Kripke Models Built from Models of Arithmetic

Paula Henk

Institute for Logic, Language and Computation (ILLC),
University of Amsterdam
P.Henk@uva.nl

Abstract. We introduce three relations between models of Peano Arithmetic (PA), each of which is characterized as an arithmetical accessibility relation. A relation R is said to be an *arithmetical accessibility relation* if for any model $\mathcal{M}$ of PA, $\mathcal{M} \models \mathsf{Pr}_\pi(\varphi)$ iff $\mathcal{M}' \models \varphi$ for all $\mathcal{M}'$ with $\mathcal{M}\ R\ \mathcal{M}'$, where $\mathsf{Pr}_\pi(x)$ is an intensionally correct provability predicate of PA. The existence of arithmetical accessibility relations yields a new perspective on the arithmetical completeness of GL. We show that any finite Kripke model for GL is bisimilar to some "arithmetical" Kripke model whose domain consists of models of PA and whose accessibility relation is an arithmetical accessibility relation. This yields a new interpretation of the modal operators in the context of PA: an arithmetical assertion φ is consistent (possible, $\Diamond\varphi$) if it holds in some arithmetically accessible model, and provable (necessary, $\Box\varphi$) if it holds in all arithmetically accessible models.

Keywords: Arithmetic, modal logic, provability logic, internal models

1 Introduction

The modal logic GL is K plus Löb's axiom $\Box(\Box A \to A) \to \Box A$. The intended meaning of $\Box A$ in the context of GL is: there exists a proof of A in[1] Peano Arithmetic (PA). To be more precise, $\Box A$ is interpreted as the arithmetical sentence $\mathsf{Pr}_\pi(\overline{\ulcorner A^* \urcorner})$, where $\mathsf{Pr}_\pi(x)$ is an intensionally correct provability predicate of PA, A^* is an arithmetical sentence, and $\overline{\ulcorner A^* \urcorner}$ is the term (of the language of PA) naming the code of A (under some assumed gödelnumbering).

As it turns out, GL captures exactly what is provable in PA, in propositional terms, about $\mathsf{Pr}_\pi(x)$. A modal formula A is said to be a *provability principle* (of PA) if every possible translation[2] of A into an arithmetical sentence is provable in PA. The system GL is arithmetically sound: if $\vdash_{\mathsf{GL}} A$, then A is a provability principle. The converse – the arithmetical completeness of GL – was proven by Robert Solovay [9]. If $\nvdash_{\mathsf{GL}} A$ for some modal formula A, then by modal completeness of GL there is a finite Kripke countermodel to A. Solovay shows

[1] Here, PA can be replaced by any recursively enumerable Σ_1-sound theory containing Elementary Arithmetic (EA).

[2] This notion of translation is made precise in Section 2.2 below.

that M can be simulated in PA. The simulation provides a translation $*$ such that A^* is not a theorem of PA.

In this article, we show that for any finite Kripke model M for GL, there is some "arithmetical" Kripke model $\mathfrak{M}_{\mathsf{big}}$ that is bisimilar to M. The domain of $\mathfrak{M}_{\mathsf{big}}$ consists of models of PA, and the accessibility relation R is one of the arithmetical accessibility relations introduced in Section 3. The choice of R guarantees that the truth values of modal formulas are independent (modulo a translation into sentences of arithmetic) of whether an element $\mathcal{M}$ in the domain of $\mathfrak{M}_{\mathsf{big}}$ is viewed as a node in the Kripke model $\mathfrak{M}_{\mathsf{big}}$ (with the modal forcing relation $\Vdash$), or as a model of PA (with the first–order satisfiability relation $\vDash$).

In order to appreciate the existence of arithmetical accessibility relations, we shall make a detour to set theory. In [4], Hamkins presents a (set–theoretic) forcing interpretation of modal logic. In this context, $\Diamond\varphi$ is interpreted as: "φ holds in some forcing extension", and $\Box\varphi$ as: "φ holds in all forcing extensions" (where φ is a statement in the language of set theory). Hamkins and Löwe [5] prove that if ZFC is consistent, then the principles of forcing provable in ZFC are exactly those derivable in the modal system S4.2.

Although Hamkins and Löwe say that they want to do for forceability what Solovay did for provability, there is an important difference between the two situations. Forcing is a relation between models of set theory – a ground model has some access to the truths of its forcing extension –, and hence it is natural to view it as an accessibility relation in a Kripke model. The interpretation of the modal operators $\Box$ and $\Diamond$ in set theory is thereby in tune with their usual modal logical meanings: $\Diamond\varphi$ means that φ holds in some successor (i.e. forcing extension), and $\Box\varphi$ that φ holds in all successors (forcing extensions). As a result, one can imagine the collection of all models of set theory, related by forcing, as an enormous Kripke model (where the valuation is given by first–order satisfiability).

Provability, on the other hand, is not a relation between models of PA. Note also that whereas the usual interpretation of $\Box\varphi$ in modal logic involves universal quantification (over all accessible worlds), the interpretation of $\Box\varphi$ in PA is an *existential* sentence: there exists a proof of φ. Similarly, in the context of PA, the interpretation of $\Diamond\varphi$ switches from existential to universal: φ is consistent, i.e. all proofs are not proofs of $\neg\varphi$. It is therefore natural to ask whether there is a relation R between models of PA that is the analogue of forcing in the context of ZFC. As in the set theoretic case, we would like to view the collection of all models PA, related by this relation R, as an enormous Kripke model, and we would furthermore want this new interpretation of the modalities to be in tune with the old one. The desideratum for R is then: $\mathcal{M} \vDash \mathsf{Pr}_\pi(\ulcorner\varphi\urcorner) \Leftrightarrow$ for all $\mathcal{M}'$ with $\mathcal{M}\ R\ \mathcal{M}'$ it holds that $\mathcal{M}' \vDash \varphi$. A relation R satisfying the above property is said to be an *arithmetical accessibility relation.*

The existence of arithmetical accessibility relations yields a new interpretation of the modal operators in the context of PA. The traditional meaning of $\Box\varphi$ – "there exists a proof of φ" – is equivalent to an interpretation of $\Box\varphi$ as: "φ holds in all arithmetically accessible models". Similarly, the traditional meaning of $\Diamond\varphi$ – "φ is consistent" – is equivalent to an interpretation of $\Diamond\varphi$ as: "φ holds in some arithmetically accessible model".

The next section contains the preliminaries. Our examples of arithmetical accessibility relations are introduced in Section 3, and Section 4 establishes that any finite Kripke model for GL is bisimilar to some "arithmetical" Kripke model whose nodes are models of PA. The arithmetical completeness of GL is an easy consequence[3] of this. Finally, in Section 5 we will see that the structural properties of the big arithmetical Kripke frame depend on the exact choice of the arithmetical accessibility relation.

2 Preliminaries

This section sketches the basic notions and results needed for our main result. Section 2.1 contains the arithmetical preliminaries, Section 2.2 deals with modal logic, in particular the system GL, and Section 2.3 introduces the notion of an internal model.

2.1 Arithmetic

We work in a first–order language with $\neg, \rightarrow$ and $\forall$ as primitive connectives; the connectives $\wedge$, $\vee$, $\leftrightarrow$ and $\exists$ are assumed to be defined from the primitive connectives in the usual way. We assume a Hilbert–style axiomatization of first–order logic, with modus ponens as the only rule of inference.

Our official signature Σ of arithmetic is relational; it includes:

- a binary relation symbol E (equality[4])
- a unary relation symbol Z (being equal to zero)
- a binary relation symbol S ($\mathsf{S}xy$ being interpreted as: $x + 1 = y$)
- a ternary relation symbol A ($\mathsf{A}xyz$ being interpreted as: $x + y = z$)
- a ternary relation symbol M ($\mathsf{M}xyz$ being interpreted as: $x \times y = z$)

We also use Σ to refer to the language of arithmetic, i.e. the first–order language based on the signature of arithmetic. We shall use lower case Greek letters for the sentences and formulas of Σ.

The theory of Peano Arithmetic (PA) is a first–order theory in the language of arithmetic. It contains axioms stating that E is a congruence relation (since equality is treated as a non–logical symbol), the basic facts about the relations Z, S, A and M (for example that $\mathsf{Z}x$ implies $\neg\mathsf{S}yx$ for all y), as well as a functionality axiom – with respect to E – for each of the above relations (for Z, this amounts

[3] Our proof of Solovay's Theorem is not "new" – the construction of the bisimulation makes crucial use of the most important ingredients of the original proof. Solovay's Theorem will thus remain among the important theorems in mathematical logic which have "essentially" only one proof (see [3]).

[4] Note that we treat equality as a non–logical symbol. This makes it straightforward to allow equality to be translated by something else than equality when defining the notion of a relative translation (Definition 6).

to the uniqueness of the element in its extension). Finally, the axioms of PA include induction for all formulas in the language Σ of arithmetic.

