \succ i} \ell_j$. So, by the induction hypothesis, we have $\Box\sigma^\star(\psi)$.

Suppose $i \nVdash \Box\psi$. Then, for some $j \succ i$, $j \nVdash \psi$. We reason in $\mathsf{R}^-_{2,\mathcal{F}}$. Suppose ℓ_i. It follows that $\Diamond\ell_j$. Ergo, by the induction hypothesis, $\Diamond\neg\,\sigma^\star(\psi)$. Hence, $\neg\,\Box\sigma^\star(\psi)$. $\Box$

Remark 5.8 The most naive attempt to avoid the use of $\vdash h_i \to \Box h_i$ is to replace $\Box\neg\,\ell_i$ by $\Box \bigvee_{j \succ i} \ell_j$ (where we read $\Diamond$ as $\neg\,\Box\neg$) everywhere in the above definitions and arguments. This certainly will give us the $(i \Vdash \Box\psi)$-part in the proof of Theorem 5.7 for free. It may amuse the reader to try this and to see where and why precisely it goes wrong.

5.4 Application to arithmetic In this subsection, we articulate what Theorem 5.7 tells us about a theory U with arithmetic N.

Consider a theory U and an arithmetic N in U. Suppose that $\deg(N) = \alpha$ and $U \vdash (\mathsf{T}^1_2)^N$. Suppose $\mathsf{GL}_\alpha \nvdash \varphi$. Let $\mathcal{K}$ be a finite counter Kripke model with frame $\mathcal{F}$. We choose $\mathcal{K}$ in such a way that the set of worlds is $\{0, \ldots, n-1\}$, the root is 0, and $0 \nVdash \varphi$. Note that the depth of the root must be $k \leq \alpha$.

Let τ be the N-interpretation of the language of $\mathsf{R}^-_{\mathcal{F}}$ that is generated by $\ell_i \mapsto L^N_i$. (We are only interested in τ on the closed fragment of $\mathsf{R}^-_{\mathcal{F}}$.) Clearly, $\tau(h_i)$ is of the form H^N_i, where H_i is S^1_2-provably equivalent to an $\exists\Pi^{\mathsf{b}}_1$-sentence. Let $\sigma^\star$ be the interpretation of the language of provability logic in the closed fragment of $\mathsf{R}^-_{2,\mathcal{F}}$ generated by $p \mapsto \bigvee_{i \Vdash p} \ell_i$. We take the interpretation ν of the language of provability logic into the language of U to be $\tau \circ \sigma^\star$. Thus, ν is the N-interpretation generated by $p \mapsto \bigvee_{i \Vdash p} L^N_i$. We assume that $U \vdash \bigwedge(H_i \to \Box_U H_i)$, in other words, that $\mathsf{R}^-_{2,\mathcal{F}}$ is N-valid. We show that $U \nvdash \nu(\varphi)$.

Suppose $U \vdash \nu(\varphi)$. Since $U \vdash L^N_0 \to \neg\,\nu(\varphi)$, we find that $U \vdash \neg\,L^N_0$. It follows that $U \vdash \Box^N_U \neg\,L^N_0$, and, hence $U \vdash \bigvee_{j<n} L^N_j$. Since $U \vdash \neg\,L^N_0$, we find $U \vdash \bigvee_{0<j<N} L^N_j$. Thus, since each $j > 0$ satisfies $\Box^{k-1}\bot$, we find $U \vdash \Box^{N,k-1}_U\bot$, quod non. We may conclude that $U \nvdash \nu(\varphi)$.

The following theorem is an immediate consequence of these considerations.

Theorem 5.9 *Consider a theory U and an arithmetic N in U. We suppose that $U \vdash (\mathsf{T}^1_2)^N$ and $U \vdash S^N \to \Box^N_U S^N$, for all sentences S in $\exists\Pi^{\mathsf{b}}_1$. Then* $\mathsf{prl}(N) = \mathsf{GL} + \Box^{\mathsf{deg}(N)}\bot$.

5.5 All arithmetics of a theory In this subsection, we apply the Solovay argument to all arithmetics of a theory U. We first show how we can improve our local arithmetic.

Consider any set of Σ_1-sentences $\mathcal{S}$ with n elements. Let $C_{\mathcal{S}} := \bigwedge_{S\in\mathcal{S}}(S \to \Box_{\mathsf{S}^1_2} S)$. Let J be an S^1_2-cut such that $\mathsf{S}^1_2 \vdash \forall x \in J\, \exists y\ 2^{2^x} = y$. We define $J_0 := \mathsf{ID}$, $J_{k+1} := \mathsf{ID}\langle C_{\mathcal{S}}\rangle(J \circ J_k)$, $J_{\mathcal{S}} := J_n$. The following argument is taken from Visser [33].

Lemma 5.10 *We have* $\mathsf{S}^1_2 \vdash C_{\mathcal{S}}^{J_{\mathcal{S}}}$.

Proof Reason in S^1_2. If we have $C_{\mathcal{S}}$, clearly $J_n = \mathsf{ID}$, and we are done. Otherwise, for some S in $\mathcal{S}$, we have S and $\neg\Box_{\mathsf{S}^1_2} S$. So, inside J, the sentence S will be false. It follows that, inside J, the number of true S's from $\mathcal{S}$ is at least one less than inside ID. Now the game repeats itself inside J for J_{n-1}. Each time we have $\neg C_{\mathcal{S}}$, we move inside J and loose at least one true S. If at some point, we have $C_{\mathcal{S}}$, we are done. Otherwise we end up with zero true S's and we have $C_{\mathcal{S}}$ in J_n. (Since n is standard the whole argument can be spelled out with big disjunctions, and so forth.) □

We have the following theorem.

Theorem 5.11 *Consider any theory U. We have* $\mathsf{prov}_{\mathsf{all}}(U) = \mathsf{GL} + \Box^{\mathsf{deg}(U)}\bot$.

We note that the result also is valid for the case when $\mathsf{deg}_{\mathsf{all}}(U) = 0$, that is, when either U is inconsistent or S^1_2 is not interpretable in U.

Proof Consider any theory U, and suppose $\mathsf{deg}_{\mathsf{all}}(U) = \alpha$. It is easily seen that $\mathsf{GL} + \Box^\alpha\bot \subseteq \mathsf{prl}(U)$.

Suppose $\mathsf{GL}_\alpha \nvdash \varphi$. Then, there is a finite Kripke model $\mathcal{K}$ with nodes $\{0, \ldots, n-1\}$ and with root 0, such that $0 \nVdash \varphi$ and $d(0) \le \alpha$.

Since $\mathsf{deg}_{\mathsf{all}}(U) := \sup(\{\mathsf{deg}(M) \mid M \text{ is an arithmetic in } U\})$ and $d(0)$ is finite, we can find an arithmetic N_0 with $d(0) \le \mathsf{deg}(N_0) \le \alpha$. We can shorten N_0 to an arithmetic $N_1 \preceq N_0$ in which we have T^1_2 (see, e.g., [15]; in fact, we can shorten N_0 to a cut on which we have $\mathrm{I}\Delta_0 + \Omega_1$). We note that $\mathsf{deg}(N_1) \ge \mathsf{deg}(N_0) \ge d(0)$.

We simultaneously construct a cut N in N_0 and the L_i using the Gödel fixed-point lemma. We find L_i such that

$$\mathsf{S}^1_2 \vdash \ L_i \leftrightarrow \Big(\Box_U \neg L_i^{N_1\circ J_{\mathcal{H}}} \wedge \bigwedge_{j\succ i} \Diamond L_j^{N_1\circ J_{\mathcal{H}}} \wedge \bigwedge_{j\parallel i}\ \bigvee_{k\preceq i,\, k\parallel j} \Box_U \neg L_k^{N_1\circ J_{\mathcal{H}}} < \Box_U \neg L_j^{N_1\circ J_{\mathcal{H}}}\Big).$$

Here

- $H_i := (\Box_U \neg L_i^{N_1\circ J_{\mathcal{H}}} \wedge \bigwedge_{j\parallel i}\ \bigvee_{k\preceq i,\, k\parallel j} \Box_U \neg L_k^{N_1\circ J_{\mathcal{H}}} < \Box_U \neg L_j^{N_1\circ J_{\mathcal{H}}})$,
- $\mathcal{H} := \{H_0, \ldots, H_{n-1}\}$.

We note that we have indeed a valid application of the fixed-point lemma since $\mathcal{H}$ occurs "inside the box."

We take $N := N_1 \circ J_{\mathcal{H}}$. We note that $\mathsf{S}^2_1 \vdash \Box_{\mathsf{S}^2_1} H_i \to \Box_U H_i^N$. Hence, we find that $U \vdash H_i^N \to \Box^N_U H_i^N$. Moreover, $U \vdash (\mathsf{T}^1_2)^N$, since T^1_2 is downwards preserved over $\preceq$. Finally, $\mathsf{deg}(N) \ge \mathsf{deg}(N_1) \ge d(0)$.

We now employ the interpretation ν of Section 5.4 using the L_i constructed above. We find $U \nvdash \nu(\varphi)$. $\square$

5.6 Theories with a Σ_1-sound arithmetic In this subsection we provide a sufficient condition for a theory to contain an arithmetic N with $\mathsf{prl}(N) = \mathsf{GL}$. We will use the following facts.

Fact 5.12 Suppose that N is an arithmetic in U. Then $U \rhd (U + \Box_U^N \bot)$.

This insight was first due Solomon Feferman in his classical paper [10]. The simple proof below was discovered independently by Per Lindström (see [11]) and the author [31].

Proof Suppose that N is an arithmetic in U. We have $U + \neg\Box_U^N \bot \vdash \Diamond_U^N(\Box_U^N \bot)$, by Löb's theorem. Hence, it follows that $(U + \neg\Box_U^N \bot) \rhd (U + \Diamond_U^N \Box_U^N \bot)$. So, using a Henkin interpretation, we may conclude that $(U + \neg\Box_U^N \bot) \rhd (U + \Box_U^N \bot)$. On the other hand, we trivially have $(U + \Box_U^N \bot) \rhd (U + \Box_U^N \bot)$. Thus, using a disjunctive interpretation, we find that $U \rhd (U + \Box_U^N \bot)$. $\square$

Fact 5.13 Suppose $U \rhd V$. Let N be an arithmetic in V. Then $U \rhd (V + \Box_U^N \bot)$.

Proof Suppose $M : U \rhd V$. We apply Fact 5.12 to the arithmetic $M \circ N$ to find the desired result. $\square$

We note that Facts 5.12 and 5.13 can be considered as nice and general formulations of the second incompleteness theorem. Suppose that, for some arithmetic N in U, we would have $U \vdash \mathsf{con}^N(U)$. Since, by Fact 5.12, we have $U \rhd (U + \Box_U^N \bot)$, it follows that U is inconsistent.

Fact 5.14 Suppose that $U \rhd V$, and suppose that U contains a Σ_1-sound arithmetic N; that is, for all Σ_1-sentences S, if $U \vdash S^N$, then S is true. Then $U \rhd_{\mathsf{faith}} V$.

This fact is a direct consequence of Theorem B.4 of Appendix B. It was first proved in Visser [36]. The basic idea of the proof is due to Per Lindström. We prove the following theorem.

Theorem 5.15 *Suppose that U contains a Σ_1-sound arithmetic N_0. Then there is an N in U such that* $\mathsf{prl}(N) = \mathsf{GL}$.

