

The omega-rule interpretation of transfinite provability logic

David Fernández Duque,
Universidad de Sevilla

Joost J. Joosten
Universitat de Barcelona

August 17, 2018

Abstract

Given a recursive ordinal Λ , the transfinite provability logic GLP_Λ has for each $\xi < \Lambda$ a modality $[\xi]$ with the intention of representing a sequence of provability predicates of increasing strength. One possibility is to read $[\xi]\phi$ as ϕ is provable in T using an ω -rule of depth ξ , where T is a second-order theory extending ACA_0 .

In this paper we shall formalize this notion in second-order arithmetic. Our main results are that, under some fairly general conditions for T , the logic GLP_Λ is sound and complete for the resulting interpretation.

1 Introduction

One compelling and particularly successful interpretation of modal logic is to think of $\Box\phi$ as the formula ϕ is provable, where provability is understood within a formal theory T capable of coding syntax. This was suggested by Gödel; indeed, if we use $\Diamond\phi$ as a shorthand for $\neg\Box\neg\phi$, the Second Incompleteness Theorem could be written as $\Diamond\top \rightarrow \Diamond\Box\perp$. It took some time, however, for a complete set of axioms to be assembled, namely until Löb showed $\Box(\Box\phi \rightarrow \phi) \rightarrow \Box\phi$ to be valid. It took longer still for the resulting calculus to be proven complete by Solovay [12]. The resulting modal logic is called GL (for Gödel-Löb).

Later, Japaridze [10] enriched the language of GL by adding a sequence of provability modalities $[n]$, for $n < \omega$. The modality $[0]$ is now used as before to state that ϕ is derivable within some fixed formal theory T , while higher modalities represent provability in stronger and stronger theories. There are many arithmetic interpretations for Japaridze's logic, and one of them also stems from an idea of Gödel, who introduced the notion of a theory T being ω -consistent: T is ω -consistent whenever for any formula ϕ , if $T \vdash \phi(\bar{n})$ for all $n \in \mathbb{N}$, then $T \not\vdash \exists x \neg \phi(x)$. Dually to this notion one can define a notion of

ω -provability: ϕ is ω -provable in T whenever $T + \neg\phi$ is ω -inconsistent. One may then interpret $[1]\phi$ as ϕ is ω -provable; a detailed discussion of this is given in Boolos [6].

One can then go on to interpret the higher modalities by using iterated ω -rules. This idea was already explored by Japaridze and gives an interpretation for which the polymodal logic GLP_ω is sound and complete; Ignatiev [9] and Beklemishev [4] later improved on this result. This logic is much more powerful than GL , and indeed Beklemishev has shown how it can be used to perform an ordinal analysis of Peano Arithmetic and its natural subtheories [1].

Our (hyper)arithmetical interpretations will be a straightforward generalization of Japaridze's where we read $[\alpha]_T\phi$ as *The formula ϕ is derivable in T using ω -rules of nesting depth at most α* ; we shall make this precise later. We do this by considering a well-ordering \prec on the naturals and defining a logic GLP_\prec . This is a variation of GLP_Δ already studied by the authors and Beklemishev [2, 8], with the sole difference that we shall represent ordinals as natural numbers rather than appending them as external entities.

Our main result is that GLP_\prec is sound and complete for arithmetical interpretations on ‘suitable’ theories T ; we will mainly work with extensions of ACA_0 , but as we shall discuss later, it is possible to work over a weaker base theory.

Plan of the paper Section 2 gives a quick review of the logics GLP_\prec as well as their Kripke semantics, and Section 3 of second-order arithmetic. Section 4 formalizes the notion of iterated ω -provability in second-order arithmetic; this notion is most naturally interpreted in *introspective theories*, introduced in Section 5. Section 6 proves that GLP_\prec is sound for our interpretation. In order to prove completeness, Section 7 gives a brief review of the modal logic J , which is used in the completeness proof provided in Section 8. Finally, Appendix B discusses our choice of base theory and Appendix A possible variations on the notion of iterated ω -provability.

2 The logic GLP_\prec

Formulas of the language $L_{[\cdot]}$ are built from \perp and countably many propositional variables $p \in \mathbb{P}$ using Boolean connectives \neg, \wedge and a modality $[\xi]$ for each natural number ξ . As is customary, we use $\langle\xi\rangle$ as a shorthand for $\neg[\xi]\neg$.

If \prec is a binary relation on the naturals, the logic GLP_\prec is given by the following rules and axioms:

1. all propositional tautologies,
2. $[\xi](\phi \rightarrow \psi) \rightarrow ([\xi]\phi \rightarrow [\xi]\psi)$ for all ξ ,
3. $[\xi]([\xi]\phi \rightarrow \phi) \rightarrow [\xi]\phi$ for all ξ ,
4. $\langle\zeta\rangle\phi \rightarrow \langle\xi\rangle\phi$ for $\xi \prec \zeta$,

5. $\langle \xi \rangle \phi \rightarrow [\zeta] \langle \xi \rangle \phi$ for $\xi \prec \zeta$,

6. Modus Ponens, Substitution and Necessitation: $\frac{\phi}{[\xi]\phi}$.

We will normally be interested in the case where \prec is a well-order, in which case it is known that $\langle \xi \rangle \top$ is consistent with GLP_\prec for all ξ (see [8]). In case \prec is a recursive well-order of order-type Λ we shall often write GLP_Λ instead of GLP_\prec making the necessary definitional changes for finite order types Λ .