All first–order models considered in this article are models of PA, or expansions of models of PA. From now on, the word "model" will refer to such a model. The domains of the arithmetical Kripke frames constructed in Section 4 consist of models whose signature contains only predicates included in the signature of arithmetic. We shall therefore sometimes also refer to such models as "worlds". If $\mathcal{M}$ is a model, we use the symbol $\Sigma_{\mathcal{M}}$ to refer to the signature of $\mathcal{M}$. A model $\mathcal{M}$ is said to be *inductive* if it satisfies the induction axioms in the language $\Sigma_{\mathcal{M}}$. If $\Sigma_{\mathcal{M}} = \Sigma_{\mathcal{M}'}$, we write $\mathcal{M} \equiv \mathcal{M}'$ to mean that $\mathcal{M}$ and $\mathcal{M}'$ are elementarily equivalent, i.e. that for every sentence φ of $\Sigma_{\mathcal{M}}$, $\mathcal{M} \vDash \varphi \Leftrightarrow \mathcal{M}' \vDash \varphi$. We write $\mathcal{M} \cong \mathcal{M}'$ if $\mathcal{M}$ and $\mathcal{M}'$ are isomorphic. If φ is a formula whose free variables are among $x_0, \ldots x_n$, and $m_0, \ldots m_k$ is a sequence of elements of $\mathcal{M}$, with $n \leq k$, we write $\mathcal{M} \vDash \varphi[m_1, \ldots, m_k]$ to mean that $\mathcal{M}$ satisfies φ when x_j is interpreted as a name for m_j.

In practice, we shall often speak of the formulas of Σ as containing terms built up from the constant symbol $\mathsf{0}$, a unary function symbol S, and binary function symbols $+$ and $\times$. Such formulas can be transformed into proper formulas of Σ by the term–unwinding algorithm[5]. We define for each natural number n a term $\overline{n}$ of our unofficial language by letting $\overline{0} = 0$, and $\overline{n+1} = \mathsf{S}\overline{n}$. We shall often also write n instead of $\overline{n}$, and $x = y$ instead of $\mathsf{E}xy$.

We assume as given some standard gödelnumbering of the syntactical objects of Σ. If φ is a formula, we write $\ulcorner \varphi \urcorner$ for the code of φ, and similarly for terms. We shall often identify a formula with its code, thus writing for example $\varphi(\psi)$ instead of $\varphi(\overline{\ulcorner \psi \urcorner})$.

We also assume as given formulas of Σ that express relations between syntactical objects (of Σ and its expansions) and operations on them. We will use self–explanatory notation for such formulas; for example, the formula form expresses the property of being (the code of) a formula, and the formula var the property of being (the code of) a variable. We write $\forall \varphi \in \mathsf{form}\, \alpha(\varphi)$ instead of $\forall x\, (\mathsf{form}(x) \to \alpha(x))$, and similarly for other syntactical objects such as variables or sentences.

We shall assume that all formulas representing relations between and operations on syntactical objects do so in an *intensionally correct* way. By this,

[5] For example, the formula $x + \mathsf{S}y = \mathsf{S}(x + y)$ becomes

$$\exists z, u, w\, (\mathsf{S}yz \wedge \mathsf{A}xzw \wedge \mathsf{A}xyu \wedge \mathsf{S}uw)$$

or, equivalently (by the functionality axioms),

$$\forall z, u, w\, (\mathsf{S}yz \wedge \mathsf{A}xzw \wedge \mathsf{A}xyu \to \mathsf{S}uw) \ .$$

The process of unravelling only adds a block of existential (universal) quantifiers in front of a formula. Given that the complexity of a formula is measured e.g. by depth of quantifier alternations, it will therefore not increase as a result of applying the term–unwinding algorithm (as long as the original formula contains at least one quantifier).

we mean that the definitions of these relations and operations in PA mimic the corresponding "informal" recursive definitions, and that the relevant recursion equations are provable in PA. For example, we assume that there is a formula sbst_1 of Σ such that

$$\vdash_{\mathsf{PA}} \forall t \in \mathsf{term}\ (\mathsf{sbst}_1(t, \mathsf{Z}v_0) = \mathsf{Z}t)\ , \tag{1}$$

and similarly for other atomic formulas. Furthermore,

$$\vdash_{\mathsf{PA}} \forall t \in \mathsf{term}, \forall \varphi, \psi \in \mathsf{form}\ (\mathsf{sbst}_1(t, \varphi \wedge \psi) = \mathsf{conj}(\mathsf{sbst}_1(t, \varphi), \mathsf{sbst}_1(t, \psi)))\ , \tag{2}$$

where $\mathsf{conj}(\varphi, \psi)$ is an intensionally correct representation of the function computing the gödelnumber of the conjunction of the formulas coded by its inputs. Equations like (2) are also assumed to be provable for the other connectives and for the quantifiers.

Given that the formula sbst_1 is intensionally correct in the above sense, it is also *extensionally correct*, i.e. it has the right extension in the standard model, or in other words it behaves as intended with respect to the codes of standard sentences and terms. This means that for any formula φ and for any term t,

$$\vdash_{\mathsf{PA}} \mathsf{sbst}_1(\overline{\ulcorner t \urcorner}, \overline{\ulcorner \varphi \urcorner}) = \overline{\mathrm{Sbst}_1(\ulcorner t \urcorner, \ulcorner \varphi \urcorner)}\ , \tag{3}$$

where Sbst_1 is the primitive recursive function with:

$$\mathrm{Sbst}_1(m, n) = \begin{cases} \ulcorner [t/v_0]\varphi \urcorner \text{ if } n = \ulcorner \varphi \urcorner,\ m = \ulcorner t \urcorner, \text{ and } t \text{ is free for } v_0 \text{ in } \varphi \\ 0 \text{ otherwise} \end{cases}$$

In general, if a formula expresses a property in an intensionally correct way, then it also expresses this property in an extensionally correct way[6]. On the other hand, there exist formulas which are extensionally correct with respect to a property but fail to express this property in an intensionally correct way (for an example, see p.68 of [2]).

Throughout this article, λ denotes a formula that expresses the property of being an axiom of first–order logic, and π denotes a formula that expresses the property of being an axiom of PA. Both λ and π can be taken to be Δ_1. Given λ and π, the proof predicate Prf_π of PA is constructed in the usual way, and is thus an intensionally correct representation of the relation

$$\{(n, p) \mid p \text{ codes a } \mathsf{PA}\text{–proof of the formula with gödelnumber } n\}\ . \tag{4}$$

[6] The concept of extensional correctness is found in the literature under various names, for example *binumerability* [2], [7] or *representability* [1]. A formula φ of the signature Σ is said to binumerate a relation R (on the natural numbers) in PA if for all n, Rn implies $\vdash_{\mathsf{PA}} \varphi(n)$, and *not* Rn implies $\vdash_{\mathsf{PA}} \neg\varphi(n)$. A weaker notion of extensional correctness is expressed by the notion of *numerability*, or *weak representability*, where a formula φ is said to numerate a relation R in PA in case Rn holds if and only if $\vdash_{\mathsf{PA}} \varphi(n)$.

The provability predicate Pr_π is obtained by letting: $\mathsf{Pr}_\pi(x) := \exists y \, \mathsf{Prf}_\pi(x, y)$. We shall often omit the subscript, writing simply Pr for Pr_π. We write $\mathsf{Con}(\varphi)$ for the sentence $\neg\mathsf{Pr}(\neg\varphi)$, and Con for $\mathsf{Con}(\top)$.

Under the assumption that PA is Σ_1–sound, the formula $\mathsf{Pr}(x)$ expresses the property of being a theorem of PA in an extensionally correct way, i.e. for any formula φ of the language Σ it holds that[7]

$$\vdash_{\mathsf{PA}} \varphi \;\Leftrightarrow\; \vdash_{\mathsf{PA}} \mathsf{Pr}(\varphi) \ . \tag{5}$$

By intensional correctness of the proof predicate, we have furthermore

$$\vdash_{\mathsf{PA}} \forall \varphi, \psi \in \mathsf{form} \; (\mathsf{Pr}(\mathsf{impl}(\varphi, \psi)) \to (\mathsf{Pr}(\varphi) \to \mathsf{Pr}(\psi))) \ , \tag{6}$$

where impl is an intensionally correct representation of the function computing the gödelnumber of $\varphi \to \psi$, given as input the gödelnumbers of φ and ψ. Finally it can be shown that

$$\vdash_{\mathsf{PA}} \forall \varphi \in \mathsf{form} \; (\mathsf{Pr}(\varphi) \to \mathsf{Pr}(\mathsf{Pr}(\varphi))) \ , \tag{7}$$

The externally quantified versions[8] of items (6) and (7) together with the left to right direction of (5) are often referred to as the *Hilbert–Bernays–Löb derivability conditions.*

The following theorem (Theorem 4.6.v in [2]) states that inside PA, properties of theorems of PA can be proven by induction on the complexity of PA–proofs.

Theorem 1. *Let α be a formula of Σ. Then*

$$\begin{aligned} \vdash_{\mathsf{PA}} \forall \varphi \in \mathsf{form} \, [(\pi(\varphi) \to \alpha(\varphi)) \wedge (\lambda(\varphi) \to \alpha(\varphi))] \\ \wedge \forall \varphi, \psi \in \mathsf{form} \; [\alpha(\varphi) \wedge \alpha(\mathsf{impl}(\varphi, \psi)) \to \alpha(\psi)] \\ \to \; \forall \varphi \in \mathsf{form} \, (\mathsf{Pr}(\varphi) \to \alpha(\varphi)). \end{aligned}$$

Remark 1. The proof uses the induction axiom for the formula α. When working in an expansion $\mathcal{M}$ of a model of PA, Theorem 1 can be applied with a formula α of the signature $\Sigma_{\mathcal{M}}$ given that induction holds in $\mathcal{M}$ for α.