Proof Suppose that U contains a Σ_1-sound arithmetic N_0. We can find an interpretation of T^1_2 by shortening N_0. By Fact 5.13, we find $U \rhd (\mathsf{T}^1_2 + \Box_U \bot)$. Let W be T^1_2 plus sentential Σ_1-completeness for U with respect to each arithmetic M in U. Since $\mathsf{T}^1_2 + \Box_U \bot$ extends W, we find $U \rhd W$. By Fact 5.14, we can find a K such that $K : U \rhd_{\mathsf{faith}} W$. Finally, let $N := \mathsf{S}^1_2 \downarrow K$. Since W is a true theory, N is a Σ^0_1-sound arithmetic in U. Hence $\mathsf{deg}(N) = \infty$. By Theorem 5.9 we find that $\mathsf{prl}(N) = \mathsf{GL}$. $\square$

6 Deep Arithmetics

In this section we study the fine structure of the arithmetics of a finitely axiomatized sequential theory. Finitely axiomatized sequential theories have many surprising properties. The present section builds on and extends a line earlier work, to wit: Smoryński [27], Pudlák [25], Krajíček [19], and Visser [35], [36].

We have the following definition. Suppose that A is a finitely axiomatized sequential theory. (We confuse these theories with their single axiom.) Let N be an arithmetic in A.

- N is Σ_1*-veracious in* A if and only if
$$\mathsf{S}^1_2 \vdash \forall S \in \Sigma_1\text{-}\mathsf{sent}\,\big(\Box_A S^N \to \Box_{\mathsf{S}^1_2}\big(\mathsf{con}_{\rho(A)}(A) \to S\big)\big).$$
Thus, we see that Σ_1-veracity is the S^1_2-verifiable Σ_1-conservativity of N over $\mathcal{E}_{\mathsf{S}^1_2(\mathsf{S}^1_2+\mathsf{con}_{\rho(A)}(A))}$:
- N is *strong in* A if and only if $A \vdash \mathsf{con}^N_{\rho(A)}(A)$;
- N is *deep in* A if and only if N is both Σ_1-veracious and strong in A.

Σ_1-veracity is connected to Σ_1-soundness: this is elucidated by the following theorem.

Theorem 6.1 *Suppose that A is a finitely axiomatized sequential theory and N is Σ_1-veracious in A. Then,*
$$\mathrm{I}\Delta_0 + \mathsf{supexp} + \mathsf{con}(A) \vdash \forall S \in \Sigma_1\text{-}\mathsf{sent}\,\big(\Box_A S^N \to \mathsf{true}(S)\big).$$
Here true *is a* Σ_1*-truth predicate.*

Proof By [42], the theory EA, also known as $\mathrm{I}\Delta_0 + \mathsf{exp}$, proves uniform Π_2-reflection for cut-free provability in S^1_2. Hence, $\mathrm{I}\Delta_0 + \mathsf{supexp}$ proves uniform Π_2-reflection for ordinary provability in S^1_2. Our theorem is immediate from this. $\Box$

In the definition of Σ_1-*veracious theory* we may replace $\rho(A)$ by any $m \geq \rho(A)$, in the light of the following theorem.

Theorem 6.2 *Let A be a finitely axiomatized sequential theory. Suppose that $m \geq \rho(A)$. We have*
$$\mathsf{S}^1_2 \vdash \forall S \in \Sigma_1\text{-}\mathsf{sent}\,\big(\Box_{\mathsf{S}^1_2}\big(\mathsf{con}_m(A) \to S\big) \leftrightarrow \Box_{\mathsf{S}^1_2}\big(\mathsf{con}_{\rho(A)}(A) \to S\big)\big).$$

Proof The right-to-left direction of our theorem is trivial.

To go from m-provability to $\rho(A)$-provability we need to eliminate standard (proof-theoretical) cuts. So we only need a multiexponential transformation.[11] Thus, there is an S^1_2-cut J, such that $\mathsf{S}^1_2 \vdash \mathsf{con}_{\rho(A)}(A) \to \mathsf{con}^J_m(A)$.

Reason in S^1_2. Consider any Σ_1-sentence S. Suppose that $\Box_{\mathsf{S}^1_2}(\mathsf{con}_m(A) \to S)$. So, $\Box_{\mathsf{S}^1_2}(\mathsf{con}_m(A) \to S)^J$. It follows that $\Box_{\mathsf{S}^1_2}(\mathsf{con}^J_m(A) \to S^J)$. We have $\Box_{\mathsf{S}^1_2}(S^J \to S)$, since we can find a p-time transformation of S into a proof of $(S^J \to S)$ by the obvious recursion on the construction of S. (Note that the recursion is over Σ_1-formulas rather than sentences.) Hence, $\Box_{\mathsf{S}^1_2}(\mathsf{con}^J_m(A) \to S)$. Thus, we may conclude that $\Box_{\mathsf{S}^1_2}(\mathsf{con}_{\rho(A)}(A) \to S)$. $\Box$

To illustrate the theorem, let us briefly consider the case where $S := \bot$. We have
$$\mathsf{S}^1_2 \vdash \Box_{\mathsf{S}^1_2}\mathsf{incon}_m(A) \leftrightarrow \Box_{\mathsf{S}^1_2}\mathsf{incon}_{\rho(A)}(A).$$
We note that the "guarding" boxes are essential. We definitely do *not* generally have that S^1_2 proves the equivalence $\mathsf{incon}_m(A)$ and $\mathsf{incon}_{\rho(A)}(A)$, which would involve a multiexponential transformation.

If an arithmetic is deep, we can strengthen the implication in Σ_1-veracity to a bi-implication.

Theorem 6.3 *Let A be a finitely axiomatized sequential theory. Suppose that N is a deep arithmetic in A. We have*

$$\mathsf{S}^1_2 \vdash \forall S \in \Sigma_1\text{-}\mathsf{sent}\,\big(\Box_A S^N \leftrightarrow \Box_{\mathsf{S}^1_2}\big(\mathsf{con}_{\rho(A)}(A) \to S\big)\big).$$

We leave the simple proof to the reader.

Theorem 6.4 *Let A be a finitely axiomatized sequential theory. Each of the following classes is S^1_2-verifiably downwards closed under $\preceq$: the Σ_1-veracious theories, the strong theories, and the deep theories.*

We leave the simple proof to the reader. Iterated inconsistencies take a simple form for Σ_1-veracious arithmetics, as will be proved in the next theorem.

Theorem 6.5 *Suppose that A is sequential and that N is a Σ_1-veracious arithmetic in A. We have*

$$(\dagger_n) \quad \mathsf{S}^1_2 \vdash \Box_A \Box_A^{N,n} \bot \leftrightarrow \Box^n_{\mathsf{S}^1_2} \Box_A \bot.$$

Proof We prove our theorem by induction on n. The case $n = 0$ is trivial. Suppose that we have $(\dagger_n)$. Note that it follows that

$$(\dagger^*_n) \quad \mathsf{S}^1_2 \vdash \Box_{\mathsf{S}^1_2} \Box_A \Box_A^{N,n} \bot \leftrightarrow \Box_{\mathsf{S}^1_2} \Box^n_{\mathsf{S}^1_2} \Box_A \bot.$$

We prove $(\dagger_{n+1})$. We have in S^1_2,

$$\begin{aligned}
\Box_A \Box_A^{N,n+1} \bot &\to \Box_{\mathsf{S}^1_2}\big(\mathsf{con}_{\rho(A)}(A) \to \Box_A \Box_A^{N,n} \bot\big) && (5)\\
&\to \Box_{\mathsf{S}^1_2}(\Box_{A,\rho(A)} \bot \vee \Box_A \Box_A^{N,n} \bot) && (6)\\
&\to \Box_{\mathsf{S}^1_2} \Box_A \Box_A^{N,n} \bot && (7)\\
&\to \Box_{\mathsf{S}^1_2} \Box^n_{\mathsf{S}^1_2} \Box_A \bot && (8)\\
&\to \Box^{n+1}_{\mathsf{S}^1_2} \Box_A \bot && (9)\\
&\to \Box_{\mathsf{S}^1_2} \Box^n_{\mathsf{S}^1_2} \Box_A \bot && (10)\\
&\to \Box_{\mathsf{S}^1_2} \Box_A \Box_A^{N,n} \bot && (11)\\
&\to \Box_A \Box_A^N \Box_A^{N,n} \bot && (12)\\
&\to \Box_A \Box_A^{N,n+1} \bot. && (13)
\end{aligned}$$

Here step (5) follows by Σ_1-veracity. Steps (8) and (11) use $(\dagger^*_n)$. Finally, Step (12) uses the fact that we have S^1_2 inside N. □

Corollary 6.6 *Suppose that A is a finitely axiomatized sequential theory and that N is a Σ_1-veracious arithmetic in A. We have*

(i) $A \vdash \Box_A^{N,n+1} \bot \leftrightarrow (\Box^n_{\mathsf{S}^1_2} \Box_A \bot)^N$;

(ii) $\mathrm{I}\Delta_0 + \mathsf{supexp} \vdash \Box_A \Box_A^{N,n} \bot \leftrightarrow \Box_A \bot$.

In the next theorem, we establish the existence of lots of deep arithmetics in a finitely axiomatized sequential theory. The proof of the theorem employs a form of the Friedman–Goldfarb–Harrington fixed point. See [36] for a discussion of this fixed point.

Theorem 6.7 *For every finitely axiomatized sequential theory A, and, for every arithmetic N_0 in A, there is a deep arithmetic N in A with $N \preceq N_0$. This theorem is verifiable in* S^1_2.

The arithmetic N in the theorem is dependent on A and N_0. Suppose that we have a deep arithmetic N for A. Of course, we can extend A to $B := (A + \mathsf{incon}^N(A))$. Consider the arithmetic $N' : \mathsf{S}^1_2 \to B$ that is based on the same translation as N, that is, τ_N. Clearly, N' will not be deep. However, the theorem tells us that we can shorten N' in such a way that we obtain a deep arithmetic for B.

Proof Let A be a finitely axiomatized sequential theory, and let N_0 be an arithmetic in A.

Let true be a Σ_1-truth predicate. For the construction of such a truth predicate, see [15, Chapter V(5)]. The two classical works on this subject are [20] and [23]. We will use the following two properties of true: for S in Σ_1,

(T1) $\mathsf{S}^1_2 \vdash \mathsf{true}(S) \to S$.

(T2) Suppose that J is an S^1_2-cut such that $\mathsf{S}^1_2 \vdash x \in J \to 2^{2^x}$ exists; then, we have $\mathsf{S}^1_2 \vdash S^J \to \mathsf{true}(S)$.

We remind the reader of the witness comparison ordering. We define, for any $C = \exists x\, C_0(x)$ and $D = \exists y\, D_0(y)$,

- $C \leq D := \exists x\, (C_0(x) \wedge \forall y < x\, \neg D_0(y))$,
- $C < D := \exists x\, (C_0(x) \wedge \forall y \leq x\, \neg D_0(y))$,
- $(C \leq D)^\perp := (D < C)$, and $(C < D)^\perp := (D \leq C)$.

By the Gödel fixed-point lemma, we find R such that, for a suitably chosen m:

$$\mathsf{S}^1_2 \vdash R \leftrightarrow \mathsf{true}(S) \leq \Box_{A,m} R^{N_0}.$$

We note that the complexity $\rho(R)$ of R is *not* dependent on S and m, since numerals do not change the complexity of a formula even if the numeral is given a relational representation. Moreover, for any B and K, $\rho(B^K)$ is linear in $\rho(B)$. Hence, we may choose m so large that $\mathsf{max}(\rho(A), \rho(R^{N_0})) \leq m$.

We choose N_1 to be an initial segment of N_0 such that

(U1) $A \vdash \Box^{N_1}_{A,m} B \to B$, for any B with $\rho(B) \leq m$;

(U2) $A \vdash (\forall S \in \Sigma_1\text{-}\mathsf{sent}\, (\mathsf{true}(S) \to \mathsf{true}(S) \leq \mathsf{true}(S)))^{N_1}$;
in other words, A proves that, in N_1, if $\mathsf{true}(S)$ is witnessed, then $\mathsf{true}(S)$ has a minimal witness.