We shall also work with Kripke semantics. A *Kripke frame* is a structure $\mathfrak{F} = \langle W, \langle R_i \rangle_{i < I} \rangle$, where W is a set and $\langle R_i \rangle_{i < I}$ a family of binary relations on W . A *valuation* on \mathfrak{F} is a function $[\cdot] : L_{[\cdot]} \rightarrow \mathcal{P}(W)$ such that

$$\begin{aligned} [\perp] &= \emptyset \\ [-\phi] &= W \setminus [\phi] \\ [\phi \wedge \psi] &= [\phi] \cap [\psi] \\ [\langle i \rangle \phi] &= R_i^{-1} [\phi]. \end{aligned}$$

A *Kripke model* is a Kripke frame equipped with a valuation $[\cdot]$. Note that propositional variables may be assigned arbitrary subsets of W . Clearly, a valuation is uniquely determined once we have fixed its values for the propositional variables. As usual, ϕ is *satisfied* on $\langle \mathfrak{F}, [\cdot] \rangle$ if $[\phi] \neq \emptyset$, and *valid* on $\langle \mathfrak{F}, [\cdot] \rangle$ if $[\phi] = W$.

It is well-known that polymodal GL is sound for \mathfrak{F} whenever R_i^{-1} is well-founded and transitive, in which case we write R_i^{-1} as $<_i$. However, constructing models of GLP_Λ is substantially more difficult than constructing models of GL, since the full logic GLP_Λ is not sound and complete for any class of Kripke frames. In Section 7 we will circumvent this problem by working in Beklemishev's J, a slightly weaker logic that is complete for a manageable class of Kripke frames.

3 Second-order arithmetic

Aside from the modal language $L_{[\cdot]}$, we will work mainly in the language L_\forall^2 of second-order arithmetic.

We fix some primitive recursive Gödel numbering mapping a formula $\psi \in L_\forall^2$ to its corresponding Gödel number $\ulcorner \psi \urcorner$, and similarly for terms and sequents of formulas (used to represent derivations). Moreover, we fix some set of *numerals* which are terms so that each natural number n is denoted by exactly one numeral written as \bar{n} . Since we will be working mainly inside theories of arithmetic, we will often identify ψ with $\ulcorner \psi \urcorner$ or even with $\overline{\ulcorner \psi \urcorner}$ for that matter.

Our results are not very sensitive to the specific choice of primitive symbols; however, to simplify notation, we will assume we have the following terms available:

1. A term $\langle x, y \rangle$ which returns a code of the ordered pair formed by x and y .
2. A term $x[y/z]$ which, when x codes a formula ϕ , y a variable v and z a term t , returns the code of the result of substituting t for v in ϕ . Otherwise, its value is unspecified, for example it could be the default \top . We shall often just write $\phi(t)$ for this term if the context allows us to.
3. A term $x \rightarrow y$ which, when x, y are codes for ϕ, ψ , returns a code of $\phi \rightarrow \psi$, and similarly for other Boolean operators or quantifiers. The context should always clarify if we use the symbol \rightarrow as a term or as a logical connective.
4. A term \bar{x} mapping a natural number to the code of its numeral.
5. For every formula ϕ , a term $\phi(\dot{x})$ which, given a natural number n , returns the code of the outcome of $\phi[x/\bar{n}]$, i.e., the code of $\phi(\bar{n})$.

We will also use this notation in the metalanguage. The only purpose of assuming these terms exist is to shorten complex formulas, as the graphs of all these functions are definable by low level arithmetic formulas over most standard arithmetic languages.

As is customary, we use Δ_0^0 to denote the set of all formulas (possibly with set parameters) where all quantifiers are “bounded”, that is, of the form $\forall x < y \phi$ or $\exists x < y \phi$. We simultaneously define $\Sigma_0^0 = \Pi_0^0 = \Delta_0^0$ and Σ_{n+1}^0 to be the set of all formulas of the form $\exists x_0 \dots \exists x_n \phi$ with $\phi \in \Pi_n^0$ and similarly Π_{n+1}^0 to be the set of all formulas of the form $\forall x_0 \dots \forall x_n \phi$ with $\phi \in \Sigma_n^0$. We denote by Π_ω^0 the union of all Π_n^0 ; these are the *arithmetic formulas*.

The classes Σ_n^1, Π_n^1 are defined analogously but using second-order quantifiers and setting $\Sigma_0^1 = \Pi_0^1 = \Delta_0^1 = \Pi_\omega^0$. It is well-known that every second-order formula is equivalent to another in one of the above forms. If Γ is a set of formulas, we denote by $\hat{\Gamma}$ the subset of Γ where no set variables appear free.

We will say a theory T is *representable* if there is a $\hat{\Delta}_0^0$ formula $\text{Proof}_T(x, y)$ which holds if and only if x codes a derivation in T of a formula coded by y ; in general we assume all theories to be representable. We may also assume without loss of generality that any derivation d is a derivation of a unique formula ϕ , for example by representing d as a finite sequence of formulas whose last element is ϕ . Also, we assume that every formula that is derivable has arbitrarily large derivations; this is generally true of standard proof systems, for example one may add many copies of an unused axiom or many redundant cuts. Whenever it does not lead to confusion we will work directly with codes rather than formulas; if ϕ is a natural number (supposedly coding a formula) we use $\square_T \phi$ as a shorthand for $\exists y \text{Proof}_T(y, \phi)$.

It is important in this paper to keep track of the second-order principles that are used; below we describe the most important ones. We use $<$ to denote the standard ordering on the naturals and Γ denotes some set of formulas:

$$\begin{array}{lll}
\Gamma\text{-CA} & \exists X \forall x(x \in X \leftrightarrow \phi(x)) & \text{where } \phi \in \Gamma; \\
\text{I-}\Gamma & \phi(\overline{0}) \wedge \forall x(\phi(x) \rightarrow \phi(x + \overline{1})) \rightarrow \forall x\phi(x) & \text{where } \phi \in \Gamma; \\
\text{Ind} & \forall x(\forall y < x y \in X \rightarrow x \in X) \rightarrow \forall x x \in X.
\end{array}$$

We assume *all* theories extend two-sorted first-order logic, so that they include Modus Ponens, Generalization, etc., as well as Robinson's Arithmetic, i.e. Peano Arithmetic without induction.