2.2 Modal Logic

We denote by $\mathcal{L}_\Box$ the language of propositional modal logic. We use upper case Latin letters for the formulas of $\mathcal{L}_\Box$.

Definition 1 (Kripke model). *A Kripke frame is a tuple $\langle W, R \rangle$, where $W \neq \emptyset$ and $R \subseteq W \times W$. A Kripke model is a triple $M = \langle F, \Vdash \rangle$, where F is a Kripke frame, and $\vDash$ is a forcing relation on W satisfying the usual clauses for the connectives, and for all $w \in W$,*

$$w \Vdash \Box\varphi \;:\Leftrightarrow\; y \Vdash \varphi \textit{ for all } y \textit{ with } wRy \ . \tag{8}$$

[7] The assumption of Σ_1–soundness is needed for the direction from right to left.
[8] I.e. versions where the universal quantifiers range only over *standard* sentences.

If $F = \langle W, R \rangle$ is a frame, we write $F \Vdash A$ in case for all models $\langle F, \Vdash \rangle$ and for all $w \in W$, $w \Vdash A$.

When do two nodes in two Kripke models satisfy the same modal formulas? A sufficent structural condition is given by the notion of bisimulation.

Definition 2 (Bismulation). *Let $M = \langle W, R, V \rangle$ and $M' = \langle W', R', V' \rangle$ be two Kripke models. A binary relation $Z \subseteq W \times W'$ is a bisimulation between M and M' if the following conditions are satisfied:*

- (at) *If wZw', then w and w' satisfy the same propositional letters*
- (back) *If wZw' and wRv, then there exists v' in M' with $w'R'v'$ and vZv'*
- (forth) *If wZw' and $w'R'v'$, then there exists v in M with wRv and vZv'*

When Z is a bisimulation between M and M', and wZw', we say that w and w' are bisimilar. The following theorem states that modal satisfiability is indeed invariant under bisimulations.

Theorem 2. *Let M and M' be Kripke models, and let w, w' be nodes of M and M' respectively. If w is bisimilar to w', then for every modal formula A, we have that $M, w \Vdash A$ iff $M', w' \Vdash A$.*

We now introduce the modal system GL, named after Gödel and Löb, and also known as provability logic.

Definition 3 (GL). *The axioms of* GL *are all tautologies of propositional logic, and*

1. $\Box(A \to B) \to (\Box A \to \Box B)$
2. $\Box(\Box A \to A) \to \Box A$.

The rules of GL *are modus ponens, and necessitation:* $\vdash_{\mathsf{GL}} A \Rightarrow \vdash_{\mathsf{GL}} \Box A$.

In other words, GL is K plus Löb's axiom $\Box(\Box A \to A) \to \Box A$. GL is known to be sound and complete with respect to transitive irreflexive finite trees.

Theorem 3 (Modal Completeness of GL). *Let $\mathcal{K}$ be the class of frames that are transitive irreflexive finite trees. Then* $\vdash_{\mathsf{GL}} A \Leftrightarrow \forall F\, [F \in \mathcal{K} \Rightarrow F \Vdash A]$.

The notion of an arithmetical realization below will be used to translate formulas of $\mathcal{L}_\Box$ to sentences of the language of arithmetic.

Definition 4 (Arithmetical realization). *A realization $*$ is a function from the propositional letters of $\mathcal{L}_\Box$ to sentences of Σ. The domain of $*$ is extended to all $\mathcal{L}_\Box$–formulas by requiring:*

1. $(\bot)^* = \bot$
2. $(A \to B)^* = A^* \to B^*$
3. $(\Box A)^* := \mathsf{Pr}(\ulcorner A^* \urcorner)$

Definition 5 (Provability logic). *A modal formula A is a provability principle of* PA *if for all realizations $*$,* $\vdash_{\mathsf{PA}} A^*$. *The provability logic of* PA, (PrL(PA)), *is the set of all provability principles of* PA, *or a logic that generates it.*

The following theorem states that GL is arithmetically sound and complete, i.e. it is the provability logic of PA.

Theorem 4. $\mathsf{PrL}(\mathsf{PA}) = \mathsf{GL}$.

Proof sketch. For the direction $\mathsf{GL} \subseteq \mathrm{PrL}(\mathsf{PA})$ (arithmetical soundness), one has to check that the axioms of GL are provable in PA under all realizations. This follows from the Hilbert–Bernays–Löb derivability conditions introduced in Section 2.1 (to see that Löb's Theorem is provable under all realizations, one also uses the Gödel–Carnap Fixed Point Lemma).

The proof of $\mathrm{PrL}(\mathsf{PA}) \subseteq \mathsf{GL}$ (arithmetical completeness) is due to Robert Solovay [9]. Given a modal formula A with $\nvdash_{\mathsf{GL}} A$, we need a realization $*$ with $\nvdash_{\mathsf{PA}} A^*$. The idea of Solovay's proof is to simulate in PA a Kripke model $M = \langle \{1, \ldots, n\}, R, V \rangle$ for GL, with $M, 1 \nVdash A$ (note that M exists by Theorem 3). This is done by constructing sentences $\sigma_0, \ldots, \sigma_n$ of the language Σ of arithmetic such that, intuitively, σ_i corresponds to the node i of M (the sentence σ_0 is used as an auxilliary). We will refer to the sentences $\sigma_0 \ldots, \sigma_n$ as the *Solovay sentences*. The arithmetical realization $*$ is defined as: $p^* := \bigvee_{i:M,i \Vdash p} \sigma_i$. M is then simulated in PA in the sense that for all $A \in \mathcal{L}_\Box$,

$$M, i \Vdash A \;\Rightarrow\; \vdash_{\mathsf{PA}} \sigma_i \to A^* \tag{9}$$

$$M, i \Vdash \neg A \;\Rightarrow\; \vdash_{\mathsf{PA}} \sigma_i \to \neg A^* \tag{10}$$

The proof of (9) and (10) uses the following properties of the Solovay sentences:

1. $\vdash_{\mathsf{PA}} \sigma_i \to \neg\sigma_j$ if $i \neq j$
2. $\vdash_{\mathsf{PA}} \sigma_i \to \mathsf{Con}(\sigma_j)$ if iRj, or if $i = 0$ and $j = 1$
3. $\vdash_{\mathsf{PA}} \sigma_i \to \mathsf{Pr}(\bigvee_{j:iRj} \sigma_j)$ for $i \geq 1$

Furthermore, we have that for all $0 \leq i \leq n$, the sentence σ_i is independent from PA. We shall use the above properties in Section 4 to prove that any finite GL–model is bisimilar to some arithmetical Kripke model, obtaining the arithmetical completeness of GL as a corollary.

2.3 Internal Models

This section introduces the notion of an internal model. Let $\mathcal{M}$ and $\mathcal{M}'$ be models of PA. Roughly speaking, $\mathcal{M}'$ is an *internal model* of $\mathcal{M}$ (or: internal to $\mathcal{M}$) if the domain of $\mathcal{M}'$ is definable in $\mathcal{M}$, and the interpretations of the atomic formulas of $\Sigma_{\mathcal{M}'}$ are given by formulas of $\Sigma_{\mathcal{M}}$.

Definition 6 (Relative translation). *Let Θ be a signature. A relative translation from Σ to Θ is a tuple $\langle \delta, \tau \rangle$, where δ is a Θ–formula with one free variable, and τ a mapping from relation symbols R of Σ to formulas R^τ of Θ. We require the number of free variables in R^τ to be equal to the arity of R. The domain of τ is extended to all formulas of Σ by letting:*

1. $(Rx_0 \dots x_n)^\tau = R^\tau x_0 \dots x_n$ *for an* $n+1$*–ary relation symbol* R
2. $(\varphi \to \psi)^\tau = \varphi^\tau \to \psi^\tau$
3. $\bot^\tau = \bot$
4. $(\forall x\, \varphi)^\tau = \forall x\, (\delta(x) \to \varphi^\tau)$

If j is a relative translation, we refer to the components of j by δ_j and τ_j. Note that although Σ in the above definition can taken to be any signature, we shall only consider cases where the language of the internal model is indeed the language of arithmetic.

Definition 7 (Internal model). *Let* $\mathcal{M}$ *be a model, and let* $j = \langle \delta, \tau \rangle$ *be a relative translation from* Σ *to* $\Sigma_{\mathcal{M}}$*. We say that* j defines an internal model of $\mathcal{M}$ *if* $\mathcal{M} \vDash \varphi^{\tau_j}$ *for every axiom* φ *of* PA*.*

If a relative translation j *defines an internal model of* $\mathcal{M}$*, we denote by* $\mathcal{M}^j$ *the following model:*

- $\mathrm{dom}(\mathcal{M}^j) := \{a \in \mathrm{dom}(\mathcal{M}) \mid \mathcal{M} \vDash \delta_j(a)\} \,/ \sim$,
  where $a \sim b :\Leftrightarrow \mathcal{M} \vDash (\mathsf{E}ab)^{\tau_j}$
- *If* $\mathbf{a}, \mathbf{b} \in \mathrm{dom}(\mathcal{M}^j)$*, let* $\mathcal{M}^j \vDash \mathsf{S}\mathbf{ab} :\Leftrightarrow \mathcal{M} \vDash (\mathsf{S}ab)^{\tau_j}$ *for some* $a \in \mathbf{a}$*,* $b \in \mathbf{b}$*, and similarly for other atomic formulas*

Note that if j defines an internal model of $\mathcal{M}$ then – by our choice of axioms of PA – we have that in $\mathcal{M}$, E^τ is a congruence relation, and the relations defined by the formulas Z^τ, S^τ, A^τ, and M^τ are functional relative to E^τ.