We can always find such an N_1 since (i) we have a truth predicate for formulas of complexity $\leq m$ and since (ii) we can interpret $\mathrm{I}\Delta_0 + \Omega_1$ in S^1_2.

Let J be an S^1_2-cut such that $\mathsf{S}^1_2 \vdash x \in J \to 2^{2^x}$ exists. We take $N := N_1 \circ J$.

We note that N_1 is strong and that, hence, N is strong. We show that N is Σ_1-veracious.

We reason, for the rest of the proof, in S^1_2. Consider any Σ_1-sentence S. Suppose $\Box_A S^N$. It follows, by (T2) that $\Box_A(\mathsf{true}(S))^{N_1}$. Ergo, by (U2), $\Box_A(R \vee R^\perp)^{N_1}$. Thus, $\Box_A(R \vee \Box_{A,m} R^{N_0})^{N_1}$. Hence, by (U1), $\Box_A(R^{N_1} \vee R^{N_0})$. Since N_1 is a cut of N_0, we get $\Box_A R^{N_0}$.

By Lemma 3.7, we may conclude that (†) $\Box_{\mathsf{S}^1_2}\Box_{A,m} R^{N_0}$.

We can find an S^1_2-cut $J^\star$, on which we have T^1_2, so that if something is A-provable with a proof in $J^\star$, then there is a minimal proof. We can arrange that $J^\star$

is so small that $\mathsf{S}^1_2 \vdash x \in J^\star \to 2^{2^x}$ exists. We find from (†): $\Box_{\mathsf{S}^1_2}(\Box_{A,m} R^{N_0})^{J^\star}$. Hence, $\Box_{\mathsf{S}^1_2}(R \vee R^\perp)^{J^\star}$. We may conclude that

$$\Box_{\mathsf{S}^1_2}\big(\mathsf{true}(S) \vee (R^\perp \wedge \Box_{A,m} R^{N_0})\big)^{J^\star}.$$

Since we have Σ_1-completeness in the presence of double exponentiation, it follows that

$$\Box_{\mathsf{S}^1_2}\big(\mathsf{true}(S) \vee \Box_{A,m}(R^\perp \wedge R)^{N_0}\big).$$

Hence, $\Box_{\mathsf{S}^1_2}(S \vee \Box_{A,m}\bot)$, or, in a different formulation: $\Box_{\mathsf{S}^1_2}(\mathsf{con}_m(A) \to S)$.

By Theorem 6.2, we may conclude that $\Box_{\mathsf{S}^1_2}(\mathsf{con}_{\rho(A)}(A) \to S)$.

We have proved that, for every finitely axiomatized sequential theory A, and, for every arithmetic N_0 in A, there is a deep arithmetic N in A with $N \preceq N_0$. To see that this argument is verifiable in S^1_2, we have to see that the construction of N from N_0 is feasible. We note that m in our argument remains standard even if S is nonstandard. As a consequence, for example, $\delta_N = \varphi(\delta_{N_0}, \mathsf{Z}_{N_0}, \mathsf{S}_{N_0}, \ldots)$, where φ is a fixed standard context. Thus N will be p-time in N_0. $\Box$

Discussion 6.8 Clearly, the second incompleteness theorem implies that adding $\mathsf{con}(U)$ to a consistent theory U that contains an arithmetic gives us a stronger theory, a theory that is, so to say, *one gödel* stronger. However, it is clear that we have to ask: to what arithmetic in U are we adding the consistency statement?

Consider GB, and let neumann be the interpretation of S^1_2 in the finite von Neumann ordinals. Clearly,

$$\mathsf{PA} \nvdash \mathsf{con}(\mathsf{GB}) \to \mathsf{con}\big(\mathsf{GB} + \mathsf{con}^{\mathsf{neumann}}(\mathsf{GB})\big).$$

In fact, by the second incompleteness theorem, GB cannot prove this statement with respect to the neumann-interpretation. However, for a Σ_1-veracious arithmetic N in GB, we have

$$\mathrm{I}\Delta_0 + \mathsf{supexp} \vdash \mathsf{con}(\mathsf{GB}) \to \mathsf{con}\big(\mathsf{GB} + \mathsf{con}^N(\mathsf{GB})\big).$$

Thus, in which theories the relative consistency of a theory plus its consistency statement with respect to that theory can be verified is dependent on the chosen arithmetic. Adding $\mathsf{con}^N(\mathsf{GB})$ adds less strength to GB than adding $\mathsf{con}^{\mathsf{neumann}}(\mathsf{GB})$ does.

So the *gödel* is not such a good unit when we define it as *how much stronger a theory becomes when we add its consistency statement*. My proposal would be to take as the theory that is *one gödel* stronger: $\mathsf{S}^1_2 + \mathsf{con}(U)$. Note that the strength of $\mathsf{S}^1_2 + \mathsf{con}(U)$ still depends on the chosen axiomatization of U.

In the case of a finitely axiomatized sequential theory A and a Σ_1-veracious arithmetic N in A, we have

$$\mathsf{S}^1_2 \vdash \mathsf{con}\big(A + \mathsf{con}^N(A)\big) \leftrightarrow \mathsf{con}\big(\mathsf{S}^1_2 + \mathsf{con}(A)\big).$$

So, by the measure of S^1_2-verifiable relative consistency, adding the consistency statement for a Σ_1-veracious arithmetic in A is adding one gödel. Note that there are no arithmetics in, say, PA with the same property.

The next theorem shows that under a verifiability condition, Theorem 6.7 can be strengthened to theories that are mutually interpretable with a finitely axiomatized sequential theory.

Theorem 6.9 *Suppose that A is a consistent finitely axiomatized sequential theory and that U is any theory. Suppose $K : A \rhd U$ and $M : U \rhd A$. Then, there is an arithmetic N in U, such that*

$$\mathsf{S}^1_2 \vdash K : A \rhd_{\mathsf{thm}} U \to \forall S \in \Sigma_1\text{-}\mathsf{sent}\,\big(\Box_U S^N \to \Box_{\mathsf{S}^1_2}(\mathsf{con}_{\rho(A)}(A) \to S)\big).$$

Here $\rhd_{\mathsf{thm}}$ stands for theorems-interpretability, where we demand that the interpreting theory prove the translations of the theorems of the interpreted theory. In the context of arithmetics without Σ_1-collection this notion is not provably equivalent to the usual notion of axioms-interpretability.

Proof Suppose $K : A \rhd U$ and $M : U \rhd A$. We find $\Box_{\mathsf{S}^1_2}(M : U \rhd A)$. Consider any arithmetic N_0 in A. We note that $N_1 := N_0MK$ is also an arithmetic in A.[12] Let N_2 be an arithmetic in A such that $N_2 \preceq N_0$ and $N_2 \preceq N_1$. We may assume that in N_2 we have $\mathrm{I}\Delta_0 + \Omega_1$, or a sufficiently large, finitely axiomatized part of $\mathrm{I}\Delta_0 + \Omega_1$.

Let k be the complexity of $(\mathsf{true}(x))^{N_0}$, where true is the Σ_1-truth predicate. By Theorem 3.8, we can choose m so large that, S^1_2-verifiably,

$$\forall B \in \Gamma_k\,\big(\Box_A(\Box^{N_2}_{A,m} B \to B) \to \Box_{A,m} B\big). \tag{14}$$

Let $N_3 \leq N_0$ be an arithmetic in A such that, verifiably in S^1_2,

$$\forall B \in \Gamma_m\, \Box_A(\Box^{N_3}_{A,m} B \to B). \tag{15}$$

Let N_4 be a cut of N_3 such that

$$\Box_A \forall x \in N_4\, \exists y \in N_3\, 2^{2^x} = y. \tag{16}$$

Finally, we take $N := N_4M$. So, N is an arithmetic in U. By the Gödel fixed-point lemma, we find R such that

$$\Box_{\mathsf{S}^1_2}\big(R \leftrightarrow \mathsf{true}(S) \leq \Box_{A,m} R^{N_0}\big). \tag{17}$$

We reason in S^1_2. We have, for all Σ_1-sentences S,

$$\begin{aligned}
\Box_A\ \big(S^{N_4} &\to \big(\mathsf{true}(S)\big)^{N_3} && (18)\\
&\to (R \vee R^{\perp})^{N_3} && (19)\\
&\to (R^{N_0} \vee \Box^{N_3}_{A,m} R^{N_0}) && (20)\\
&\to R^{N_0}\ \big). && (21)
\end{aligned}$$

Suppose $K : A \rhd_{\mathsf{thm}} U$. We also have $M : U \rhd A$. Consider any S, and suppose $\Box_U S^N$. It follows that $\Box_A S^{N_4MK}$. By the previous result, we may conclude that $\Box_A R^{N_0MK}$, that is, $\Box_A R^{N_1}$. From this, we have

$$\begin{aligned}
\Box_A\ (\Box^{N_2}_{A,m} R^{N_0} &\to R^{N_1} \wedge \Box^{N_2}_{A,m} R^{N_0} && (22)\\
&\to R^{N_2} && (23)\\
&\to R^{N_0}). && (24)
\end{aligned}$$

So, by equation (14), we have $\Box_A R^{N_0}$. It follows that $\Box_{\mathsf{S}^1_2}\Box_{A,m} R^{N_0}$. We now may repeat the reasoning of the proof of Theorem 6.7. So we get

$$\Box_{\mathsf{S}^1_2}\big(\mathsf{con}_{\rho(A)}(A) \to S\big),$$

and we are done. $\Box$

7 An Example

In this section, we provide an example of a sequential theory W, such that the degrees of the arithmetics in W are finite and cofinal in ω. So, for every n there is an arithmetic in W with degree $k \geq n$, but there is no arithmetic in W with degree ∞.

We start with a consistent finitely axiomatized sequential theory A. Pick any arithmetic N in A. Let Σ be the signature of A, and let Θ be the signature of arithmetic.

Let $\tau : \Theta \to \Sigma$ be a translation. We define $\tilde{\tau} := \langle \mathsf{S}^1_2, \tau\langle(\mathsf{S}^1_2)^\tau\rangle\tau_N, A\rangle$. Here we assume that the axioms of identity are an explicit part of the axiomatization of S^1_2. It is easily seen that $\tilde{\tau}$ is an arithmetic in A. We assign to any translation $\tau : \Theta \to \Sigma$ a Gödel number $\mathsf{gn}(\tau)$. We define

$$W := A + \big\{(\Box^{\mathsf{gn}(\tau)}_{\mathsf{S}^1_2}\Box_A\bot)^{\tilde{\tau}} \mid \tau : \Theta \to \Sigma\big\}.^{13}$$

We note that there is a p-time algorithm to decide whether a sentence is of the form $(\Box^{\mathsf{gn}(\tau)}_{\mathsf{S}^1_2}\Box_A\bot)^{\tilde{\tau}}$. So, W is Δ^{b}_1-axiomatized.[14]

Consider any arithmetic K in W. We have $W \vdash (\Box^{\mathsf{gn}(\tau_K)}_{\mathsf{S}^1_2}\Box_A\bot)^{\tilde{\tau}_K}$. Clearly, in W, the interpretations K and $\tilde{\tau}_K \uparrow W$ coincide; hence, by an easy induction, $W \vdash \Box^{K,\mathsf{gn}(\tau_K)+1}_W\bot$. So, each arithmetic K in W has a finite degree.

We show that, for any n, the theory W contains an arithmetic with degree $\geq n$. Consider any number n.

Let $N^\star$ be a strong arithmetic in A that has an initial embedding in all arithmetics $\tilde{\tau}$ in A with $\mathsf{gn}(\tau) \leq n$. Let $N^\circ \preceq (N^\star \uparrow (A + \Box^{N^\star}_A\bot))$ be a deep arithmetic in $A + \Box^{N^\star}_A\bot$. Let $N_\circ := \tilde{\tau}_{N^\circ}$. We note that $N_\circ$ is an arithmetic in A.