Another principle that will be relevant to us is *transfinite recursion*, but this is a bit more elaborate to describe. For simplicity let us assume that L_\forall^2 contains only monadic set-variables; binary relations and functions can be represented by coding pairs of numbers. It will be convenient to establish a few conventions for working with binary relations in second-order arithmetic. First, let us write *R is a binary relation* and *f is a function*:

$$\begin{aligned}
\mathbf{rel}(R) &= \forall x(x \in R \rightarrow \exists y \exists z(x = \langle y, z \rangle)), \\
\mathbf{funct}(f) &= \mathbf{rel}(f) \wedge \forall x \exists! y(\langle x, y \rangle \in f).
\end{aligned}$$

Here, $\exists!$ is the standard abbreviation for *there exists a unique*.

Also for simplicity, we may write $n R m$ if R represents a relation and $\langle n, m \rangle \in R$ as well as $n \not R m$ for $\neg(\langle n, m \rangle \in R)$, or $n = f(m)$ if $\langle m, n \rangle \in f$ and f is meant to be interpreted as a function. Further, it is possible to work with a second-order equality symbol, but it suffices to define $X \equiv Y$ by $\forall x(x \in X \leftrightarrow y \in Y)$.

It will also be important to represent ordinals in second-order arithmetic. For this we will reserve a set-variable \prec . Here, we will need to express *the relation \prec is a linear order* and *the relation \prec is well-ordered*, as follows:

$\mathbf{linear}(\prec) :$

$$\forall x(\neg(x \prec x) \wedge \forall y(x \prec y \vee y \prec x \vee y = x)) \wedge \forall x, y, z(x \prec y \wedge y \prec z \rightarrow x \prec z);$$

$\mathbf{wo}(\prec) :$

$$\mathbf{linear}(\prec) \wedge \forall X \left(\exists x(x \in X) \rightarrow \exists y \forall z(z \prec y \rightarrow \neg(z \in X)) \right).$$

We will use Greek letters for natural numbers when viewed as ordered under \prec . When it is clear from context we may use natural numbers to represent finite ordinals, so that, for example, 0 is the least element under \prec , independently of whether it truly corresponds to the natural number zero.

We shall often want that the elementary properties of \prec be provable, for example,

$$\xi \prec \zeta \rightarrow \square_T \xi \prec \zeta.$$

This can be guaranteed if we work with recursive well-orders, in which case we assume T contains an axiom $\forall x, y (x \prec y \leftrightarrow \sigma(x, y))$ for some $\hat{\Sigma}_1^0$ formula $\sigma(x, y)$.

Transfinite recursion is the principle that sets may be defined by iterating a formula along a well-order. To formalize this, let us consider a set X whose elements are of the form $\langle \xi, x \rangle$. Write X_ξ for $\{x \mid \langle \xi, x \rangle \in X\}$ and $X_{\prec\xi}$ for $\{x \mid \exists \zeta \prec \xi \langle \zeta, x \rangle \in X\}$. Then, given a set of formulas Γ we define

$$\text{TR-}\Gamma \quad \text{wo}(\prec) \rightarrow \exists X \forall \xi \forall x \left(x \in X_\xi \leftrightarrow \phi(x, X_{\prec\xi}) \right) \quad \text{for } \phi \in \Gamma.$$

With this, we may define the following systems of arithmetic:

$$\begin{aligned} \text{RCA}_0 &:= \text{I-}\Sigma_1^0 + \Delta_0^0\text{-CA} \\ \text{ACA}_0 &:= \text{Ind} + \Pi_\omega^0\text{-CA} \\ \text{ATR}_0 &:= \text{Ind} + \text{TR-}\Pi_\omega^0. \end{aligned}$$

We list these from weakest to strongest, but even ATR_0 is fairly weak in the realm of second-order arithmetic. For convenience we will work mainly in ACA_0 , but later discuss how our techniques could be pushed down even to below RCA_0 at the cost of slightly stronger transfinite induction.

The system ATR_0 is relevant because we will define iterated provability by recursion over the well-order \prec . However, as we shall see, we require much less than the full power of arithmetic transfinite recursion.

In various proofs we wish to reason by transfinite induction. By $\text{TI}(\prec, \phi)$ we denote the transfinite induction axiom for ϕ along the ordering \prec :

$$\text{TI}(\prec, \phi) := \forall \xi (\forall \zeta \prec \xi \phi(\zeta) \rightarrow \phi(\xi)) \rightarrow \forall \xi \phi(\xi).$$

We will write $\phi\text{-CA}$ instead of $\{\phi\}\text{-CA}$, i.e., the instance of the comprehension axiom stating that $\{x \mid \phi(x)\}$ is a set. The following lemma tells us that we have access to transfinite induction for formulas of the right complexity:

Lemma 3.1. *In any second order arithmetic theory containing predicate logic we can prove*

$$\text{wo}(\prec) \wedge \neg \phi\text{-CA} \rightarrow \text{TI}(\prec, \phi).$$

Proof. Reason in T and assume $\text{wo}(\prec) \wedge \neg \phi\text{-CA}$. We prove $\text{TI}(\prec, \phi)$ by contradiction. Thus, suppose that $\exists \lambda \neg \phi(\gamma)$. As $\{\xi \mid \neg \phi(\xi)\}$ is a set, we can apply $\text{wo}(\prec)$ to obtain the minimal such λ . Clearly for this minimal λ we do not have $\forall \zeta \prec \lambda \phi(\zeta) \rightarrow \phi(\lambda)$. \square

4 Nested ω -rules

In this section we shall formalize the notion of iterated ω -rules inside second-order arithmetic. In Boolos ([6]) it is noted that multiple *parallel* applications of the ω -rule do not add extra strength. For example, the rule that allows us to conclude σ from

$$\begin{aligned} \forall n &\vdash \psi(\bar{n}) \\ \forall m &\vdash \forall x \psi(x) \rightarrow \phi(\bar{m}) \\ &\vdash \forall x \phi(x) \rightarrow \sigma \end{aligned}$$

can actually be derived by a single application of the ω rule.