We say that $\mathcal{M}'$ is an internal model of $\mathcal{M}$, and write $\mathcal{M} \rhd \mathcal{M}'$, if some relative translation j to $\Sigma_{\mathcal{M}}$ defines an internal model of $\mathcal{M}$, and $\mathcal{M}'$ is (modulo isomorphism) this internal model, i.e. $\mathcal{M}' \cong \mathcal{M}^j$. In this context, we shall often refer to $\mathcal{M}$ as the *external model*. The following theorem is a well–known basic fact about the internal model relation.

Theorem 5. *Let* $\mathcal{M}$ *be a model, and suppose that a relative translation* j *from* Σ *to* $\Sigma_{\mathcal{M}}$ *defines an internal model of* $\mathcal{M}$*. Then for any formula* $\varphi(x_0, \dots, x_n)$ *of the language* Σ *of arithmetic,*

$$\mathcal{M}^j \vDash \varphi[\mathbf{a_0}, \dots, \mathbf{a_n}] \Leftrightarrow \mathcal{M} \vDash \varphi^{\tau_j}[a_0, \dots, a_n] \text{ for } a_0 \in \mathbf{a_0}, \dots, a_n \in \mathbf{a_n} \ . \tag{11}$$

In particular, $\mathcal{M}^j \vDash \varphi \Leftrightarrow \mathcal{M} \vDash \varphi^{\tau_j}$ *for any sentence* φ *of* Σ*.*

Proof. By induction on the complexity of $\varphi(x_0, \dots, x_n)$.

Note that if j defines an internal model of $\mathcal{M}$ (as in Definition 7,) then it follows by Theorem 5 that $\mathcal{M}^j \vDash \varphi$ for every axiom φ of PA. Hence $\mathcal{M} \rhd \mathcal{M}'$ implies that $\mathcal{M}'$ is a model of PA.

3 t–internal and t–associated models

We introduce three arithmetical accessibility relations between models of PA, all of which are based on the $\mathtt{t}$*–internal model relation.* Roughly speaking, an internal model $\mathcal{M}'$ of $\mathcal{M}$ is a t*–internal model* of $\mathcal{M}$ if $\mathcal{M}$ has a truth–predicate for sentences of $\Sigma_{\mathcal{M}'}$, and furthermore the axioms of PA (including the nonstandard ones) are in the extension of the truth predicate. We shall generalize this relation by allowing the truth predicate to be definable in $\mathcal{M}$ with parameters, or definable in some inductive expansion of $\mathcal{M}$. As before, we shall only consider cases where the language of the internal model is the language Σ of arithmetic.

3.1 Definitions

Before defining the t–internal model relation, it is useful to mention a technicality. Suppose that j defines an internal model of $\mathcal{M}$. As suggested above, we want to express that $\mathcal{M}$ has a truth predicate tr for the internal model $\mathcal{M}^j$. This means in particular that for any sentence φ of the language Σ (remember that the signature of $\mathcal{M}^j$ is assumed to be Σ), $\mathcal{M} \vDash \varphi^{\tau_j} \leftrightarrow \mathsf{tr}(\varphi)$. We want to arrive at this statement by an induction on the complexity of φ. In order to express that tr behaves as expected with respect to the atomic formulas, we would like to say (in $\mathcal{M}$), for example, that whenever $m \in \mathcal{M}^j$, then $(\mathsf{Z}m)^\tau$ if and only if the code of $\mathsf{Z}m$ is in the extension of tr. However, a truth predicate should apply to gödelnumbers of *sentences*. Hence when inside the truth predicate, we want to associate to m (the code of some) constant naming it.

A simple way to achieve this is to stipulate that m is the code of its own name, and in general that elements of $\mathcal{M}^j$ are codes of their own names. Intuitively, this means that inside $\mathcal{M}$ we will be working with the language $\Sigma \cup \{c_m \mid m \in \mathcal{M}^j\}$, where for all $m \in \mathcal{M}^j$, $\ulcorner c_m \urcorner = m$. We assume our gödelnumbering to be done in such a way that this does not lead to ambiguities[9], for example the domain of $\mathcal{M}^j$ (as given by δ_j) should be disjoint from the set of codes of terms of Σ (in $\mathcal{M}$).[10]

The formulas representing properties of the syntactical objects of the language $\Sigma \cup \{c_m \mid m \in \mathcal{M}^j\}$ in $\mathcal{M}$ are distinguished by using the subscript δ_j (since the domain of $\mathcal{M}^j$ is given by the domain formula δ_j), and similarly we refer to the language $\Sigma \cup \{c_m \mid m \in \mathcal{M}^j\}$ as Σ_{δ_j}. Thus for example the formula sent_{δ_j} expresses the property of being a sentence of the language $\Sigma_{\delta_j} = \Sigma \cup \{c_m \mid m \in \mathcal{M}^j\}$.

Let $\mathcal{M}$ be a model, and let $j = \langle \delta, \tau \rangle$ be a relative translation to $\Sigma_{\mathcal{M}}$. Fix in $\mathcal{M}$ (the code of) a (possibly nonstandard) formula φ of the language Σ_δ. We write $\mathsf{fv}(\varphi)$ for the set of free variables of φ (according to $\mathcal{M}$). Let $\mathsf{tr}(x)$ be a (standard)

[9] To see how ambiguities could arise in principle, suppose (in $\mathcal{M}$) that 17 is in the extension of δ_j (i.e. 17 is in the domain of the internal model), but also that 17 is the code of the constant 0. Suppose also that we use sequences to code syntactical objects; for example the code of the sentence $\mathsf{Z}c$ (where c is a constant) is the number $\langle \ulcorner \mathsf{Z} \urcorner, \ulcorner c \urcorner \rangle$ (where $\langle m, n \rangle$ is the code of the pair (m, n)). Then, given the number $\langle \mathsf{Z}, 17 \rangle$, it is not clear whether it should be parsed as coding the sentence $\mathsf{Z}0$ (of the language Σ), or instead the sentence $\mathsf{Z}c_{17}$ (of the language $\Sigma \cup \{c_m \mid m \in \mathcal{M}^j\}$).

[10] Note that this requirement excludes the possibility that $\mathcal{M}$ and $\mathcal{M}^j$ share the same domain. However since we think of models modulo isomorphism, we can still allow the internal model relation to be e.g. reflexive.

formula with one free variable. If $\mathsf{fv}(\varphi) = \{v_0\}$, we write $\mathsf{tr}(\varphi(x))$ as shorthand for $\mathsf{tr}\,(\mathsf{sbst}_1(x, \varphi(v_0)))$, thus e.g. $\mathsf{tr}(\mathsf{Z}x)$ stands for $\mathsf{tr}(\mathsf{sbst}_1(x, \overline{\ulcorner \mathsf{Z}v_0 \urcorner}))$. If x is in the extension of δ, it is the code of the constant c_x, and hence $\mathsf{sbst}_1(x, \overline{\ulcorner \mathsf{Z}v_0 \urcorner})$ is the code of the Σ_δ*–sentence* $\mathsf{Z}c_x$. More generally, if φ is a formula with n free variables (where n is standard), we write $\mathsf{tr}(\varphi(x_0, \ldots, x_{n-1}))$ as shorthand for

$$\mathsf{tr}(\mathsf{sbst_n}(x_0, \ldots, x_{n-1}, \overline{\ulcorner \varphi(v_0, \ldots, v_{n-1}) \urcorner})) \ , \tag{12}$$

where $\mathsf{sbst_n}$ is an intensionally correct representation of the function corresponding to the simultaneous substitution of n terms in a formula.

Definition 8 (t–internal model). *Let $\mathcal{M}$ be a model, and $j = \langle \delta, \tau \rangle$ a relative translation from Σ to $\Sigma_{\mathcal{M}}$. We say that j* defines a t–internal model of $\mathcal{M}$ *if there is a formula* tr *of the signature $\Sigma_{\mathcal{M}}$ with one free variable, such that the following sentences are satisfied in $\mathcal{M}$:*

1. $\forall x, y\,(\delta(x) \wedge \delta(y) \to ((\mathsf{S}xy)^\tau \leftrightarrow \mathsf{tr}(\mathsf{S}xy)))$, *similarly for other atomic formulas*
2. $\forall \varphi \in \mathsf{sent}_\delta, \forall \psi \in \mathsf{sent}_\delta\,(\mathsf{tr}(\varphi \to \psi) \leftrightarrow (\mathsf{tr}(\varphi) \to \mathsf{tr}(\psi)))$
3. $\forall \varphi \in \mathsf{sent}_\delta\,(\mathsf{tr}(\neg\varphi) \leftrightarrow \neg\mathsf{tr}(\varphi))$
4. $\forall \varphi \in \mathsf{form}_\delta, \forall u \in \mathsf{var}\,(\mathsf{fv}(\varphi) \subseteq \{u\} \to (\mathsf{tr}\,(\forall u\, \varphi) \leftrightarrow \forall x\,(\delta(x) \to \mathsf{tr}(\varphi(x)))))$
5. $\forall \varphi \in \mathsf{sent}\,(\pi(\varphi) \to \mathsf{tr}(\varphi))$

We refer to the formula tr as the *truth predicate* (for the internal model), and write tr_j for the truth predicate that comes with a relative translation j as in Definition 8. The next theorem states that the formula tr_j is indeed a well–behaved truth predicate – modulo the translation τ_j – for the language Σ.