We want to show that $W \nvdash \Box^{N_\circ\uparrow W,n}_W\bot$. This will be a direct consequence of the following claim.

Claim *We have* $\mathsf{S}^1_2 \vdash \Box_W\Box^{N_\circ\uparrow W,n}_W\bot \to \Box^n_{\mathsf{S}^1_2}\Box_A\bot$.

We first show how our desired result follows from the claim. Suppose $W \vdash \Box^{N_\circ\uparrow W,n}_W\bot$. Since S^1_2 is a true theory, the claim gives us, by applying reflection a number of times, $\Box_A\bot$. Quod non. Note that this argument can be formalized in $\mathrm{I}\Delta_0 + \mathsf{supexp}$.

Proof of the claim By our conventions, we may write $\Box^{N_\circ\uparrow W,n}_W\bot$ as $\Box^{N_\circ,n}_W\bot$. We will apply this convention to increase readability. We prove by induction that, for each $j \leq n$,

$$(\$_j) \quad \mathsf{S}^1_2 \vdash \Box_W\Box^{N_\circ,j}_W\bot \to \Box^j_{\mathsf{S}^1_2}\Box_A\bot.$$

For the case $j = 0$, we have to prove $\mathsf{S}^1_2 \vdash \Box_W\bot \to \Box_A\bot$. *We reason in* S^1_2. Suppose $\Box_W\bot$. Consider any W-proof p of $\bot$. If p only employs the axiom A, we are done. Suppose that p employs at least one axiom of the form $(\Box^{\mathsf{gn}(\tau)}_{\mathsf{S}^1_2}\Box_A\bot)^{\tilde{\tau}}$. Let X be the set of such axioms employed in p. By our assumption X is not empty.

We construct an arithmetic M in A, such that $M \preceq \tilde{\tau}$, for all $\tau \in X$. Suppose $\tau_0 \in X$. For each τ we construct an initial embedding F_τ of an initial segment J_τ of $\tilde{\tau}_0$ in $\tilde{\tau}$. This construction is uniform in τ, and $|F_\tau|$ is linear in $|\tau|$. We take M to be the intersection of the J_τ.

The definition of M involves, for example, a conjunction $\delta_M(x) :\leftrightarrow \bigwedge_{\tau\in X} J_\tau(x)$. Why is this conjunction not too big? Under reasonable assumptions, we have

$$\mathsf{gn}(\tau) \le \mathsf{gn}\big((\Box^{\mathsf{gn}(\tau)}_{\mathsf{S}^1_2}\Box_A\bot)^{\tilde\tau}\big) \le p.$$

Moreover, $|J_\tau|$ of is linear in $|\mathsf{gn}(\tau)|$ and, hence, linear in $|p|$. Moreover the size of X is $< |p|$. Hence $|M|$ is bounded by $L(|p|) \times |p|$, where L is a standard linear term. So M is below $a \cdot (\omega_1(p))^b$, for some standard a and b.

Clearly, each axiom in X is implied by $\Box^M_A\bot$. So, we have $\Box_{A+\Box^M_A\bot}\bot$ and, hence $\Box_A\bot$, by the second incompleteness theorem.

An alternative way to prove this is to adapt the proof of the second incompleteness theorem as follows. We still reason inside S^1_2. By the Gödel fixed-point lemma, we construct G such that

$$\Box_{\mathsf{S}^2_1}\Big(G \leftrightarrow \neg\,\Box_A \bigvee_{\tau\in X} G^{\tilde\tau}\Big).$$

Note that the big disjunction exists inside S^1_2, since the set X is derived from p. We have

$$\begin{aligned}
\Box_{\mathsf{S}^1_2}\Big(\ \neg G &\to \Box_A \bigvee_{\tau\in X} G^{\tilde\tau} && (25)\\
&\to \Big(\Box_A \bigvee_{\tau\in X} G^{\tilde\tau} \wedge \Box_A \bigwedge_{\nu\in X} \Box^{\tilde\nu}_A \bigvee_{\tau\in X} G^{\tilde\tau}\Big) && (26)\\
&\to \Big(\Box_A \bigvee_{\tau\in X} G^{\tilde\tau} \wedge \Box_A \bigwedge_{\nu\in X} \neg\, G^{\tilde\nu}\Big) && (27)\\
&\to \Box_A\bot\ \Big). && (28)
\end{aligned}$$

Step (26) uses the fact that we have $\exists\Sigma^{\mathsf{b}}_1$-completeness for every $\tilde\tau$.

From our assumption on X, it clearly follows that $\Box_{A+\bigwedge_{\tau\in X}\Box^{\tilde\tau}_A\bot}\bot$. Hence, we find $\Box_A \bigvee_{\tau\in X} \mathsf{con}^{\tilde\tau}(A)$, and, so, by (28), $\Box_A \bigvee_{\tau\in X} G^{\tilde\tau}$. We may conclude that $\Box_A\bot$.

The nice feature of this second argument is that it does not use sequentiality.

We stop reasoning in S^1_2.

We now prove $(\$_{j+1})$, for $j+1 \le n$, where we use the induction hypothesis $(\$_j)$.

We reason again in S^1_2. Suppose $\Box_W\Box^{N_\circ, j+1}_W\bot$. Our induction hypothesis, $\$_j$, gives us $\Box_W(\Box^j_{\mathsf{S}^1_2}\Box_A\bot)^{N_\circ}$. Let p be a proof witnessing $\Box_W(\Box^j_{\mathsf{S}^1_2}\Box_A\bot)^{N_\circ}$. Let X be the set of all τ such that $(\Box^{\mathsf{gn}(\tau)}_{\mathsf{S}^1_2}\Box_A\bot)^{\tilde\tau}$ occurs as an axiom in p. We clearly have

$$\mathsf{proof}_{A+\{(\Box^{\mathsf{gn}(\tau)}_{\mathsf{S}^1_2}\Box_A\bot)^{\tilde\tau} \mid \tau\in X\}}\big(p, (\Box^j_{\mathsf{S}^1_2}\Box_A\bot)^{N_\circ}\big). \tag{29}$$

Let X_0 be the set of elements τ of X with $\mathsf{gn}(\tau) \le j$, and let X_1 be the set of elements of X with $\mathsf{gn}(\tau) > j$. Since $j+1 \le n$, we find that $N^\star \preceq \tilde\tau$, for any τ with $\mathsf{gn}(\tau) \le j$. It follows that, inside $\Box_{\mathsf{S}^1_2}$, $\Box^{N^\star}_A\bot$ implies $(\Box^{\mathsf{gn}(\tau)}_{\mathsf{S}^1_2}\Box_A\bot)^{\tilde\tau}$, for each τ in X_0.

Reasoning as in the case $j = 0$, we can find an arithmetic $M^\star$ in A, such that $M^\star \preceq \tilde\tau$, for all $\tau \in X_1$. Moreover, we can choose $M^\star$ in such a way that it is deep.

We have to move in a careful way, at this point, to compensate for our lack of induction. Clearly, we can find a standard cut J such that

$$\Box_{S^1_2} \forall z \in J\, (\Box^{j}_{S^1_2} \Box_A \bot \to \Box^{j+z}_{S^1_2} \Box_A \bot). \tag{30}$$

If follows that for p-time computable f with standard code:

$$\forall \tau \in X_1, \quad \exists q < f\big(\mathsf{gn}(\tau)\big)\ \mathsf{proof}_{S^1_2}\big(q, (\Box^{j}_{S^1_2} \Box_A \bot \to \Box^{\mathsf{gn}(\tau)}_{S^1_2} \Box_A \bot)\big). \tag{31}$$

We note that we can find a p-time computable g with standard code such that

$$\forall \tau \in X_1, \quad \exists r < g\big(\mathsf{gn}(\tau)\big)\ \mathsf{proof}_A\big(r, (\Box^{M^\star, \mathsf{gn}(\tau)}_{S^1_2} \Box_A \bot \to \Box^{\tilde{\tau}, \mathsf{gn}(\tau)}_{S^1_2} \Box_A \bot)\big). \tag{32}$$

Using equations (31) and (32), we can transform the proof p of equation (29) to a proof s witnessing the following provability:

$$\Box_{A + \Box_A^{N^\star} \bot + (\Box^{j+1}_{S^1_2} \Box_A \bot)^{M^\star}} (\Box^{j}_{S^1_2} \Box_A \bot)^{N_\circ}. \tag{33}$$

Hence, using $\boxplus B$ for $B \wedge \Box B$,

$$\boxplus_{S^1_2} \big(\Box_A (\Box^{j+1}_{S^1_2} \Box_A \bot)^{M^\star} \to \Box_{A + \Box_A^{N^\star} \bot} (\Box^{j}_{S^1_2} \Box_A \bot)^{N_\circ}\big). \tag{34}$$

Since $N_\circ \uparrow (A + \Box_A^{N^\star} \bot) = N^\circ$, we have

$$\boxplus_{S^1_2} \big(\Box_A (\Box^{j+1}_{S^1_2} \Box_A \bot)^{M^\star} \to \Box_{A + \Box_A^{N^\star} \bot} (\Box^{j}_{S^1_2} \Box_A \bot)^{N^\circ}\big). \tag{35}$$

Since $M^\star$ is deep, we find

$$\boxplus_{S^1_2} \big(\Box_{S^1_2} \big(\mathsf{con}_{\rho(A)}(A) \to \Box^{j+1}_{S^1_2} \Box_A \bot\big) \to \Box_{A + \Box_A^{N^\star} \bot} (\Box^{j}_{S^1_2} \Box_A \bot)^{N^\circ}\big). \tag{36}$$

Hence,

$$\boxplus_{S^1_2} \big(\Box^{j+2}_{S^1_2} \Box_A \bot \to \Box_{A + \Box_A^{N^\star} \bot} (\Box^{j}_{S^1_2} \Box_A \bot)^{N^\circ}\big). \tag{37}$$

Since N° is deep, we get

$$\boxplus_{S^1_2} \big(\Box^{j+2}_{S^1_2} \Box_A \bot \to \Box_{S^1_2} \big(\mathsf{con}_{\rho(A + \Box_A^{N^\star} \bot)}(A + \Box_A^{N^\star} \bot) \to \Box^{j}_{S^1_2} \Box_A \bot\big)\big). \tag{38}$$

It follows that

$$\boxplus_{S^1_2} \big(\Box^{j+2}_{S^1_2} \Box_A \bot \to \Box_{S^1_2} (\Box_A \neg \Box_A^{N^\star} \bot \vee \Box^{j}_{S^1_2} \Box_A \bot)\big). \tag{39}$$

By the second incompleteness theorem, we have

$$\boxplus_{S^1_2} \big(\Box^{j+2}_{S^1_2} \Box_A \bot \to \Box_{S^1_2} (\Box_A \bot \vee \Box^{j}_{S^1_2} \Box_A \bot)\big). \tag{40}$$

Ergo, we have

$$\boxplus_{S^1_2} (\Box^{j+2}_{S^1_2} \Box_A \bot \to \Box^{j+1}_{S^1_2} \Box_A \bot). \tag{41}$$

So, by Löb's theorem,

$$(\Box^{j+2}_{S^1_2} \Box_A \bot \to \Box^{j+1}_{S^1_2} \Box_A \bot) \wedge \Box^{j+2}_{S^1_2} \Box_A \bot. \tag{42}$$

We may conclude the following:

$$\Box^{j+1}_{S^1_2} \Box_A \bot. \tag{43}$$

This is what we wanted to prove. $\Box$

We end this section by providing a bit of information on the relationship between A and W. By Theorem 6.9, it follows that W is not interpretable in A. On the other hand, it turns out that W is model-interpretable in A. Consider any model $\mathcal{M}$ of A. In case when for all arithmetics N in $\mathcal{M}$ we have $\Box_A^N \bot$, we find that $M \models W$. So we can take the identity translation to provide an inner model of M. Suppose, for some N, we have $\mathcal{M} \models \mathsf{con}^N(A)$. We note that the proof of ($\$_0$) works for an arbitrary arithmetic. So we have $\mathcal{M} \models \mathsf{con}^N(W)$. We use the Henkin interpretation to provide an inner model of W.