However, when we admit slightly less uniformity by allowing ψ to depend on m in this rule, and adding the premises $\forall n \vdash \psi_m(\bar{n})$ we get our notion of 2-provability. More generally, we may iterate this process to generate a hierarchy or stronger and stronger notions of ξ -provability for a recursive ordinal ξ . It is the *nesting depth* that gives extra strength and not the number of applications.

We will use $[\lambda]_T^\prec \phi$ to denote our representation of *The formula ϕ is provable in T using one application of an ω -rule of depth λ (according to \prec)*. The desired recursion for such a sequence of provability predicates is given by the following equivalence.¹

$$[\lambda]_T^\prec \phi \leftrightarrow \left(\square_T \phi \vee \exists \psi \exists \xi \prec \lambda (\forall n [\xi]_T^\prec \psi(\dot{n}) \wedge \square_T (\forall x \psi(x) \rightarrow \phi)) \right). \quad (1)$$

As a first step in such a formalization, we will use a set X as an ‘iterated provability class’ IPC for short. Its elements are codes of pairs $\langle \lambda, \phi \rangle$, with λ a code for an ordinal and ϕ a code for a formula; we use $[\lambda]_X \phi$ as a shorthand for $\langle \lambda, \phi \rangle \in X$ and $\langle \lambda \rangle_X \phi$ for $\langle \lambda, \neg\phi \rangle \notin X$. Clearly, any IPC will depend on a parameter \prec whose intended interpretation is a well-ordering on the naturals. We then define a formula $\text{IPC}_T^\prec(X)$ (‘iterated provability class’) as a formalization of:

$[\lambda]_X \phi$ if and only if

1. $\lambda = 0$ and $\square_T \phi$, or
2. there is a formula $\psi(x)$ and an ordinal $\xi \prec \lambda$ such that
 - (a) for each $n < \omega$, $[\xi]_X \psi(\bar{n})$, and
 - (b) $\square_T (\forall x \psi(x) \rightarrow \phi)$.

Intuitively, we understand $\text{IPC}_T^\prec(X)$ as stating “ X is an iterated provability predicate” and in the remainder of this text we will use both ‘class’ or ‘predicate’ to refer to IPCs. Let us enter in a bit more detail:

Definition 4.1. Define $\text{Rule}_T^\prec(\lambda, \phi \mid X)$ by

$$\exists \psi \exists \xi \prec \lambda (\forall n [\xi]_X \psi(\dot{n}) \wedge \square_T (\forall x \psi(x) \rightarrow \phi))$$

and let $\text{IPC}_T^\prec(X)$ be the formula

$$\forall z [z \in X \leftrightarrow \exists \lambda \exists \phi (\langle z, \phi \rangle = \langle \lambda, \phi \rangle \wedge (\square_T \phi \vee \text{Rule}_T^\prec(\lambda, \phi \mid X)))].$$

Then, $[\lambda]_T^\prec \phi$ is the Π_1^1 -formula $\forall X (\text{IPC}_T^\prec(X) \rightarrow [\lambda]_X \phi)$.

¹There are other reasonable ways of defining this recursion. In Appendix A, we shall discuss some possible alternatives.

Note that the formulas $[\lambda]_X\phi$ and $\langle\lambda\rangle_X\phi$ are independent of T and of \prec and are merely of complexity Δ_0^0 . Note also that for r.e. theories T we have that $\text{Rule}_T^\prec(\lambda, \phi \mid X)$ is a Σ_2^0 -formula whence $\text{IPC}_T^\prec(X)$ is a Π_3^0 -formula. We can write the definition of $\text{IPC}_T^\prec(X)$ more succinctly as

$$\text{IPC}_T^\prec(X) \leftrightarrow \forall\xi, \phi \left([\xi]_X\phi \leftrightarrow (\square_T\phi \vee \text{Rule}_T^\prec(\xi, \phi \mid X)) \right).$$

From the definition of our provability predicates we easily obtain monotonicity:

Lemma 4.2. *Given theories U, T where U extends ACA_0 and T is representable, we have that*

$$U \vdash (\xi \prec \zeta) \rightarrow ([\xi]_T^\prec\phi \rightarrow [\zeta]_T^\prec\phi).$$

Proof. We reason within U . Suppose $[\xi]_T^\prec\phi$ holds as well as $\text{IPC}_T^\prec(X)$. Letting x be any variable not appearing in ϕ we have that $[\xi]_X\phi(\bar{n})$ holds for all $n < \omega$ as well as $\square_T(\forall x\phi \rightarrow \phi)$. Thus, $[\zeta]_X\phi$ and, since X was arbitrary, $[\zeta]_T^\prec\phi$ holds. \square

Corollary 4.3. *Let U and T be as in Lemma 4.2. If $T \vdash \phi$, then $T \vdash [\lambda]_T^\prec\phi$.*

Proof. If $T \vdash \phi$ then $\square_T\phi$ is a true $\hat{\Sigma}_1^0$ sentence whence $U \vdash \square_T\phi$. Thus also $U \vdash [0]_T^\prec\phi$ and by monotonicity (Lemma 4.2), we get $U \vdash [\lambda]_T^\prec\phi$. \square

Our notion of ξ -provability, $[\xi]_T^\prec$, is a very weak one as it has a universal quantification over all provability predicates X and it may be the case that there are no such predicates. Dually, the notion of consistency $\langle\xi\rangle_T^\prec$ is very strong as it in particular asserts the existence of a provability predicate. In particular, we always provably have

$$\square_T\phi \rightarrow [0]_T^\prec\phi. \tag{2}$$

However, we can in general not prove $[0]_T^\prec\phi \rightarrow \square_T\phi$. Nevertheless, the two notions of provability coincide under the assumption that a provability predicate exists:

Lemma 4.4. *Given a representable theory T and a theory U that extends ACA_0 we have*

$$U \vdash \exists X \text{IPC}_T^\prec(X) \rightarrow (\square_T\phi \leftrightarrow [0]_T^\prec\phi).$$

Proof. The proof is straightforward and is left to the reader. \square

In the field of formalized provability one often uses formalized Σ_1^0 completeness (see e.g. [6]):

Lemma 4.5. *Let T be some representable arithmetic theory with induction for all $\hat{\Delta}_0^0$ formulas (i.e., Δ_0^0 without free set variables) and where exponentiation is provably total. Let σ be a $\hat{\Sigma}_1^0$ formula. We have*

$$T \vdash \sigma \rightarrow \square_T\sigma.$$

Of course, it makes really no sense to speak of provable Σ_1^0 completeness where we allow free set variables. In particular we cannot apply it to our notion $[0]_X\phi$, which is why Lemma 4.4 will often be useful.