Theorem 6. *Suppose that j defines a* t*–internal model of $\mathcal{M}$, and let φ be a formula of Σ whose free variables are among $x_0, \ldots, x_n$. Then the following sentence*[11] *is satisfied in $\mathcal{M}$:*

$$\delta_j(x_0) \wedge \ldots \wedge \delta_j(x_n) \to (\varphi^{\tau_j}\,(x_0, \ldots, x_n) \leftrightarrow \mathsf{tr}_j(\varphi(x_0, \ldots, x_n))) \ . \tag{13}$$

In particular, $\mathcal{M} \models \varphi^{\tau_j} \leftrightarrow \mathsf{tr}_j(\varphi)$ for any sentence φ of Σ.

Proof. For readability, we shall drop the subscript j from δ, τ, and tr. The proof is by (external) induction on the complexity of φ. The base cases hold by Definition 8. The inductive cases for $\to$ and $\neg$ follow easily by using that tr and τ commute with the propositional connectives. We treat the universal case, assuming $n = 0$ for simplicity. Let $\varphi(x)$ be the formula $\forall y\, \psi(x, y)$. Argue in $\mathcal{M}$:

$$\begin{aligned} & \delta(x) \wedge \delta(y) \to (\psi(x,y)^\tau \leftrightarrow \mathsf{tr}(\psi(x,y))) \\ \to\ & \delta(x) \to (\forall y\,(\delta(y) \to \psi(x,y)^\tau) \leftrightarrow \forall y\,(\delta(y) \to \mathsf{tr}(\psi(x,y)))) \\ \leftrightarrow\ & \delta(x) \to ((\forall y\, \psi(x,y))^\tau \leftrightarrow \mathsf{tr}(\forall y\, \psi(x,y))) \ , \end{aligned}$$

where the first line is the induction assumption, the second follows by logic, and the third line follows by the properties of τ and tr.

[11] The free variables are assumed to be bound by universal quantifiers.

Suppose that j defines a t–internal model of $\mathcal{M}$. By Theorem 6 and item 5 of Definition 8, we have that $\mathcal{M} \models \varphi^{\tau_j}$ for every axiom φ of PA. Hence, in particular j defines an internal model of $\mathcal{M}$. We say that $\mathcal{M}'$ is a t–internal model of $\mathcal{M}$, and write $\mathcal{M} \rhd_{\mathsf{t}} \mathcal{M}'$, if some relative translation j defines a t–internal model of $\mathcal{M}$, and $\mathcal{M}' \cong \mathcal{M}^j$ (where $\mathcal{M}^j$ is as in Definition 7).

Allowing the truth predicate to contain parameters from the external model yields the notion of a t–internal model with parameters.

Definition 9 (t–internal model with parameters). *Let $\mathcal{M}$ be a model, and suppose that $j = \langle \delta, \tau \rangle$ is a relative translation to $\Sigma_{\mathcal{M}}$. We say that j* defines a t–internal model of $\mathcal{M}$ with parameters *if there is a formula tr of the signature $\Sigma_{\mathcal{M}}$ with two free variables, and some*[12] *$m \in \mathsf{dom}(\mathcal{M})$ such that items 1–5 of Definition 8 are satisfied in $\mathcal{M}$, if the remaining free variable in tr is interpreted as a name*[13] *for m.*

It is clear that an analogue of Theorem 6 holds for the case where j defines a t–internal model of $\mathcal{M}$ with parameters. Hence if j defines a t–internal model of $\mathcal{M}$ with parameters, it also defines an internal model of $\mathcal{M}$ (note that if φ is an axiom of PA, then φ and also φ^{τ_j} are sentences, whence $\mathcal{M} \models \varphi^{\tau_j}[m]$ iff $\mathcal{M} \models \varphi^{\tau_j}$). We say that $\mathcal{M}'$ is a t–internal model of $\mathcal{M}$ with parameters, and write $\mathcal{M} \rhd_{\mathsf{tpar}} \mathcal{M}'$, if some relative translation j defines a t–internal model of $\mathcal{M}$ with parameters, and $\mathcal{M}' \cong \mathcal{M}^j$.

Finally, allowing the truth predicate to be non–definable in the language of the external model yields the notion of a t–associated model.

Definition 10 (t–associated model). *Let $\mathcal{M}$ be a model, and $j = \langle \delta, \tau \rangle$ a relative translation from Σ to $\Sigma_{\mathcal{M}}$. We say that j* defines a t–associated model of $\mathcal{M}$ *if there is some inductive expansion $\mathcal{M}^+$ of $\mathcal{M}$, and some formula tr of $\Sigma_{\mathcal{M}^+}$ with one free variable such that items 1–5 of Definition 8 hold in $\mathcal{M}^+$.*

Note that if j defines a t–associated model of $\mathcal{M}$ and $\mathcal{M}^+$ is the inductive expansion of $\mathcal{M}$ as in Definition 10, then j defines a t–internal model of $\mathcal{M}^+$. Hence by Theorem 6, $\mathcal{M}^+ \models \varphi^{\tau_j} \leftrightarrow \mathsf{tr}_j(\varphi)$ for any sentence φ of Σ. Furthermore since for all φ, φ^{τ_j} is a sentence of $\Sigma_{\mathcal{M}}$, we have that $\mathcal{M}^+ \models \varphi^{\tau_j}$ iff $\mathcal{M} \models \varphi^{\tau_j}$. Hence by item 5 of Definition 8, we have that $\mathcal{M} \models \varphi^{\tau_j}$ for every axiom φ of PA, and thus in particular j defines an internal model of $\mathcal{M}$. We say that $\mathcal{M}'$ is a t–associated model of $\mathcal{M}$, and write $\mathcal{M} \blacktriangleright_{\mathsf{t}} \mathcal{M}'$, if some relative translation j defines a t–associated model of $\mathcal{M}$, and $\mathcal{M}' \cong \mathcal{M}^j$.

Remark 2. If $\mathcal{M}'$ is a t–associated model of $\mathcal{M}$, then the interpretations of the atomic formulas of $\Sigma_{\mathcal{M}'}$ are given by formulas of $\Sigma_{\mathcal{M}}$, and some inductive expansion $\mathcal{M}^+$ of $\mathcal{M}$ has a truth predicate for $\mathcal{M}'$. For the purposes of this article (in particular for proving Theorem 9 below), we could also have chosen a more general definition, where the interpretations of the atomic formulas of $\Sigma_{\mathcal{M}'}$ are only

[12] Due to the availability of coding, allowing one parameter is as strong as allowing an arbitrary finite number of parameters.

[13] For example for item 5 we require that $\mathcal{M} \models \forall \varphi \in \mathsf{sent}\,(\pi(\varphi) \to \mathsf{tr}(\varphi, y))[m]$.

required to be given by formulas of $\mathcal{M}^+$. Equivalently, we could postulate that the interpretations of the atomic formulas are only defined by the truth predicate in the first place. For example, Z^τ would then be the formula $\mathsf{tr}(\mathsf{sbst}_1(x, \overline{\ulcorner \mathsf{Z}\mathsf{v}_0 \urcorner}))$.

Remark 3. It is easy to see that $\mathcal{M} \rhd_{\mathsf{t}} \mathcal{M}'$ implies $\mathcal{M} \rhd_{\mathsf{tpar}} \mathcal{M}'$, and $\mathcal{M} \rhd_{\mathsf{tpar}} \mathcal{M}'$ implies $\mathcal{M} \blacktriangleright_{\mathsf{t}} \mathcal{M}'$. As we will see in Section 5, the reverse implications fail.

3.2 Arithmetical Accessibility Relations

In this section, we shall use the term "worlds" to refer to models of PA whose signature contains only predicates of the signature Σ of arithmetic. A relation R between worlds is said to be an *arithmetical accessibility relation* if for any world $\mathcal{M}$, $\mathcal{M} \vDash \mathsf{Pr}_\pi(\varphi)$ iff $\mathcal{M}' \vDash \varphi$ for all $\mathcal{M}'$ with $\mathcal{M}\ R\ \mathcal{M}'$. We shall use theorems 7 and 8 below to show that each of the relations $\rhd_{\mathsf{t}}$, $\rhd_{\mathsf{tpar}}$ and $\blacktriangleright_{\mathsf{t}}$ between worlds is an arithmetical accessibility relation in this sense.