Appendix A: More Details on the Basics

In this appendix we explain some basic notions in somewhat more detail.

A.1 Translations and interpretations The notion of interpretation that we will employ in this paper will be *m-dimensional interpretation without parameters*. There are two extensions of this notion: we can consider piecewise interpretations, and we can add parameters. We refrain from considering piecewise interpretations. We explain why in Section A.3. We sketch a few basic ingredients of adding parameters in Section A.4. We explain why, in the sequential case, addition of parameters makes no difference for the provability logic of all arithmetics of a given theory in Remark 3.10.

Consider two signatures Σ and Θ. An m-dimensional translation $\tau : \Sigma \to \Theta$ is a quadruple $\langle \Sigma, \delta, \mathcal{F}, \Theta \rangle$, where $\delta(v_0, \ldots, v_{m-1})$ is a Θ-formula and where for any n-ary predicate P of Σ, $\mathcal{F}(P)$ is a formula $A(\vec{v}_0, \ldots, \vec{v}_{n-1})$ in the language of signature Θ, where $\vec{v}_i = v_{i0}, \ldots, v_{i(m-1)}$. In the case of both δ and A all free variables are among the variables shown. Moreover, if $i \neq j$ and $k \neq \ell$, then v_{ik} is syntactically different from $v_{j\ell}$.

We demand that we have $\vdash \mathcal{F}(P)(\vec{v}_0, \ldots, \vec{v}_{n-1}) \to \bigwedge_{i<n} \delta(\vec{v}_i)$. Here $\vdash$ is provability in predicate logic. This demand is inessential, but it is convenient to have.

We define B^τ as follows:

- $(P(x_0, \ldots, x_{n-1}))^\tau := \mathcal{F}(P)(\vec{x}_0, \ldots, \vec{x}_{n-1})$;
- $(\cdot)^\tau$ commutes with the propositional connectives;[15]
- $(\forall x\, A)^\tau := \forall \vec{x}\, (\delta(\vec{x}) \to A^\tau)$;
- $(\exists x\, A)^\tau := \exists \vec{x}\, (\delta(\vec{x}) \wedge A^\tau)$.

There are two worries about this definition. First, what variables $\vec{x}_i$ on the side of the translation A^τ correspond with x_i in the original formula A? The second worry is that substitution of variables in δ and $\mathcal{F}(P)$ may cause variable clashes. These worries are never important in practice: we choose "suitable" sequences $\vec{x}$ to correspond to variables x, and we avoid clashes by α-conversions. However, if we want to give precise definitions of translations and, for example, of composition of translations, these problems come into play. We will address these problems elsewhere.

We allow identity to be translated to a formula that is not identity. There is some tension between this choice and the treatment of identity as a logical constant. The reason is that the notion of logical constant can do several kinds of work. It may be *obligatory in the language* and it may be *preserved under translation*. For identity we only ask that it is obligatory.

There are several important operations on translations.

- The identity translation is id_Σ. We take $\delta_{\mathsf{id}_\Sigma}(v) := v = v$ and $\mathcal{F}(P) := P(\vec{v})$.

- We can compose translations. Suppose $\tau : \Sigma \to \Theta$ and $\nu : \Theta \to \Lambda$. Then $\nu \circ \tau$ or $\tau\nu$ is a translation from Σ to Λ. We define
  - $\delta_{\tau\nu}(\vec{v}_0, \ldots, \vec{v}_{m_\tau - 1}) := \bigwedge_{i<m_\tau} \delta_\nu(\vec{v}_i) \wedge (\delta_\tau(v_0, \ldots, v_{m_\tau - 1}))^\nu$;
  - $P_{\tau\nu}(\vec{v}_{0,0}, \ldots, \vec{v}_{0,m_\tau - 1}, \ldots, \vec{v}_{n-1,0}, \ldots, \vec{v}_{n-1,m_\tau - 1}) := \bigwedge_{i<n, j<m_\tau} \delta_\nu(\vec{v}_{i,j}) \wedge (P(v_0, \ldots, v_{n-1})^\tau)^\nu$.
- Let $\tau, \nu : \Sigma \to \Theta$, and let A be a sentence of signature Θ. We define the disjunctive translation $\sigma := \tau\langle A\rangle\nu : \Sigma \to \Theta$ as follows. We take $m_\sigma := \mathsf{max}(m_\tau, m_\nu)$. We write $\vec{v} \restriction n$, for the restriction of $\vec{v}$ to the first n variables, where $n \leq \mathsf{length}(\vec{v})$:
  - $\delta_\sigma(\vec{v}) := (A \wedge \delta_\tau(\vec{v} \restriction m_\tau)) \vee (\neg A \wedge \delta_\nu(\vec{v} \restriction m_\nu))$;
  - $P_\sigma(\vec{v}_0, \ldots, \vec{v}_{n-1}) := (A \wedge P_\tau(\vec{v}_0 \restriction m_\tau, \ldots, \vec{v}_{n-1} \restriction m_\tau)) \vee (\neg A \wedge P_\nu(\vec{v}_0 \restriction m_\nu, \ldots, \vec{v}_{n-1} \restriction m_\nu))$.

Note that in the definition of $\tau\langle A\rangle\nu$ we used a padding mechanism. In case, for example, $m_\tau < m_\nu$, the variables $v_{m_\tau}, \ldots, v_{m_\nu - 1}$ are used "vacuously" when we have A. If we had piecewise interpretations, where domains are built up from pieces with possibly different dimensions, we could avoid padding by building the domain of disjoint pieces with different dimensions.

A translation relates signatures; an interpretation relates theories. An interpretation $K : U \to V$ is a triple $\langle U, \tau, V\rangle$, where U and V are theories and $\tau : \Sigma_U \to \Sigma_V$. We demand that for all axioms A of U, we have $V \vdash A^\tau$.

In the context of the formalization of interpretability, we have to distinguish between *axioms-interpretability*, which is the notion we just introduced and *theorems-interpretability*, where we demand that for all theorems A of U, we have $V \vdash A^\tau$. In the real world these notions are equivalent, but we need a principle like Σ_1-collection to prove that, so, for example Buss's theory S^1_2 does not "know" this equivalence. See [32] for more information about this matter.

Remark A.1 The design choice to make interpretations a transition between theories has many advantages. It allows us to build various categories of theories and interpretations; it allows us to have a decent model functor on categories of theories and interpretations; in various arguments, it reminds us *where we are*, and so forth. However, in some cases, the typing regime is somewhat stifling. For example, if you have an interpretation $K : U \to V$ and an extension W of V, then it would seem that K is also an interpretation of U in W. The typing regime forces us to say that it is a lifting $K \uparrow W : U \to W$; that is the interpretation based on τ_K, and so forth. In this paper we will remain faithful to the typing regime, but we will alleviate it a bit by the convention below.

- Suppose $K : U \to V$. We often write A^K for A^{τ_K}, in the context of a theory W that extends V.

Here are some further definitions.

- We write $\overline{U}$ for the set of theorems of U. Suppose $K : U \to V$. We write $\overline{K} := \{A \mid V \vdash A^K\}$. We note that $\overline{U} \subseteq \overline{K}$. If $\overline{K} = \overline{U}$, we will say that K is *faithful*.
- $\mathsf{ID}_U : U \to U$ is the interpretation $\langle U, \mathsf{id}_{\Sigma_U}, U\rangle$.
- Suppose $\overline{U} \subseteq \overline{V}$. Then, $\mathcal{E}_{UV} : U \to V$ is $\langle U, \mathsf{id}_{\Sigma_U}, V\rangle$.
- Suppose $K : U \to V$ and $M : V \to W$. Then, $KM := M \circ K : U \to W$ is $\langle U, \tau_M \circ \tau_K, W\rangle$.

- Suppose $K : W \to U$ and $U \subseteq V$. We write $K \uparrow V$ for $\mathcal{E}_{UV} \circ K$.
- Suppose $M : V \to Z$ and $U \subseteq V$. We write $U \downarrow M$ for $M \circ \mathcal{E}_{UV}$.
- Suppose $K : U \to (V + A)$ and $M : U \to (V + \neg A)$. Then $K\langle A\rangle M : U \to V$ is the interpretation $\langle U, \tau_K\langle A\rangle\tau_M, V\rangle$. In an appropriate category $K\langle A\rangle M$ is a special case of a product.

The notation $K : U \to V$ is inspired by the idea of interpretations as arrows in a category. There is also an intuition of interpretability as a generalization of provability. The traditional notation and notions associated to this intuition are as follows:

- $K : U \lhd V$ stands for $K : U \to V$.
- $K : V \rhd U$ stands for $K : U \to V$.
- $U \lhd V$ stands for $\exists K\ K : U \lhd V$; we say that U is *interpretable* in V.
- $V \rhd U$ stands for $\exists K\ K : V \rhd U$; we say that V *interprets* U.
- $U \lhd_{\mathsf{loc}} V$ means that all finitely axiomatized subtheories U_0 of U are interpretable in V; we say that U is *locally interpretable* in V.
- $U \lhd_{\mathsf{mod}} V$ means that, for every $\mathcal{M} \models V$, there is a translation τ such that $\tau(\mathcal{M}) \models U$; we say that U is *model interpretable* in V.

A.2 i-morphisms Consider an interpretation $K : U \to V$. We can view this interpretation as a uniform way of constructing internal models $\tau_K(\mathcal{M})$ of U from models $\mathcal{M}$ of V. This construction gives us the contravariant model functor as soon as we have defined an appropriate category of interpretations.

Now consider two interpretations $K, M : U \to V$. Between the inner models $\tau_K(\mathcal{M})$ and $\tau_M(\mathcal{M})$ we have the usual structural morphisms of models. We are interested in the case when these morphisms are V-definable and uniform over models. This idea leads to the following definition. An *i-morphism* $M : K \to M$ is a triple $\langle K, F(\vec{u}, \vec{v}), M\rangle$, where $F(\vec{u}, \vec{v})$ is a V-formula and where $\vec{u}$ has length m_K and $\vec{v}$ has length m_M. We demand that

- $V \vdash F(\vec{u}, \vec{v}) \to (\delta_K(\vec{u}) \wedge \delta_M(\vec{v}))$,
- $V \vdash \delta_K(\vec{u}) \to \exists \vec{v}\ (\delta_M(\vec{v}) \wedge F(\vec{u}, \vec{v}))$,
- $V \vdash (\vec{u}_0 =_K \vec{u}_1 \wedge F(\vec{u}_0, \vec{v}_0) \wedge F(\vec{u}_1, \vec{v}_1)) \to \vec{v}_0 =_M \vec{v}_1$,
- $V \vdash (\vec{u}_0 =_K \vec{u}_1 \wedge \vec{v}_0 =_M \vec{v}_1 \wedge F(\vec{u}_0, \vec{v}_0)) \to F(\vec{u}_1, \vec{v}_1)$,
- $V \vdash (P_K(\vec{u}_0, \ldots, \vec{u}_{n-1}) \wedge \bigwedge_{i<n} F(\vec{u}_i, \vec{v}_i)) \to P_M(\vec{v}_0, \ldots, \vec{v}_{n-1})$.

Clearly, $F : K \to M$ is an i-morphism if and only if, for all models $\mathcal{M}$ of V, $F^{\mathcal{M}}$ represents a morphism of models from $\tau_K(\mathcal{M})$ to $\tau_M(\mathcal{M})$.