A useful fact is that $[\lambda]_T^\prec\phi$ is well-defined in the following sense:

Lemma 4.6. *Given theories U, T where U extens ACA₀ and T is representable, we have that U proves*

$$\text{wo}(\prec) \rightarrow \forall X \forall Y \left(\text{IPC}_T^\prec(X) \wedge \text{IPC}_T^\prec(Y) \rightarrow \forall x (x \in X \leftrightarrow x \in Y) \right).$$

Proof. This follows by a simple induction over \prec . □

Moreover, under the assumption of $\text{wo}(\prec)$ we can also show a useful monotonicity property of our provability predicates.

Lemma 4.7. *Let U be some theory extending ACA₀ and let T and T' be representable theories. Write $T \subseteq T'$ as a shorthand for $\forall \phi \square_T \phi \rightarrow \square_{T'} \phi$.*

Then,

$$U + \text{wo}(\prec) \vdash T \subseteq T' \rightarrow ([\lambda]_T^\prec \phi \rightarrow [\lambda]_{T'}^\prec \phi).$$

Proof. Reason in $U + \text{wo}(\prec)$ and assume that $T \subseteq T'$. By an easy induction on λ it is shown that if both $\text{IPC}_T^\prec(X)$ and $\text{IPC}_{T'}^\prec(X')$, then $\langle \lambda, \phi \rangle \in X \rightarrow \langle \lambda, \phi \rangle \in X'$. □

5 Introspective theories

For a theory T to be able to reason about non-trivial facts of iterated provability at all, it is necessary for it to at least “believe” that such a notion exists. For strong theories this is not an issue, but there is no reason to assume that ACA₀ or any weaker theory is capable of proving that we have provability predicates. Hence we shall pay attention to those theories that do have them, and we shall call them *introspective theories*.

Definition 5.1 (Introspective theory). *An arithmetic theory T is \prec -introspective if $T \vdash \exists X \text{IPC}_T^\prec(X)$.*

We defined provability operators by transfinite recursion, and as such it should be no surprise that ATR₀ is introspective:

Lemma 5.2. *Given an elementarily presented theory T ,*

$$\text{ATR}_0 \vdash \text{wo}(\prec) \rightarrow \exists X \text{IPC}_T^\prec(\prec, X).$$

In particular, ATR₀ is introspective.

However, we wish to work over much weaker theories than ATR₀, which may not be introspective. Our strategy will be to consider some sort of an “introspective closure”, but do not wish for it to become much stronger than the original theory. Fortunately, this is not too difficult to achieve.

Definition 5.3. We define the \prec -introspective closure of T as the theory \overline{T} given by $T + \exists X \text{IPC}_T^\prec(X)$.

Below, we use the term “Gödelian” somewhat informally as being susceptible to Gödel’s second incompleteness theorem; for example, it could be taken to mean sound, representable and extending RCA_0 .

Lemma 5.4. T is equiconsistent with \overline{T} , provided T is Gödelian and contains Δ_0^0 comprehension.

Proof. Clearly the consistency of \overline{T} implies the consistency of T . For the other direction we use that if T is Gödelian, then T is equiconsistent with $T' := T + \square_{T\perp}$. We claim that $T' \vdash \exists X \text{IPC}_T^\prec(X)$ so that $T' \supseteq \overline{T}$ whence

$$\begin{aligned} \text{Con}(T) &\Rightarrow \text{Con}(T') \\ &\Rightarrow \text{Con}(\overline{T}). \end{aligned}$$

Indeed, reasoning within T , if T were inconsistent, then $\square_T \phi$ for every formula ϕ . It follows that if X is an iterated provability operator, then $[\lambda]_X \phi$ for all λ and ϕ ; hence the trivial set consisting of all pairs $\langle \lambda, \phi \rangle$ is an iterated provability operator, and by Δ_0^0 comprehension, it forms a set. \square

There is still a danger of sliding down a slippery-slope, where \overline{T} is itself not introspective, thus needing to generate a sequence of theories that is each “introspective over the previous”. Fortunately, this is not the case. In order to show this we need a technical lemma reminiscent of the Deduction Theorem.

Definition 5.5. Let X be an iterated provability operator, so that $\text{IPC}_T^\prec(X)$ holds. We define the set X given θ – which we denote by $\{X|\theta\}$ – as

$$\langle \lambda, \phi \rangle \in \{X|\theta\} \iff \langle \lambda, \theta \rightarrow \phi \rangle \in X.$$

The technical lemma that we shall now prove tells us in particular that introspection is preserved under taking finite extensions.

Lemma 5.6. Let U be some theory containing Δ_0^0 comprehension, and let T be representable. Then

$$U \vdash \text{IPC}_T^\prec(Y) \rightarrow \text{IPC}_{T+\theta}^\prec(\{Y|\theta\}).$$

Proof. We reason in U and assume $\text{IPC}_T^\prec(Y)$. By Δ_0^0 comprehension we see that $\{Y|\theta\}$ is a set. We need to show that $[\lambda]_{\{Y|\theta\}} \phi \leftrightarrow \square_{T+\theta} \phi \vee \text{Rule}_{T+\theta}^\prec(\lambda, \phi \mid \{Y|\theta\})$. Since

$$\begin{aligned} [\lambda]_{\{Y|\theta\}} \phi &\leftrightarrow [\lambda]_Y (\theta \rightarrow \phi) \\ &\leftrightarrow \square_T (\theta \rightarrow \phi) \vee \text{Rule}_T^\prec(\lambda, \theta \rightarrow \phi \mid Y), \end{aligned}$$

and since $\square_{T+\theta} \phi \leftrightarrow \square_T (\theta \rightarrow \phi)$, it suffices to show $\text{Rule}_{T+\theta}^\prec(\lambda, \phi \mid \{Y|\theta\}) \leftrightarrow \text{Rule}_T^\prec(\lambda, \theta \rightarrow \phi \mid Y)$.