Theorem 7. *Suppose that j defines a* t*–associated model of $\mathcal{M}$, and let $\mathcal{M}^+$ be an inductive expansion of $\mathcal{M}$ with the truth predicate* tr_j *(as in Definition 10). Then*

$$\mathcal{M}^+ \vDash \forall\varphi \in \mathsf{form}\,(\mathsf{Pr}_\pi(\varphi) \to \forall \mathsf{a} \in \mathsf{as}_{\delta_\mathsf{j}}\ \mathsf{tr}_j(\varphi[\mathsf{a}]))\ \ . \tag{14}$$

where a *is an assignment from the variables of Σ to elements in the extension of δ_j, and $\varphi[\mathsf{a}]$ denotes the* sentence[14] *of the language Σ_{δ_j} obtained by simultaneously substituting the constant $c_{\mathsf{a}(v_i)}$ for v_i.*

Proof sketch. For readability, we shall drop the subscript j from δ and tr. Since induction holds in $\mathcal{M}^+$ by assumption, we can use Theorem 1 to prove the statement, taking as $\alpha(\varphi)$ the formula $\forall \mathsf{a} \in \mathsf{as}_{\delta_\mathsf{j}}\ \mathsf{tr}_j(\varphi[\mathsf{a}])$. Thus it suffices to show that the following sentences are satisfied in $\mathcal{M}^+$:

1. $\forall\varphi,\psi \in \mathsf{form}(\forall \mathsf{a} \in \mathsf{as}_\delta\ \mathsf{tr}((\varphi \to \psi)[\mathsf{a}]) \wedge \forall \mathsf{a} \in \mathsf{as}_\delta\ \mathsf{tr}(\varphi[\mathsf{a}]) \to \forall \mathsf{a} \in \mathsf{as}_\delta\ \mathsf{tr}(\psi[\mathsf{a}]))$
2. $\forall\varphi \in \mathsf{form}(\pi(\varphi) \to \forall \mathsf{a} \in \mathsf{as}_\delta\ \mathsf{tr}(\varphi[\mathsf{a}]))$
3. $\forall\varphi \in \mathsf{form}(\lambda(\varphi) \to \forall \mathsf{a} \in \mathsf{as}_\delta\ \mathsf{tr}(\varphi[\mathsf{a}]))$

Modulo some facts[15] concerning the assignments in as_δ that we assume to hold in $\mathcal{M}^+$, items 1 and 2 are consequences of items 2 and 5 of Definition 8 respectively. For item 3, we have to show (in $\mathcal{M}^+$) that whenever a formula φ is an axiom of first–order logic, then φ is true under every assignment in as_δ. This follows from the fact that tr commutes with the propositional connectives and the quantifiers (together with the above mentioned facts concerning the assignments in as_δ).

Note that if j defines a t–associated model of $\mathcal{M}$ and $\mathcal{M}^+$ is the inductive expansion of $\mathcal{M}$ with the truth predicate tr_j, then by Theorem 7 we have in particular

$$\mathcal{M}^+ \vDash \forall\varphi \in \mathsf{sent}\,(\mathsf{Pr}_\pi(\varphi) \to \mathsf{tr}_j(\varphi))\ \ . \tag{15}$$

[14] Remember that elements of the internal model are assumed to function simultaneously as codes of their own names.

[15] For example, it should hold in $\mathcal{M}^+$ that if φ is a *sentence* of the language Σ_δ, then $\mathsf{tr}(\varphi)$ if and only if $\forall \mathsf{a} \in \mathsf{as}_\delta\ \mathsf{tr}(\varphi[\mathsf{a}])$.

We shall use (15) to establish that for any world and for any sentence φ of Σ, $\mathcal{M} \vDash \mathsf{Pr}_\pi(\varphi)$ implies $\mathcal{M}' \vDash \varphi$ for all $\mathcal{M}'$ with $\mathcal{M} \blacktriangleright_{\mathsf{t}} \mathcal{M}'$. In order to establish that the other direction holds, we shall make use of the well–known fact that if $\mathcal{M} \vDash \mathsf{Con}(\varphi)$, then $\mathcal{M}$ has an internal model where φ is true. The proof is by formalizing the Completeness Theorem for first–order logic[16] in PA. By examining the proof, one can see that the internal model constructed in this process is actually a t–internal model. The (definable) formula representing the Henkin set in $\mathcal{M}$ can be seen as a truth predicate, and furthermore it defines a model of $\mathsf{PA} + \varphi$ in $\mathcal{M}$ as required in Definition 8. Thus we have the following:

Theorem 8 (Arithmetized Completeness). *Let $\mathcal{M} \vDash \mathsf{PA}$, and $\mathcal{M} \vDash \mathsf{Con}(\varphi)$. Then there exists some $\mathcal{M}'$ with $\mathcal{M} \rhd_{\mathsf{t}} \mathcal{M}'$ and $\mathcal{M}' \vDash \varphi$.*

We are now in a position to prove that each of the relations $\rhd_{\mathsf{t}}$, $\rhd_{\mathsf{tpar}}$ and $\blacktriangleright_{\mathsf{t}}$ is an arithmetical accessibility relation.

Theorem 9. *Let $\mathcal{M}$ be a model of signature Σ, and $R \in \{\rhd_t, \rhd_{\mathsf{tpar}}, \blacktriangleright_{\mathsf{t}}\}$. For any sentence φ of Σ,*

$$\mathcal{M} \vDash \mathsf{Pr}(\varphi) \Leftrightarrow \textit{ for all } \mathcal{M}' \textit{ with } \mathcal{M}\ R\ \mathcal{M}',\ \mathcal{M}' \vDash \varphi\ . \tag{16}$$

Proof. Fix a sentence φ of Σ. Suppose that $\mathcal{M} \vDash \mathsf{Pr}(\varphi)$, and let $\mathcal{M}'$ be such that $\mathcal{M}\ R\ \mathcal{M}'$. By Remark 3 it suffices to show the claim for the case that $\mathcal{M} \blacktriangleright_{\mathsf{t}} \mathcal{M}'$. So suppose that $\mathcal{M} \blacktriangleright_{\mathsf{t}} \mathcal{M}'$, let j be a relative translation that defines a t–associated model of $\mathcal{M}$ with $\mathcal{M}^j \cong \mathcal{M}'$, and let $\mathcal{M}^+$ be the inductive expansion of $\mathcal{M}$ with the truth predicate. By Theorem 7, $\mathcal{M}^+ \vDash \mathsf{tr}_j(\varphi)$. By Theorem 6 this implies $\mathcal{M}^+ \vDash \varphi^{\tau_j}$, and thus also $\mathcal{M} \vDash \varphi^{\tau_j}$. By Theorem 5 it follows that $\mathcal{M}^j \vDash \varphi$, and thus also $\mathcal{M}' \vDash \varphi$ (since $\mathcal{M}' \cong \mathcal{M}^j$). For the other direction, assume that $\mathcal{M} \vDash \neg\mathsf{Pr}(\varphi)$. By Theorem 8, there is some $\mathcal{M}'$ with $\mathcal{M} \rhd_{\mathsf{t}} \mathcal{M}'$ and $\mathcal{M}' \vDash \neg\varphi$. By Remark 3, we also have that $\mathcal{M} \rhd_{\mathsf{tpar}} \mathcal{M}'$ and $\mathcal{M} \blacktriangleright_{\mathsf{t}} \mathcal{M}'$.

4 A New Perspective on Solovay's Theorem

Arithmetical accessibility relations can be used to construct big arithmetical Kripke frames whose nodes are models of PA with signature Σ. We shall show that any GL–model is bisimilar to a Kripke model based on such a Kripke frame, obtaining the arithmetical completeness of GL as a corollary.

Definition 11 (Arithmetical Kripke frame). *An arithmetical Kripke frame is a structure $\mathcal{F}_{\mathsf{big}} = \langle W_{\mathsf{big}}, R_{\mathsf{big}} \rangle$, where W_{big} is the collection of worlds modulo isomorphism, and $R_{\mathsf{big}} \in \{\rhd_t, \rhd_{\mathsf{tpar}}, \blacktriangleright_{\mathsf{t}}\}$.*

Remark 4. An alternative but completely legitimate option is to work with arithmetical Kripke frames whose nodes are complete theories (in the language of PA)

[16] This was first noted in [10], and more carefully articulated in [2].

extending PA. The definitions in the previous section can be adjusted so as to define a relation between complete theories, and also the arguments leading to Theorem 9 would work analogously as in the case of models. We have chosen the model–theoretic approach since this yields a more natural definition of our triplet of arithmetical accessibility relations, in particular of the relation $\blacktriangleright_{\mathsf{tpar}}$ where the truth predicate is allowed to contain parameters from the external model. Another option would be to take as W_{big} the collection of all models of PA modulo *elementary equivalence*. In fact, the two alternative options are equivalent in the sense that the resulting arithmetical Kripke frames are isomorphic — to every complete theory corresponds a class of elementary equivalent models and vice versa.

Although $\mathcal{F}_{\mathsf{big}}$ is a Kripke frame, it is far from obvious that it is a GL–frame. In fact, we will see in Section 5 below that R_{big} fails to be conversely well-founded in all of the three cases (taking $\mathcal{F}_{\mathsf{big}}$ with $\blacktriangleright_{\mathsf{t}}$, the arithmetical Kripke frame even contains a reflexive point).