Two i-morphisms $F, G : K \to M$ are *i-equal*, when $V \vdash \forall \vec{u}, \vec{v}\, (F(\vec{u}, \vec{v}) \leftrightarrow G(\vec{u}, \vec{v}))$.

In the obvious way, we can define the identity i-morphism $\mathsf{Id}_K : K \to K$, composition of i-morphisms, i-isomorphisms, and so forth. One can show that these operations preserve i-equality. Moreover, i-isomorphisms really are isomorphisms in the categories given by these operations.

We will say that two interpretations K, M are *i-equivalent* when there is an i-isomorphism between them, that is, when they are i-isomorphic.

We will *not* divide out i-equivalence of interpretations. This enables us to use the notation τ_M meaningfully, to speak about the dimension of an interpretation, and so forth. However, we demand that operations on interpretations preserve i-equivalence. It is easy to see that, for example, the operation $K, M \mapsto K\langle A\rangle M$ preserves i-equivalence. Moreover, if K and M are i-equivalent, then $\overline{K} = \overline{M}$.

One can show, by a simple compactness argument, that K and M are i-isomorphic if and only if, for every $\mathcal{M} \models V$, there is an F such that $F^{\mathcal{M}}$ represents an isomorphism between $\tau_K(\mathcal{M})$ and $\tau_M(\mathcal{M})$.

The category INT_1 is the category of theories (as objects) and interpretations modulo i-equivalence (as arrows). One may show that we have indeed defined a category. The relation of i-equivalence is preserved by composition. Two theories U and V are *bi-interpretable* if they are isomorphic in INT_1. Wilfrid Hodges [16, p. 222] calls this notion *homotopy*.

Thus, U and V are bi-interpretable if there are interpretations $K : U \to V$ and $M : V \to U$, so that $M \circ K$ is i-isomorphic to ID_U and $K \circ M$ is i-isomorphic to ID_V. We call the pair K, M a *bi-interpretation* between U and V. One can show that the components of a bi-interpretation are faithful interpretations. Many good properties of theories like finite axiomatizability, decidability, and κ-categoricity are preserved by bi-interpretations.

A.3 Piecewise interpretations There is a notion of *piecewise interpretability* where we allow the domain of the interpretation to be built up from finitely many pieces with possibly different dimensions. An example of this is the construction where we add points at infinity to a points-only version of plane geometry. We could have a piece with the original points and strict identity and a piece with pairs of distinct points with the following equivalence relation: (x, y) is equivalent to (u, v) if there is no point w that is both collinear with x and y and with u and v. Of course, we can replace this interpretation by a one-piece interpretation that is isomorphic (in an appropriate sense) to it in various obvious ways. For example, a pair (x, y) could represent a point at infinity if $x \neq y$ and the point x if $x = y$.

One can show that, if we have $V \vdash \exists x, y\ x \neq y$, then any piecewise interpretation is isomorphic (in an appropriate sense) to an interpretation without pieces (but in general with higher dimension). It follows that a theory piecewise interprets a weak arithmetic if and only if it interprets this arithmetic nonpiecewise via an interpretation that is in a relevant sense *i-equivalent* to the original one.

A.4 Parameters In general, interpretations are allowed to have parameters. We will briefly sketch how to add parameters to our framework. We first define a translation with parameters. The parameters of the translation are given by a fixed sequence of variables $\vec{w}$ that we keep apart from all other variables. A translation is defined as before, but for the fact that now the variables $\vec{w}$ are allowed to occur in the domain and in the translations of the predicate symbols in addition to the variables that correspond to the argument places. Officially, we represent a translation $\tau_{\vec{w}}$ with parameters $\vec{w}$ as a quintuple $\langle \Sigma, \delta, \vec{w}, F, \Theta \rangle$. The parameter sequence may be empty; in this case our interpretation is parameter-free.

An interpretation with parameters $K : U \to V$ is a quadruple $\langle U, \alpha, E, \tau_{\vec{w}}, V \rangle$, where $\tau_{\vec{w}} : \Sigma_U \to \Sigma_V$ is a translation and α is a V-formula containing at most $\vec{w}$ free. The formula α represents the parameter domain. For example, if we interpret the hyperbolic plain in the Euclidean plain via the Poincaré interpretation, we need two distinct points to define a circular disk. These points are parameters of the construction, the parameter domain is $\alpha(w_0, w_1) = (w_0 \neq w_1)$. (For this specific example, we can also find a parameter-free interpretation.) The formula E represents an equivalence relation on the parameter domain. In practice this is always pointwise identity for parameter sequences, but for reasons of theory one must admit other equivalence relations too. We demand the following:

- $\vdash \delta_{\tau,\vec{w}}(\vec{v}) \to \alpha(\vec{w})$;
- $\vdash P_{\tau,\vec{w}}(\vec{v}_0, \ldots, \vec{v}_{n-1}) \to \alpha(\vec{w})$;
- $V \vdash \exists \vec{w}\, \alpha(\vec{w})$;
- $V \vdash E(\vec{w}, \vec{z}) \to (\alpha(\vec{w}) \wedge \alpha(\vec{z}))$;
- V proves that E represents an equivalence relation on the sequences forming the parameter domain;
- $\vdash E(\vec{w}, \vec{z}) \to \forall \vec{x}\, (\delta_{\tau,\vec{w}}(\vec{x}) \leftrightarrow \delta_{\tau,\vec{z}}(\vec{x}))$;
- $\vdash E(\vec{w}, \vec{z}) \to \forall \vec{x}_0, \ldots, \vec{x}_{n-1}\, (P_{\tau,\vec{w}}(\vec{x}_0, \ldots, \vec{x}_{n-1}) \leftrightarrow P_{\tau,\vec{z}}(\vec{x}_0, \ldots, \vec{x}_{n-1}))$;
- for all U-axioms A, $V \vdash \forall \vec{w}\, (\alpha(\vec{w}) \to A^{\tau,\vec{w}})$.

We can lift the various operations in the obvious way. Note that the parameter domain of $N := M \circ K$ and the corresponding equivalence relation should be

- $\alpha_N(\vec{w}, \vec{u}_0, \ldots, \vec{u}_{k-1}) := \alpha_M(\vec{w}) \wedge \bigwedge_{i<k} \delta_{\tau_M}(\vec{w}, \vec{u}_i) \wedge (\alpha_K(\vec{u}))^{\tau_M,\vec{w}}$,
- $E_N(\vec{w}, \vec{u}_0, \ldots, \vec{u}_{k-1}, \vec{z}, \vec{v}_0, \ldots, \vec{v}_{k-1}) := E_M(\vec{w}, \vec{z}) \wedge \bigwedge_{i<k} \delta_{\tau_M}(\vec{w}, \vec{u}_i) \wedge \bigwedge_{i<k} \delta_{\tau_M}(\vec{w}, \vec{v}_i) \wedge (E_K(\vec{u}, \vec{v}))^{\tau_M,\vec{w}}$.

Consider interpretations $K, M : U \to V$. An i-morphism $\varphi : K \to M$ is a triple $\langle K, G, F, M \rangle$, where $G(\vec{u}, \vec{w})$ and $F(\vec{u}, \vec{w}, \vec{x}, \vec{y})$ are V-formulas.[16] We write $F^{\vec{u};\vec{w}}(\vec{x}, \vec{y})$ for F. We demand that

- V proves that G is a surjective relation between α_K/E_K and α_M/E_M;[17]
- $V \vdash F^{\vec{u};\vec{w}}(\vec{x}, \vec{y}) \to G(\vec{u}, \vec{w})$;
- V proves that if $G(\vec{u}, \vec{w})$, then $F^{\vec{u};\vec{w}}$ is a function from $\delta_K/{=_K}$ to $\delta_M/{=_M}$;
- V proves that if $E_K(\vec{u}_0, \vec{u}_1)$ and $E_M(\vec{w}_0, \vec{w}_1)$, then $F^{\vec{u}_0,\vec{w}_0}$ is the same function is $F^{\vec{u}_1,\vec{w}_1}$.

Finally, we say that two i-maps φ_0 and φ_1 are *i-equal* if V proves that G_{φ_0} and G_{φ_1} and F_{φ_0} and F_{φ_1} are the same.

The definitions of the identity i-morphism and of composition of i-morphisms are as is to be expected. We can compute what an i-isomorphism is: G is, V-verifiably, a bijection between α_K/E_K and α_M/E_M, and V proves that, if $G(\vec{u}, \vec{w})$, then $F^{\vec{u};\vec{w}}$ is a bijection between $\delta_K/{=_K}$ and $\delta_M/{=_M}$.

A.5 Complexity measures *Restricted provability* plays an important role in this paper. An n-proof is a proof from axioms with Gödel number smaller than or equal to n only involving formulas of complexity smaller than or equal to n. To work conveniently with this notion, a good complexity measure is needed. This should satisfy three conditions. (i) Eliminating terms in favor of a relational formulation should raise the complexity only by a fixed standard number. (ii) Translation of a formula via the translation corresponding to an interpretation K should raise the complexity of the formula by a fixed standard number depending only on K. (iii) The tower of exponents involved in cut elimination should be of height linear in the complexity of the formulas involved in the proof.

Such a good measure of complexity together with a verification of desideratum (iii)—a form of nesting degree of quantifier alternations—is supplied in the work of Philipp Gerhardy [12], [13]. It is also provided by Samuel Buss in his preliminary draft [7]. Buss also proves that (iii) is fulfilled.

Gerhardy's measure corresponds to the following formula classes:

- AT is the class of atomic formulas;
- $\mathsf{N}^\star_{-1} = \Sigma^\star_{-1} = \Pi^\star_{-1} := \emptyset$;
- $\mathsf{N}^\star_n ::= \mathsf{AT} \mid \neg\mathsf{N}^\star_n \mid (\mathsf{N}^\star_n \wedge \mathsf{N}^\star_n) \mid (\mathsf{N}^\star_n \vee \mathsf{N}^\star_n) \mid (\mathsf{N}^\star_n \to \mathsf{N}^\star_n) \mid \forall\, \Pi^\star_n \mid \exists\, \Sigma^\star_n$;
- $\Sigma^\star_n ::= \mathsf{AT} \mid \neg\Pi^\star_n \mid (\mathsf{N}^\star_{n-1} \wedge \mathsf{N}^\star_{n-1}) \mid (\Sigma^\star_n \vee \Sigma^\star_n) \mid (\Pi^\star_n \to \Sigma^\star_n) \mid \forall\, \Pi^\star_{n-1} \mid \exists\, \Sigma^\star_n$;
- $\Pi^\star_n ::= \mathsf{AT} \mid \neg\Sigma^\star_n \mid (\Pi^\star_n \wedge \Pi^\star_n) \mid (\mathsf{N}^\star_{n-1} \vee \mathsf{N}^\star_{n-1}) \mid (\mathsf{N}^\star_{n-1} \to \mathsf{N}^\star_{n-1}) \mid \forall\, \Pi^\star_n \mid \exists\, \Sigma^\star_{n-1}$.

We may define $\rho(A)$ as the minimal n such that A is in $\mathsf{N}^\star_n$.[18]

Samuel Buss gives the following formula classes:

- $\Sigma^*_0 = \Pi^*_0 =$ the class of quantifier-free formulas;
- $\Sigma^*_n ::= \Sigma^*_{n-1} \mid \Pi^*_{n-1} \mid \neg\, \Pi^*_n \mid (\Sigma^*_n \wedge \Sigma^*_n) \mid (\Sigma^*_n \vee \Sigma^*_n) \mid (\Pi^*_n \to \Sigma^*_n) \mid \exists\, \Sigma^*_n$;
- $\Pi^*_n ::= \Sigma^*_{n-1} \mid \Pi^*_{n-1} \mid \neg\, \Sigma^*_n \mid (\Pi^*_n \wedge \Pi^*_n) \mid (\Pi^*_n \vee \Pi^*_n) \mid (\Sigma^*_n \to \Pi^*_n) \mid \forall\, \Pi^*_n$.