But this follows easily from the tautology

$$(\theta \rightarrow (\forall x \psi(x) \rightarrow \phi)) \leftrightarrow (\forall x (\theta \rightarrow \psi(x)) \rightarrow (\theta \rightarrow \phi))$$

and the definition of $\{Y|\theta\}$. \square

As a direct consequence of this lemma we see that the introspective closure of a theory is indeed itself introspective.

Lemma 5.7. *Using Δ_0^0 comprehension one can show that \overline{T} is introspective.*

Proof. By the above Lemma 5.6, if Y is a provability predicate for T , then $\{Y \mid \exists X \text{IPC}_T^\prec(X)\}$ is a provability predicate for \overline{T} . Moreover, by Δ_0^0 comprehension, it forms a set. \square

We conclude that working with introspective theories is not too restrictive:

Corollary 5.8. *Every Gödelian theory T is equiconsistent to an introspective theory \overline{T} .*

Note that in general we may not expect for U to prove $\exists X \text{IPC}_{T+\theta}^\prec(X) \rightarrow \exists X \text{IPC}_T^\prec(X)$ since for $\theta = \square_T \perp$ the antecedent always holds (recall the proof of Lemma 5.4).

A problem that \overline{T} has is that it introduces an existential second-order formula, which may make it hard later to control the complexity of the resulting theory. Because of this, it is sometimes more convenient to work with *explicitly introspective* theories:

Definition 5.9. *Given a formal theory T , we define T^π over the language $L_V^2 + \{\pi\}$, where π is a new set-constant and $T^\pi = T + \text{IPC}_T^\prec(\pi)$.*

Much as with \overline{T} , T^π is introspective, provided T contains Δ_0^0 comprehension. In fact, we can do a bit better in this case. Recall that, given a class of formulas Γ , $\hat{\Gamma}$ denotes the set of those formulas of Γ with no open set variables; excluding π , of course, which is a constant. Then we have the following:

Lemma 5.10. *Using $\hat{\Delta}_0^0$ comprehension one can prove that T^π is introspective.*

The proof proceeds as before and we omit it. Parameter-free comprehension is very convenient in that it does not “blow up”, as it cannot be iterated; for example, Π_1^0 comprehension with set parameters is equivalent to full arithmetic comprehension, but $\hat{\Pi}_1^0$ comprehension is not.

Now that we have shown that introspective theories are not such a bad thing to work with, we will employ them freely in the next sections. Introspective theories are capable of reasoning about their own iterated provability; for example, we may prove the desired recursion as stated in (1).

Lemma 5.11. *Let T be a theory that extends ACA₀. Then, we have that*

1. $T \vdash (\square_T \phi \vee \exists \psi \exists \xi \prec \lambda (\forall n [\xi]_T^\prec \psi(n) \wedge \square_T (\forall x \psi(x) \rightarrow \phi))) \rightarrow [\lambda]_T^\prec \phi;$
2. $\overline{T} \vdash [\lambda]_T^\prec \phi \leftrightarrow (\square_T \phi \vee \exists \psi \exists \xi \prec \lambda (\forall n [\xi]_T^\prec \psi(n) \wedge \square_T (\forall x \psi(x) \rightarrow \phi)))$.

Proof. In the first item, we reason in T and need to prove $[\lambda]_T^\prec \phi$ under the assumption of the antecedent. To this end, we fix some X with $\text{IPC}_T^\prec(X)$ and show $[\lambda]_X \phi$. However, this follows directly from the definition of X being a provability predicate since we can replace $\forall n [\xi]_T^\prec \psi(\dot{n})$ in the antecedent by $\forall n [\xi]_X \psi(\dot{n})$.

For the remaining implication in the second item we reason as follows. In case $\lambda = 0$ we get the implication by Lemma 4.4. In case $\lambda > 0$, from the definition we see that for any provability predicate X we have

$$[\lambda]_X \phi \rightarrow \left(\square_T \phi \vee \exists \psi \exists \xi \prec \lambda (\forall n [\xi]_X \psi(\dot{n}) \wedge \square_T (\forall x \psi(x) \rightarrow \phi)) \right).$$

From this we obtain

$$[\lambda]_T^\prec \phi \rightarrow \forall X \left[\text{IPC}_T^\prec(X) \rightarrow \left(\square_T \phi \vee \exists \psi \exists \xi \prec \lambda (\forall n [\xi]_X \psi(\dot{n}) \wedge \square_T (\forall x \psi(x) \rightarrow \phi)) \right) \right],$$

from which the claim directly follows. \square

6 Soundness

In this section we shall see that indeed GLP_\prec is sound for its arithmetic interpretation. In Lemma 4.2 we have already seen the soundness of the monotonicity axiom $[\xi]\phi \rightarrow [\zeta]\phi$ for $\xi \prec \zeta$. For the remaining axioms we will transfinite induction over \prec so we define, given a second-order theory T , a new theory T^\prec as

$$T^\prec := T + \text{wo}(\prec).$$

We will assume that T contains ACA_0 , although in Appendix B we shall discuss this choice. Since introspection is closed under taking finite extensions both \overline{T}^\prec and \overline{T}^\prec are introspective (though not necessarily equivalent); for all our arguments below it is irrelevant which one we use.

Let us first check the soundness of the basic distribution axiom.

Lemma 6.1. *Given theories U, T where U extends ACA_0 and T is representable, then*

$$U^\prec \vdash [\lambda]_T^\prec (\phi_1 \rightarrow \phi_2) \rightarrow ([\lambda]_T^\prec \phi_1 \rightarrow [\lambda]_T^\prec \phi_2).$$

Proof. We reason within U^\prec .