Given an arithmetical realization $*$, the forcing relation $\Vdash^*$ on $\mathcal{F}_{\mathsf{big}}$ is defined as follows:

$$\mathcal{M} \Vdash^* p :\Leftrightarrow \mathcal{M} \vDash p^* \ , \tag{17}$$

i.e. the propositional letter $p \in \mathcal{L}_\Box$ is forced at node $\mathcal{M}$ if and only if the arithmetical sentence p^* is satisfied in $\mathcal{M}$ (seen as a first–order model of PA). Let $\mathfrak{M}^*_{\mathsf{big}}$ denote the resulting Kripke model. As an immediate consequence of Theorem 9, we have for every sentence $A \in \mathcal{L}_\Box$ and for every $\mathcal{M} \in W_{\mathsf{big}}$,

$$\mathfrak{M}^*_{\mathsf{big}}, \mathcal{M} \Vdash^* A \Leftrightarrow \mathcal{M} \vDash A^* \ . \tag{18}$$

This means that the forcing of modal formulas is independent, modulo the realization $*$, of whether $\mathcal{M}$ is seen as a node in the Kripke model $\mathfrak{M}^*_{\mathsf{big}}$, or as a first–order model of PA.

Let $M = \langle \{1, \ldots, n\}, R, \Vdash \rangle$ be a Kripke model for GL, and let $*$ be the Solovay realization corresponding to M, i.e. $p^* := \bigvee_{i:M,i\Vdash p} \sigma_i$, where $\sigma_0, \ldots, \sigma_n$ are the Solovay sentences. Remember from Section 2.2 that the Solovay sentences are constructed in such a way that the following hold:

1. $\vdash_{\mathsf{PA}} \sigma_i \to \neg\sigma_j$ if $i \neq j$
2. $\vdash_{\mathsf{PA}} \sigma_i \to \mathsf{Con}(\sigma_j)$ if iRj, or if $i = 0$ and $j = 1$
3. $\vdash_{\mathsf{PA}} \sigma_i \to \mathsf{Pr}(\bigvee_{j:iRj} \sigma_j)$ for $i \geq 1$
4. $\nvdash_{\mathsf{PA}} \sigma_i$ and $\nvdash_{\mathsf{PA}} \neg\sigma_i$ for all i.

The following theorem states that $\mathfrak{M}^*_{\mathsf{big}}$ is bisimilar to M.

Theorem 10. *Fix a* GL*-model* $M = \langle \{1, \ldots, n\}, R, \Vdash \rangle$*. Let* $\sigma_0, \cdots, \sigma_n$ *be the corresponding Solovay sentences, and* $p^* := \bigvee_{i:M,i\Vdash p} \sigma_i$*, for any propositional letter* p *of* $\mathcal{L}_\Box$*. The relation* $Z : W \times W_{\mathsf{big}}$ *defined as:* $(i, \mathcal{M}) \in Z :\Leftrightarrow \mathcal{M} \vDash \sigma_i$ *for* $i \geq 1$ *is a bisimulation between* M *and* $\mathfrak{M}^*_{\mathsf{big}}$*. Furthermore, for every node* i *of* M *there is some node* $\mathcal{M}$ *of* $\mathfrak{M}^*_{\mathsf{big}}$ *such that* $(i, \mathcal{M}) \in Z$ *.*

Proof. Since for all i, σ_i is independent from PA, we have for all σ_i some model $\mathcal{M}$ with $\mathcal{M} \vDash \sigma_i$, and thus for all i there is some $\mathcal{M}$ such that $(i, \mathcal{M}) \in Z$. Note also that any model where σ_0 is true is not in the range of Z. We will now verify that Z is a bisimulation.

To see that if $(i, \mathcal{M}) \in Z$ then i and $\mathcal{M}$ satisfy the same propositional letters suppose first that $M, i \Vdash p$. Then by definition of $*$, we have σ_i as a disjunct of p^*. By definition of Z, we have that $\mathcal{M} \vDash \sigma_i$, hence $\mathcal{M} \vDash p^*$, and thus by definition of $\Vdash^*$ it is the case that $\mathfrak{M}_{\mathsf{big}}, \mathcal{M} \Vdash^* p$. If on the other hand $M, i \nVdash p$, then $p^* = \sigma_{j_1} \vee, \cdots, \sigma_{j_m}$, where $i \neq j_k$ for all k. By property 1 of the Solovay sentences, we find that $\vdash_{\mathsf{PA}} \sigma_i \to \neg\sigma_{j_k}$ for all k, whence $\vdash_{\mathsf{PA}} \sigma_i \to \neg p^*$ and thus $\mathcal{M} \vDash \neg p^*$. By definition of $\Vdash^*$, we have that $\mathfrak{M}_{\mathsf{big}}, \mathcal{M} \nVdash^* p$.

To verify the *back*-condition of the bisimulation, suppose that $(i, \mathcal{M}) \in Z$ and iRj. By the assumption that $(i, \mathcal{M}) \in Z$, we have that $\mathcal{M} \vDash \sigma_i$. By property 2 of the Solovay sentences and the assumption that iRj, we have $\vdash_{\mathsf{PA}} \sigma_i \to \mathsf{Con}(\sigma_j)$, and thus $\mathcal{M} \vDash \mathsf{Con}(\sigma_j)$. By Theorem 9 there is some $\mathcal{M}'$ with $\mathcal{M}\ R_{\mathsf{big}}\ \mathcal{M}'$ and $\mathcal{M}' \vDash \sigma_j$. By definition of Z, this means that $(j, \mathcal{M}') \in Z$.

Finally, to verify the *forth*–condition suppose that $(i, \mathcal{M}) \in Z$ and let $\mathcal{M}'$ be such that $\mathcal{M}\ R_{\mathsf{big}}\ \mathcal{M}'$. Since $(i, \mathcal{M}) \in Z$, we have that $\mathcal{M} \vDash \sigma_i$. By property 3 of the Solovay sentences, $\vdash_{\mathsf{PA}} \sigma_i \to \mathsf{Pr}(\bigvee_{j:iRj} \sigma_j)$, and so $\mathcal{M} \vDash \mathsf{Pr}(\bigvee_{j:iRj} \sigma_j)$. By Theorem 9, this implies $\mathcal{M}' \vDash \bigvee_{j:iRj} \sigma_j$, i.e. there is some j with iRj and $\mathcal{M}' \vDash \sigma_j$, i.e. with $(j, \mathcal{M}') \in Z$ as required.

Corollary 1 (Arithmetical completeness of GL). $\mathrm{PrL}(\mathsf{PA}) \subseteq \mathsf{GL}$

Proof. If $\mathsf{GL} \nvdash A$ for some $A \in \mathcal{L}_{\Box}$, then by modal completeness of GL there is a GL–model M with $M = \langle \{1, \ldots, n\}, R, \Vdash \rangle$, and $M, 1 \nVdash A$. Let $*$ be the Solovay realization corresponding to M, and let Z be the bisimulation from Theorem 10. Let $\mathcal{M} \in W_{\mathsf{big}}$ be such that $(1, \mathcal{M}) \in Z$, i.e. $\mathcal{M} \vDash S_1$. Since Z is a bisimulation, we have $\mathfrak{M}^*_{\mathsf{big}}, \mathcal{M} \nVdash^* A$ by Theorem 2. By (18), this implies $\mathcal{M} \nvDash A^*$. Since $\mathcal{M}$ is a model of PA, this means that $\nvdash_{\mathsf{PA}} A^*$ as required.

5 Properties of Arithmetical Kripke Frames

This section contains observations concerning the structure of the big arithmetical Kripke frames. First, it is not difficult to see that the relations $\rhd_{\mathsf{t}}$ and $\rhd_{\mathsf{tpar}}$ are transitive. Since the definition of a t–associated model postulates a truth predicate in some inductive *expansion*, the transitivity of $\blacktriangleright_{\mathsf{t}}$ is at least not obvious.

However, different from GL–frames, the big Kripke frames are not conversely well–founded. This follows from the fact that there exists a sequence of consistent theories $\{T_n\}_{n\in\omega}$ such that $T_n \vdash \mathsf{Con}(T_{n+1})$ for all n. The existence of this sequence follows from results in [8]; see also [11] . By using the formalized completeness proof, we get a sequence of models $\{\mathcal{M}_i\}_{i\in\omega}$ with $\mathcal{M}_i \vDash T_i$ and $\mathcal{M}_i \rhd_{\mathsf{t}} \mathcal{M}_{i+1}$ for all i. By Remark 3, this is an infinite ascending chain in all our arithmetical Kripke models.

We will now show that the properties of the arithmetical Kripke frame of Definition 11 depend on the choice of R_{big}. At the same time, we establish that the relations $\rhd_{\mathsf{t}}$, $\rhd_{\mathsf{tpar}}$ and $\blacktriangleright_{\mathsf{t}}$ are increasingly stronger (see Remark 3). Remember that we use the symbol "$\equiv$" to denote elementary equivalence between worlds.