We may define $\rho(A)$ as the smallest n such that A is in Σ^*_n. This is the same measure as was employed in [35]. For our purposes it does not matter whether we use Gerhardy's or Buss's definition.

Appendix B: On Faithful Interpretability

We assume that the formalization of syntax is standard, so that the code of a subformula C of B is smaller than the code of B, and so forth. We also assume that the proof-predicate is standard, so that every proof p has a single conclusion C with $C < p$.

Theorem B.1 *Consider a theory T, and suppose that N is an arithmetic in T. Let Γ be any T-definable class of T-sentences for which T contains a definable truth predicate, say,* TRUE, *for sentences coded in N. We only need that* TRUE *satisfies Tarski's convention. We assume that the set of codes of elements of Γ has a fixed binumeration in T (which,* par abus de langage, *we call also Γ). So we assume that if $A \in \Gamma$, then $T \vdash A \in \Gamma$ and, if $A \notin \Gamma$, then $T \vdash A \notin \Gamma$. Then, there is a unary predicate $A(x)$, such that*

(i) $T \vdash A(x) \to x \in N$;
(ii) $T \vdash (A(x) \wedge x =_N y) \to A(y)$;
(iii) $T \vdash (A(x) \wedge A(y)) \to x =_N y$;
(iv) for any n, $T + A(\underline{n})$ is Γ-conservative over T; here $\underline{n}$ is the N-numeral of n.

Proof We define, for $p \in N$ and B an N-code of a formula with at most one designated variable v_0 free:

- $\Im(p, x, B) :\leftrightarrow p, x \in N \wedge \exists C \in \Gamma\, (\mathsf{proof}^N_T(p, B(x) \to C) \wedge \neg\, \mathsf{TRUE}(C))$.

Here $B(x)$ in the context of proof means the code of the result of substituting the numeral of x for v_0 in B. We find, using the Gödel fixed-point lemma, a formula A with the following property:

$$T \vdash A(x) \leftrightarrow \exists p\, \big(\Im(p, x, A) \wedge \forall q <_N p\, \forall y \in N\, \neg\, \Im(q, y, A)\big).$$

Clearly, we have (i) and (ii). We prove the uniqueness clause (iii).

Reason in T. Suppose that $x \neq_N y$ and $A(x)$ and $A(y)$. Let p be a witness for $A(x)$, and let q be a witness of $A(y)$. By our assumption about the proof predicate, we find that $p \neq_N q$, and, hence, $p <_N q$ or $q <_N p$. By the specification of A, this is impossible.

We move to the metatheory again. We prove (iv). We write $r : T \vdash E$ when r is (a code of) a T-proof of E.

We assume, to get a contradiction, that, for some n, $A(n)$ is not Γ-conservative over T. Let p be the smallest proof such that, for some n and some $C \in \Gamma$, we have $p : T \vdash A(\underline{n}) \to C$ and $T \nvdash C$. It follows that, for all $q < p$, and all m and all $D \in \Gamma$, if $q : T \vdash A(\underline{m}) \to D$, then $T \vdash D$. It follows that $T \vdash \forall q <_N \underline{p}\, \forall y \in N \neg \Im(q, y, A)$. We also find $T \vdash \neg C \to \Im(\underline{p}, \underline{n}, A)$. Ergo $T + \neg C \vdash A(\underline{n})$. In other words, we have both $T + A(\underline{n}) \vdash C$ and $T + \neg A(\underline{n}) \vdash C$. Ergo $T \vdash C$, a contradiction. □

By the specification of the above formula A it functions as a closed partial N-numerical term in T. For this reason we will write $\tau \simeq_N x$ for $A(x)$.

Theorem B.2 *Let T be a theory, and suppose that N is an arithmetic in T. Let Σ be a finite signature for predicate logic. We call predicate logic of signature Σ: FOL_Σ. Let $\alpha(x)$ be a formula in the language of T such that T proves that all elements of $\{x \mid \alpha(x)\}$ are N-codes of Σ-sentences. We write $\Box_\alpha$ for provability from the sentences coded by the elements of $\{x \mid \alpha(x)\}$. We write* $\mathsf{con}(\alpha)$ *for* $\neg \Box_\alpha \bot$.

There is an interpretation $H : (T + \mathsf{con}(\alpha)) \rhd \mathsf{FOL}_\Sigma$ such that, for any Σ-sentence A, we have $\mathcal{T} + \mathsf{con}(\alpha) + \Box_\alpha A \vdash A^H$. We say that H is a Henkin interpretation *of α.*

Proof We can see this by inspection of the usual proof of the interpretation existence lemma. The basic idea is that we formalize the Henkin construction, employing definable cuts whenever we would have used induction in PA. See, for example, [32] or [34]. □

We proceed with our upper-bound result.

Theorem B.3 *Let T be any theory. Suppose $K : T \rhd U$. Let A be any T-sentence, and let N be an arithmetic in $T + A$. Then there is an interpretation $M : T \rhd U$ such that, for any U-sentence B, $T \vdash B^M \Rightarrow T + A \vdash \Box^N_U B$.*

Proof Consider $T + A$. We first show that we may assume without loss of generality that we have a Σ_1-truth predicate for N.

By Theorem 3.9, we may shorten N to a $T + A$-definable cut N' such that $T + A$ contains a truth predicate, say, TRUE, for the Σ_1-sentences of N', that is, for every S in Σ_1, $U \vdash S^{N'} \leftrightarrow \mathsf{TRUE}(S)$, where S inside the truth predicate is coded in N'.

Note that

$$T + A \vdash \Box^{N'}_U B \quad \Rightarrow \quad T + A \vdash \Box^N_U B.$$

It follows that it is sufficient to prove our theorem for N'.

Thus, we may assume that T contains a truth predicate, say, TRUE, for the Σ_1-sentences of N.

Let τ be the partial closed term promised by Theorem B.1 for N and Σ_1. We fix some standard enumeration C_x of the U-sentences in such a way that T verifies its elementary properties with respect to N. We specify M by cases. In case we have $\neg A$, we take M equal to K. Suppose we have A. We may now work in $T + A$. Let $U^* := U + \{C_x \mid \tau \simeq_N x\}$. Note that (i) U^* is not Δ^b_1-axiomatized, that (ii) in talking about U^* we are really talking about the formula defining the axiom set, and that (iii) the definition of U^* only makes sense in the presence of A. In case $\mathsf{incon}^N(U^*)$,[19] we take M again equal to K. If $\mathsf{con}(U^*)$, we take M equal to the Henkin interpretation H of U^*. In other words, we take

$$M := H\langle A \wedge \mathsf{con}^N(U^*)\rangle K.$$ [20]

Clearly, $M : T \rhd U$. Suppose $T \vdash B^M$. Let $\neg B = C_n$. We have

$$T + A + \tau \simeq_N \underline{n} \vdash \text{“}(U + \neg B) = U^*\text{”}.$$

Here “=” stands for extensional identity. Hence,

$$T + A + (\tau = \underline{n}) + \mathsf{con}^N(U + \neg B) \vdash \neg B^M.$$

Thus, $T + A \vdash (\tau \simeq_N \underline{n}) \to \Box_U^N B$. By the Σ_1-conservativity of $\tau \simeq_N \underline{n}$, we find $T + A \vdash \Box_U^N B$. □

From Theorem B.3 we can derive a basic result about interpretability. We say that a theory U is *trustworthy* if, whenever $U \rhd V$, then $U \rhd_{\mathsf{faith}} V$.

Theorem B.4 *The following are equivalent.*

(i) U is trustworthy.
(ii) U faithfully interprets predicate logic with a binary predicate R.
(iii) For some A, $U + A$ contains a Σ_1-sound arithmetic N.

Proof Trivially (ii) follows from (i).

Suppose (ii). Let B be the single axiom of adjunctive set theory AS. In AS we can provide a Σ_1-sound interpretation M of S^1_2. Suppose that K is the promised faithful interpretation of predicate logic with a binary relation symbol R in U. Then, as is easily seen, $A := B^K$ and $N := K \circ M$ satisfy the desiderata of (iii).

Finally, we assume (iii). Suppose $K : U \rhd V$. Let M be the interpretations of V provided by Theorem B.3, such that $M : U \rhd V$ and $U \vdash B^M$ implies $U + A \vdash \Box_V^N B$. By the Σ_1-soundness of N, we may conclude that $U \vdash B^M$ implies $V \vdash B$, and we are done. □

Notes

1. See [6] or [15] for an explanation of the relevant formula classes.

2. The function $\ulcorner\cdot\urcorner$ sends a syntactical object to its Gödel number. The function $\underline{(\cdot)}$ sends a number to its numeral. We will employ efficient numerals that reflect binary notation.

3. Similarly, if U is polysequential via an m-dimensional interpretation of AS, any arithmetic N in U is i-equivalent to an m-dimensional arithmetic N.

4. The two classical works on this subject are Lesan [20] and Paris and Dimitracopoulos [23].

5. See Kaye, Paris, and Dimitracopoulos [18] and Cordon–Franco, Fernández-Margarit, and Lara-Martín [8] for details on $\mathrm{I}\Pi_1^-$.

6. This fact is folklore. I learned it first from Sergei Artemov around 1984.

7. The theory B as defined here seems to suffice. However, I am not sure that a definition using $\exists^* \Pi^{\mathsf{b}}_1$-sentences is not more natural.

8. It takes a little argument to prove the equivalence of our formulation of CFL and the formulation in [8]. For the definition it does not matter, modulo provable equivalence over T^1_2, whether the Σ_1-sentences are written in the form $\exists\Delta_0$ or $\exists^*\Delta_0$. We may also consider sentences in $\exists^*\Delta_0$ and just relativize only the first of the block of existential quantifiers to $\mathcal{J}$.

9. I was aware of two essentially different constructions for the multiple fixed-point lemma. Vincent van Oostrom, after I asked him, provided a third construction. All three constructions automatically guarantee the desired property even in the presence of nontrivial automorphisms of the frame. One reason that this happens is that the choice of substitution variables is explicitly arithmetically coded in the construction.

10. Note that the L_i are not necessarily uniquely determined by the fixed-point equations. Thus, we are looking at some choice of the L_i.

11. We use the work of [12], [13], and [7] here. See Section A.5 of Appendix A.

12. We remind the reader that N_0MK stands for $K \circ M \circ N_0$.

13. Instead of $(\Box^{\mathsf{gn}(\tau)}_{\mathsf{S}^1_2}\Box_A\bot)^{\tilde{\tau}}$, we could also have used $\Box^{\tilde{\tau},\mathsf{gn}(\tau)+1}_A\bot$, but it seems to me that the argument is a bit shorter under the present choice.

14. Alternatively, we could have constructed W using a version of Craig's trick, taking as axioms $(\underline{p} = \underline{p}) \wedge (\Box^{\mathsf{gn}(N)}_{\mathsf{S}^1_2}\Box_A\bot)^N$, where p is an A-proof that N is an arithmetic.

15. If we have a complex formula A, the translation A^τ could be satisfied in a model even if the sequences of values of the variables corresponding to the free variables in A are not in the domain of the translation in that model. One alternative option for the definition is to add a conjunction that stipulates that these sequences are in the domain. Thus, we would always have $\vdash A^\tau \to \delta_K(\vec{x})$, where $\vec{x}$ is a sequence corresponding to a free variable in A. We will refrain from doing this. The cost is that, for example, the definition of composition of translations becomes more complicated.

16. In G and F we could allow extra parameters, $\vec{z}$, the *eigenparameters* of G and F. We will refrain from doing that here to unburden the presentation a bit.

17. It seems a more logical choice to demand that G represent a function from α_K/E_K to α_M/E_M. There are also sound theoretical reasons for that choice. However, the definition of *initial embedding* that we need in Section 3 does not work under this second choice. So for the purposes of at least this paper we seem to need the definition given in the main text.