Let X be a provability predicate. We shall prove by induction on λ that

$$\forall \phi_1, \phi_2 ([\lambda]_X \phi_1 \wedge [\lambda]_X (\phi_1 \rightarrow \phi_2) \rightarrow [\lambda]_X \phi_2). \quad (3)$$

Note that by Lemma 3.1 we only need Σ_1^0 comprehension (with set parameters) to have access to this transfinite induction.

So, we assume that $[\lambda]_X \phi_1 \wedge [\lambda]_X (\phi_1 \rightarrow \phi_2)$ and let ψ_1, ψ_2 be such that

1. for each $i = 1, 2$ there is $\xi_i < \lambda$ such that for all $n < \omega$, $\langle \xi_i, \psi_i(\bar{n}) \rangle \in X$,
2. $\square_T(\forall x\psi_1(x) \rightarrow \phi_1)$,
3. $\square_T(\forall x\psi_2(x) \rightarrow (\phi_1 \rightarrow \phi_2))$.

By first-order logic we see that

$$\square_T(\forall x(\psi_1(x) \wedge \psi_2(x)) \rightarrow \phi_2). \quad (4)$$

Let $\xi = \max\{\xi_1, \xi_2\}$. By induction on $\xi \prec \lambda$ and several uses of Modus Ponens inside $[\xi]_X$ we obtain for each n that $[\xi]_X(\psi_1(\bar{n}) \wedge \psi_2(\bar{n}))$. But given that X is an IPC, this shows in combination with (4) that $[\lambda]_X\phi_2$ and we have shown (3).

To conclude the proof, we assume that $[\lambda]_T^\prec(\phi_1 \rightarrow \phi_2) \wedge [\lambda]_T^\prec\phi_1$. Thus, for an arbitrary provability predicate X we have $[\lambda]_X(\phi_1 \rightarrow \phi_2) \wedge [\lambda]_X\phi_1$ whence by (3) also $[\lambda]_X\phi_2$. As X was arbitrary, we obtain $[\lambda]_T^\prec\phi_2$. \square

With our distribution axiom at hand we can now obtain a formalized Deduction Theorem.

Lemma 6.2. *Let U be a theory extending ACA_0 and let T be representable. We have that*

$$U^\prec \vdash [\lambda]_{T+\theta}^\prec\phi \leftrightarrow [\lambda]_T^\prec(\theta \rightarrow \phi).$$

Proof. If $[\lambda]_T^\prec(\theta \rightarrow \phi)$ then, by Lemma 4.7 we also have $[\lambda]_{T+\theta}^\prec(\theta \rightarrow \phi)$. Since clearly $[\lambda]_{T+\theta}^\prec\theta$, by the distribution axiom we get $[\lambda]_{T+\theta}^\prec\phi$.

For the other direction, reason in U^\prec and assume $[\lambda]_{T+\theta}^\prec\phi$. Let X be arbitrary with $\text{IPC}_T^\prec(X)$. By Lemma 5.6 we see that $\text{IPC}_{T+\theta}^\prec(\{X|\theta\})$. Now by the assumption that $[\lambda]_{T+\theta}^\prec\phi$ we see that $[\lambda]_{\{X|\theta\}}\phi$ so that consequently $[\lambda]_X(\theta \rightarrow \phi)$. \square

So far we have shown that some of the axioms of GLP_\prec are sound for our omega-rule interpretation; Löb's axiom and the “provable consistency” axiom remain to be checked. For the former, the following lemma will be quite useful.

Lemma 6.3. *Extend GL with a new operator \blacksquare and the following axioms for all formulas ϕ , and ψ :*

1. $\vdash \square\phi \rightarrow \blacksquare\phi$,
2. $\vdash \blacksquare(\phi \rightarrow \psi) \rightarrow (\blacksquare\phi \rightarrow \blacksquare\psi)$ and,
3. $\vdash \blacksquare\phi \rightarrow \blacksquare\blacksquare\phi$,

and call the resulting system GL^\blacksquare .

Then for all ϕ ,

$$\text{GL}^\blacksquare \vdash \blacksquare(\blacksquare\phi \rightarrow \phi) \rightarrow \blacksquare\phi.$$

Proof. It is well-known that GL is equivalent to K4 plus the Löb Rule:

$$\frac{\Box\phi \rightarrow \phi}{\phi}.$$

Thus it suffices to check that this rule holds for \blacksquare . But indeed, assume that $\text{GL}^\blacksquare \vdash \blacksquare\phi \rightarrow \phi$. Then, using $\Box\phi \rightarrow \blacksquare\phi$ we obtain $\Box\phi \rightarrow \phi$, and by Löb's rule (for \Box) we see that $\text{GL}^\blacksquare \vdash \phi$, as desired. \square

Thus to show that $[\lambda]_T^\prec$ is Löbian for all λ , we need only show the following:

Lemma 6.4. *Given a recursive order \prec , theories U, T where U extends ACA₀ and T is representable, we have that*

$$U^\prec \vdash \forall\phi\forall\lambda \quad [\lambda]_T^\prec\phi \rightarrow [\lambda]_T^\prec[\dot{\lambda}]_T^\prec\dot{\phi}.$$

Proof. Reason within U^\prec . We assume $\text{IPC}_T^\prec(X)$ and will show by induction on \prec that if $[\lambda]_X\phi$, then $[\lambda]_X[\bar{\lambda}]_T^\prec\bar{\phi}$, from which the lemma clearly follows.

The base case, when $\lambda=0$, is straightforward. We assume $[0]_X\phi$ and by Lemma 4.4 we get $\Box_T\phi$ whence $\Box_T\Box_T\bar{\phi}$ by provable $\hat{\Sigma}_1^0$ -completeness of U . Consequently, by applying (2) twice (once under the box) we obtain $[0]_T^\prec[0]_T^\prec\bar{\phi}$ whence certainly also $[0]_X[\bar{0}]_T^\prec\bar{\phi}$.