Fact 11. *If $\mathcal{M} \rhd_{\mathsf{t}} \mathcal{M}'$, then $\mathcal{M} \not\equiv \mathcal{M}'$.*

To see why Fact 11 holds, suppose that j defines a t–internal model of $\mathcal{M}$. By the Gödel–Carnap Fixed Point Lemma, let γ be a sentence of the language Σ such that

$$\mathcal{M} \models \gamma \leftrightarrow \neg\mathsf{tr}(\gamma) \ . \tag{19}$$

By Theorem 6, we also have $\mathcal{M} \models \gamma^{\tau_j} \leftrightarrow \mathsf{tr}(\gamma)$. Using Theorem 5,

$$\mathcal{M}^j \models \gamma \Leftrightarrow \mathcal{M} \models \gamma^{\tau_j} \Leftrightarrow \mathcal{M} \models \mathsf{tr}(\gamma) \Leftrightarrow \mathcal{M} \not\models \gamma \ , \tag{20}$$

whence clearly $\mathcal{M} \not\equiv \mathcal{M}^j$, and thus also $\mathcal{M} \not\equiv \mathcal{M}'$ whenever $\mathcal{M} \rhd_{\mathsf{t}} \mathcal{M}'$.

In contrast, there are elementary equivalent models $\mathcal{M}$ and $\mathcal{M}'$ such that $\mathcal{M}'$ is a t–internal model of $\mathcal{M}$ with parameters.

Fact 12. *There are worlds $\mathcal{M}$ and $\mathcal{M}'$ with $\mathcal{M} \equiv \mathcal{M}'$ and $\mathcal{M} \rhd_{\mathsf{tpar}} \mathcal{M}'$.*

To establish Fact 12, let $\mathcal{N}$ be the standard model, and let Σ_c be the signature $\Sigma \cup \{c\}$, where c is a constant. One can use a standard compactness argument to show that the theory

$$T := \mathsf{Th}_{\Sigma}(\mathcal{N}) \cup \{\varphi \in c \mid \mathcal{N} \models \varphi\} + \mathsf{Con}\{\varphi \mid \varphi \in c\} \tag{21}$$

in the language Σ_c has a model $\mathcal{M}^+$ (remember that Con denotes the sentence $\neg\mathsf{Pr}(\bot)$, where Pr is an intensionally correct provability predicate of PA). Since $\mathcal{M} \models \mathsf{Con}\{\varphi \mid \varphi \in c\}$, we can use the formalized Henkin construction to find a t–internal model $\mathcal{M}'$ of $\mathcal{M}^+$ with $\mathcal{M}' \models \varphi$ for all $\varphi \in c$. Let $\mathcal{M}$ be the reduct of $\mathcal{M}^+$ to Σ, and note that $\mathcal{M} \rhd_{\mathsf{tpar}} \mathcal{M}'$ (since the construction of $\mathcal{M}'$ inside $\mathcal{M}$ uses $c^{\mathcal{M}}$ as a parameter). Since c contains the codes of all true sentences and since $\mathcal{M}$ is a model of $\mathsf{Th}_{\Sigma}(\mathcal{N})$, we have that $\mathcal{M} \equiv \mathcal{M}'$. Note that therefore also $\mathcal{M} \not\rhd_{\mathsf{t}} \mathcal{M}'$ by Fact 11.

Remark 5. As pointed out in Remark 4, we could have chosen as W_{big}, the domain of the arithmetical Kripke model, the collection of models of PA modulo *elementary equivalence*. In that case, $\langle W_{\mathsf{big}}, \rhd_{\mathsf{tpar}}\rangle$ would thus contain a reflexive point, whereas $\langle W_{\mathsf{big}}, \rhd_{\mathsf{t}}\rangle$ would not.

Finally, we give a separating example for $\rhd_{\mathsf{tpar}}$ and $\blacktriangleright_{\mathsf{t}}$.

Fact 13. *Let $\mathcal{N}$ be the standard model. Then $\mathcal{N} \blacktriangleright_{\mathsf{t}} \mathcal{N}$.*

Note first that the standard model is not a reflexive point of $\rhd_{\mathsf{tpar}}$. This is because all elements of $\mathcal{N}$ are definable, and thus $\mathcal{N} \rhd_{\mathsf{tpar}} \mathcal{M}$ implies $\mathcal{N} \rhd_{\mathsf{t}} \mathcal{M}$ for all $\mathcal{M}$. In particular $\mathcal{N} \rhd_{\mathsf{tpar}} \mathcal{N}$ would imply $\mathcal{N} \rhd_{\mathsf{t}} \mathcal{N}$, which is impossible by Fact 11.

On the other hand, the fact that the standard model $\mathcal{N}$ has a full inductive satisfaction class[17] can be used to show that there is an expansion of $\mathcal{N}$ (note that any expansion of $\mathcal{N}$ is necessarily inductive) that has a truth predicate for the sentences of $\mathcal{N}$. To make the existence of a full satisfaction class fit our definition of a t–associated model, we need to find a translation $j = \langle \delta, \tau \rangle$ such that δ is disjoint from the codes of syntactical objects of Σ, and choose τ in such a way that $\mathcal{N}^j$ is isomorphic to $\mathcal{N}$. Our convention of having the domain of $\mathcal{N}^j$ differ from that of $\mathcal{N}$ makes working out the details a bit tedious, but it is easy to convince oneself that this can be done.

6 Conclusion

We have established three examples of arithmetical accessibility relations between models of PA. We have shown how, as a result, one can see the collection of models of PA, related by one of these relations, as a big Kripke model where the forcing of modal formulas coincides with the local satisfiability of first–order sentences (modulo an arithmetical realization). We showed how this insight can be used to gain a new, model–theoretic perspective on Solovay's proof of arithmetical completeness of the modal logic GL. Finally, we have seen that the properties of the big arithmetical Kripke model are dependent on the choice of the accessibility relation as well as the domain of the Kripke model. Looking at models modulo isomorphism, the t–associated model relation ($\blacktriangleright_{\mathsf{t}}$) is the only one of the three relations giving rise to a Kripke model where the standard model is a reflexive point. If the domain of the Kripke model consists of models modulo elementary equivalence, however, then also the relation of t–internal model with parameters ($\rhd_{\mathsf{tpar}}$) has the standard model as a reflexive point.

We conclude with some pen questions.

Question 1. The arguments of this article go through if we replace PA with a Σ_1–sound theory containing $\mathsf{I}\Sigma_2$. However, Solovay's Theorem holds for all Σ_1–sound theories containing EA. Can we make our arguments work for theories weaker than $\mathsf{I}\Sigma_2$? (Σ_2–induction is used in the standard proof of the Arithmetized Completeness Theorem).

Question 2. What is the relation between the big arithmetical frames and the canonical model for GL?

Question 3. What is the modal logic of the arithmetical Kripke model if the accessibility relation is replaced by some other relation between models of PA? Some possibilities are: the internal model relation where we demand a truth predicate for the internal model (but do not require that the axioms of PA are

[17] A model $\mathcal{M}$ with signature Σ has a *full inductive satisfaction class* if there is a subset $S \subseteq \mathcal{M}$ such that S contains (possibly nonstandard) sentences of the language $\Sigma \cup \{c \mid c \in \mathcal{M}\}$, the usual Tarski's truth conditions are satisfied for S, and furthermore there is an inductive expansion of $\mathcal{M}$ to a model of the lanaguge $\Sigma \cup \{c \mid c \in \mathcal{M}\} \cup \{S\}$ (for a precise definition, see e.g. [6]).

in the extension of the truth predicate), the internal model relation, and the end-extension relation. A difference from the t–internal model relation is that these relations need not be definable by an arithmetical formula.

Acknowledgements. I thank Albert Visser who provided the insight and expertise needed to write this paper. I am grateful to Dick de Jongh for useful discussions, and for comments on several drafts of the article.

References

1. Cooper, S.B.: Computability Theory. CRC Mathematics Series, Chapman & Hall, New York, London (2004)
2. Feferman, S.: Arithmetization of Metamathematics in a General Setting. Fundamenta Mathematicae, vol 49, pp. 35–92 (1960)
3. Foundations of Mathematics mailing list `http://cs.nyu.edu/pipermail/fom/2009-August/013996.html`
4. Hamkins, J.D.: A Simple Maximality Principle. The Journal of Symbolic Logic, vol 68, pp.527-550 (2003)
5. Hamkins, J.D., Löwe, B.: The Modal Logic of Forcing. Transactions of the American Mathematical Society, vol 360, pp.1793–1817 (2008)
6. Kaye, R.:Models of Peano Arithmetic. Oxford Logic Guides, Oxford University Press (1991)
7. Lindström, P.: Aspects of Incompleteness. Lecture Notes in Logic. ASL / A K Peters, Natick, Massachusetts (2002)
8. Shavrukov, V.Yu.:Subalgebras of Diagonalizable Algebras of Theories Containing Arithmetic. Dissertationes mathematicae (Rozprawy matematyczne), vol CCCXXIII (1993)
9. Solovay, R.M.: Provability Interpretations of Modal Logic. Israel Journal of Mathematics, vol. 25, pp. 287–304 (1976)
10. Wang, H.:A Arithmetical Models for Formal Systems. Methodos, vol 3, pp. 217–232 (1951)
11. Zambella, D.: Shavrukov's Theorem on the Subalgebras of Diagonalizable Algebras for Theories Containing $I\Delta_0 + \mathsf{EXP}$. Notre Dame Journal of Formal Logic, vol 35, pp. 147–157 (1994)