18. Vincent van Oostrom gave a variant of this formulation of Gerhardy's measure in conversation.

19. In writing $(\mathsf{in})\mathsf{con}^N(U^*)$, we intend no relativization of the formula defining the axiom set, only of the proofs.

20. Strictly speaking, we should not have K here but $K \upharpoonright (T + (\neg A \vee \mathsf{incon}(U^*)))$.

References

[1] Artemov, S., and L. Beklemishev, "Provability logic," pp. 229–403 in *Handbook of Philosophical Logic, Volume 13*, 2nd edition, edited by D. Gabbay and F. Guenthner, Springer, Dordrecht, 2004. 82

[2] Beklemishev, L., and A. Visser, "On the limit existence principles in elementary arithmetic and Σ^0_n-consequences of theories," *Annals of Pure and Applied Logic*, vol. 136 (2005), pp. 56–74. Zbl 1087.03037. MR 2162847. DOI 10.1016/j.apal.2005.05.005. 90

[3] Beklemishev, L., and A. Visser, "Problems in the Logic of Provability," pp. 77–136 in *Mathematical Problems from Applied Logic, I: Logics for the XXIst Century*, edited by D. M. Gabbay, S. S. Concharov, and M. Zakharyashev, vol. 4 of *International Mathematical Series*, Springer, New York, 2006. Zbl 1100.03051. MR 2185909. DOI 10.1007/0-387-31072-X_2. 91

[4] Boolos, G., *The Logic of Provability*, Cambridge Univ. Press, Cambridge, 1993. Zbl 0891.03004. MR 1260008. 82

[5] Boolos, G., and G. Sambin, "Provability: The emergence of a mathematical modality," *Studia Logica*, vol. 50 (1991), pp. 1–23. Zbl 0742.03003. MR 1152776. DOI 10.1007/BF00370383. 82

[6] Buss, S. R., *Bounded Arithmetic: Studies in Proof Theory*, vol. 3 of *Lecture Notes*, Bibliopolis, Napoli, 1986. Zbl 0649.03042. MR 0880863. 82, 86, 91, 94, 115

[7] Buss, S., "Cut elimination *in situ*," preprint, 2013, http://math.ucsd.edu/~sbuss/ResearchWeb/index.html. 86, 112, 116

[8] Cordón-Franco, A., A. Fernández-Margarit, and F. Lara-Martín, "A note on parameter free Π_1-induction and restricted exponentiation," *Mathematical Logic Quarterly*, vol. 57 (2011), pp. 444–55. Zbl 1238.03048. MR 2838320. DOI 10.1002/malq.201010013. 95, 115, 116

[9] de Jongh, D., M. Jumelet, and F. Montagna, "On the proof of Solovay's theorem," *Studia Logica*, vol. 50 (1991), pp. 51-69. Zbl 0744.03057. MR 1152780. DOI 10.1007/BF00370387. 93

[10] Feferman, S., "Arithmetization of metamathematics in a general setting," *Fundamenta Mathematicae*, vol. 49 (1960), pp. 35–92. Zbl 0095.24301. MR 0147397. 99

[11] Feferman, S., "My route to arithmetization: The arithmetization of metamathematics," *Theoria*, vol. 63 (1997), pp. 168–81. MR 1730698. DOI 10.1111/j.1755-2567.1997.tb00746.x. 99

[12] Gerhardy, P., "Refined complexity analysis of cut elimination," pp. 212–25 in *Computer Science Logic*, edited by M. Baaz and J. Makovsky, vol. 2803 of *Lecture Notes in Computer Science*, Springer, Berlin, 2003. Zbl 1116.03318. MR 2047962. DOI 10.1007/978-3-540-45220-1_19. 86, 112, 116

[13] Gerhardy, P., "The role of quantifier alternations in cut elimination," *Notre Dame Journal of Formal Logic*, vol. 46 (2005), pp. 165–71. Zbl 1078.03045. MR 2150949. DOI 10.1305/ndjfl/1117755147. 86, 112, 116

[14] Guaspari, D., and R. M. Solovay, "Rosser sentences," *Annals of Mathematical Logic*, vol. 16 (1979), pp. 81–99. Zbl 0426.03062. MR 0530432. DOI 10.1016/0003-4843(79)90017-2. 93

[15] Hájek, P., and P. Pudlák, *Metamathematics of First-Order Arithmetic*, vol. 3 of *Perspectives in Mathematical Logic*, Springer, Berlin, 1993. Zbl 0889.03053. MR 1219738. DOI 10.1007/978-3-662-22156-3. 82, 86, 89, 98, 102, 115

[16] Hodges, W., *Model Theory*, vol. 42 of *Encyclopedia of Mathematics and Its Applications*, Cambridge Univ. Press, Cambridge, 1993. Zbl 0789.03031. MR 1221741. DOI 10.1017/CBO9780511551574. 85, 111

[17] Japaridze, G., and D. de Jongh, "The logic of provability," pp. 475–546 in *Handbook of Proof Theory*, edited by S. Buss, vol. 137 of *Studies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam, 1998. Zbl 0915.03019. MR 1640331. DOI 10.1016/S0049-237X(98)80022-0. 82

[18] Kaye, R., J. B. Paris, and C. Dimitracopoulos, "On parameter-free induction schemas," *Journal of Symbolic Logic*, vol. 53 (1988), pp. 1082–97. Zbl 0678.03025. MR 0973102. DOI 10.2307/2274606. 90, 95, 115

[19] Krajíček, J., "A note on proofs of falsehood," *Archiv für Mathematische Logik und Grundlagenforschung*, vol. 26 (1987), pp. 169–76. Zbl 0633.03056. MR 0905955. DOI 10.1007/BF02017501. 99

[20] Lesan, H., "Models of arithmetic," Ph.D. dissertation, University of Manchester, Manchester, 1978. 102, 115

[21] Lindström, P., "Provability logic: A short introduction," pp. 19–61 in *Philosophy in Sweden (Umeå, 1995)*, vol. 62 of *Theoria*, Theoria, Stockholm, 1996. Zbl 0897.03055. MR 1619332. DOI 10.1111/j.1755-2567.1996.tb00529.x. 82

[22] Mycielski, J., P. Pudlák, and A. S. Stern, "A lattice of chapters of mathematics (interpretations between theorems [theories])," *Memoirs of the American Mathematical Society*, vol. 84 (1990), no. 426. Zbl 0696.03030. MR 0997901. 86

[23] Paris, J. B., and C. Dimitracopoulos, "Truth definitions for Δ_0 formulae," pp. 317–29 in *Logic and Algorithmic (Zurich, 1980)*, vol. 30 of *Monographies de L'Enseignement Mathématique*, Université de Genève, Geneva, 1982. Zbl 0475.03033. MR 0648309. 102, 115

[24] Pudlák, P., "Some prime elements in the lattice of interpretability types," *Transactions of the American Mathematical Society*, vol. 280 (1983), pp. 255–75. Zbl 0561.03014. MR 0712260. DOI 10.2307/1999613. 86

[25] Pudlák, P., "Cuts, consistency statements and interpretations," *Journal of Symbolic Logic*, vol. 50 (1985), pp. 423–41. Zbl 0569.03024. MR 0793123. DOI 10.2307/2274231. 86, 88, 95, 99

[26] Pudlák, P., "On the length of proofs of finitistic consistency statements in first order theories," pp. 165–96 in *Logic Colloquium '84 (Manchester, England, 1984)*, vol. 120 of *Studies in Logic and the Foundations of Mathematics*, edited by J. B. Paris, A. J. Wilkie, and G. M. Wilmers, North-Holland, Amsterdam, 1986. Zbl 0619.03037. MR 0861424. DOI 10.1016/S0049-237X(08)70462-2. 88

[27] Smoryński, C., "Nonstandard models and related developments," pp. 179–229 in *Harvey Friedman's Research on the Foundations of Mathematics*, edited by L. A. Harrington, M. D. Morley, A. Scedrov, and S. G. Simpson, vol. 117 of *Studies in Logic and the Foundations of Mathematics*, North-Holland, Amsterdam, 1985. MR 0835259. DOI 10.1016/S0049-237X(09)70160-0. 99

[28] Smoryński, C., *Self-Reference and Modal Logic*, Universitext, Springer, New York, 1985. Zbl 0596.03001. MR 0807778. DOI 10.1007/978-1-4613-8601-8. 82

[29] Švejdar, V., "On provability logic," *Nordic Journal of Philosophical Logic*, vol. 4 (1999), pp. 95–116. Zbl 0352.02019. MR 1834733. 82

[30] Verbrugge, L., "Efficient metamathematics," Ph.D. dissertation, Institute for Logic, Language and Computation, University of Amsterdam, Amsterdam, 1993. 82, 91

[31] Visser, A., "Interpretability logic," pp. 175–209 in *Mathematical Logic*, edited by P. P. Petkov, Plenum, New York, 1990. MR 1083994. 99

[32] Visser, A., "The formalization of interpretability," *Studia Logica*, vol. 50 (1991), pp. 81–105. Zbl 0744.03023. MR 1152782. DOI 10.1007/BF00370389. 84, 109, 114

[33] Visser, A., "On the Σ_1^0-conservativity of Σ_1^0-completeness," *Notre Dame Journal of Formal Logic*, vol. 32 (1991), pp. 554–61. Zbl 0747.03026. MR 1146610. DOI 10.1305/ndjfl/1093635927. 82, 98

[34] Visser, A., "An inside view of EXP," *Journal of Symbolic Logic*, vol. 57 (1992), pp. 131–65. Zbl 0785.03008. MR 1150930. DOI 10.2307/2275181. 114

[35] Visser, A., "The unprovability of small inconsistency: A study of local and global interpretability," *Archive for Mathematical Logic*, vol. 32 (1993), pp. 275–98. Zbl 0795.03080. MR 1213062. DOI 10.1007/BF01387407. 86, 88, 99, 113

[36] Visser, A., "Faith and Falsity: A study of faithful interpretations and false Σ_1^0-sentences," *Annals of Pure and Applied Logic*, vol. 131 (2005), pp. 103–31. Zbl 1065.03038. MR 2097224. DOI 10.1016/j.apal.2004.04.008. 99, 101

[37] Visser, A., "Cardinal arithmetic in the style of Baron von Münchhausen," *Review of Symbolic Logic*, vol. 2 (2009), pp. 570–89. Zbl 1185.03088. MR 2568943. DOI 10.1017/S1755020309090261. 86

[38] Visser, A., "Can we make the second incompleteness theorem coordinate free?," *Journal of Logic and Computation*, vol. 21 (2011), pp. 543–60. Zbl 1262.03123. MR 2823429. DOI 10.1093/logcom/exp048. 95

[39] Visser, A., "Aspects of diagonalization and provability," Ph.D. dissertation, Utrecht University, Netherlands, 1981. 91

[40] Visser, A., "What is sequentiality?," pp. 229–69 in *New Studies in Weak Arithmetics*, vol. 211 of CSLI Lecture Notes, CSLI Publications and Presses Universitaires du Pôle de Recherche et d'Enseignement Supérieur Paris-est, Stanford, 2013. 86, 89

[41] Visser, A., "Peano Basso and Peano Corto," Mathematical Logic Quarterly, vol. 60 (2014), pp. 92–117. 95

[42] Wilkie, A. J., and J. B. Paris, "On the scheme of induction for bounded arithmetic formulas," *Annals of Pure and Applied Logic*, vol. 35 (1987), pp. 261–302. Zbl 0647.03046. MR 0904326. DOI 10.1016/0168-0072(87)90066-2. 91, 100

Faculty of Humanities, Philosophy
Utrecht University
Janskerkhof 13
3512BL Utrecht
The Netherlands
a.visser@uu.nl