Now assume that $\lambda \succ 0$ and there are $\xi \prec \lambda$ and ψ such that for all n , $[\xi]_X\psi(\bar{n})$ and $\Box_T(\forall x\psi(x) \rightarrow \phi)$.

By the induction hypothesis on $\xi \prec \lambda$, for every number n we can see that $[\xi]_X[\bar{\xi}]_T^\prec\bar{\psi}(\bar{n})$. Thus we obtain by one application of the ω -rule that

$$[\lambda]_X\forall n[\bar{\xi}]_T^\prec\bar{\psi}(\dot{n}). \quad (5)$$

Meanwhile, we have that $[0]_X\Box_T(\forall x\bar{\psi}(x) \rightarrow \bar{\phi})$ from which it follows by monotonicity that

$$[\lambda]_X\Box_T(\forall x\bar{\psi}(x) \rightarrow \bar{\phi}). \quad (6)$$

Since \prec is recursive we also have that

$$[\lambda]_X\bar{\xi} \prec \bar{\lambda}. \quad (7)$$

Putting (5), (6) and (7) together and bringing the existential quantifiers under the box we conclude that

$$[\lambda]_X\left(\exists\psi\exists\xi\prec\lambda \quad (\forall n[\xi]_T^\prec\psi(\dot{n}) \wedge \Box_T(\forall x\psi(x) \rightarrow \phi))\right).$$

By an application under the box of Lemma 5.11.1 (note that no need of introspection is required) we obtain $[\lambda]_X[\bar{\lambda}]_T^\prec\bar{\phi}$ as was to be proven. \square

In the proof of the remaining GLP _{\prec} axiom we will need for the first and only time the assumption that T is introspective.

Lemma 6.5. *If U is any theory extending ACA₀, \prec is recursive and T is representable and \prec -introspective, then*

$$U^\prec \vdash \forall\phi\forall\lambda\forall\xi\prec\lambda \quad \langle\xi\rangle_T^\prec\phi \rightarrow [\lambda]_T^\prec\langle\dot{\xi}\rangle_T^\prec\dot{\phi}.$$

Proof. We reason in U^\prec and assume that $\xi \prec \lambda$. Let us first see that it is sufficient to show that for an arbitrary provability predicate X we have $\langle \xi \rangle_X \phi \rightarrow [\lambda]_X \langle \bar{\xi} \rangle_T \bar{\phi}$.

If we wish to show $[\lambda]_T \langle \bar{\xi} \rangle_T \bar{\phi}$ we pick an arbitrary provability predicate X and set out to prove $[\lambda]_X \langle \bar{\xi} \rangle_T \bar{\phi}$. By the assumption $\langle \xi \rangle_T \phi$ we know that there is some provability predicate Y with $\langle \xi, \phi \rangle \notin Y$, that is, $\langle \xi \rangle_Y \phi$. By Lemma 4.6, we have that $X \equiv Y$, whence $\langle \xi \rangle_X \phi$ and indeed, $\langle \xi \rangle_X \phi \rightarrow [\lambda]_X \langle \bar{\xi} \rangle_T \bar{\phi}$ suffices to finish the proof.

In view of the above, we will prove $\langle \xi \rangle_X \phi \rightarrow [\lambda]_X \langle \bar{\xi} \rangle_T \bar{\phi}$ by induction on λ . For the base case, when $\lambda = 1$, we reason as follows. From $\langle 0 \rangle_X \phi$, we obtain the $\hat{\Pi}_1^0$ sentence $\Diamond_T \phi$, that is, $\forall n \neg \text{Proof}_T(n, \bar{\phi})$. Since $\neg \text{Proof}_T(n, \bar{\phi}) \in \hat{\Sigma}_1^0$, we get $\forall n \Box_T \neg \text{Proof}_T(n, \bar{\phi})$ and also $\forall n [0]_T \neg \text{Proof}_T(n, \bar{\phi})$. Then by applying an ω -rule we see that $[1]_X \Diamond_T \phi$. Since T is \prec -introspective then $[1]_X \exists Y \text{IPC}_T^\prec(Y)$, and by Lemma 4.4

$$[1]_X (\exists Y \text{IPC}_T^\prec(Y) \rightarrow (\langle \bar{0} \rangle_T \bar{\phi} \leftrightarrow \Diamond_T \bar{\phi})),$$

from which we conclude that $[1]_X \langle \bar{0} \rangle_T \bar{\phi}$.

So assume that $\lambda \succ 1$. If we have that $\langle \xi \rangle_X \phi$ then for every formula ψ either

1. for all $\eta \prec \xi$ there is $n < \omega$ such that $\langle \eta \rangle_X \neg \psi(\bar{n})$, or
2. $\Diamond_T (\forall x \psi(x) \wedge \phi)$ holds.

In the first case, by the induction hypothesis for $\xi \prec \lambda$ we can see that $\exists n [\xi]_X \langle \eta \rangle_T \neg \psi(\bar{n})$; in the second, we have that $[\xi]_X \Diamond_T (\forall x \psi(x) \wedge \phi)$. Combining these, we obtain

$$\eta \prec \xi \rightarrow [\xi]_X \left(\exists x \langle \eta \rangle_T \neg \psi(\dot{x}) \vee \Diamond_T (\forall x \psi(x) \wedge \phi) \right).$$

Since \prec is recursive, we know that $\eta \succcurlyeq \xi \rightarrow [\xi]_X \bar{\eta} \succcurlyeq \bar{\xi}$. We thus see that, for all pairs $\langle \eta, \psi \rangle$,

$$[\xi]_X \left(\bar{\eta} \succcurlyeq \bar{\xi} \vee \exists x \langle \eta \rangle_T \neg \psi(\dot{x}) \vee \Diamond_T (\forall x \psi(x) \wedge \phi) \right).$$

By one application of the ω -rule to all pairs $\langle \eta, \psi \rangle$ (represented as natural numbers) we obtain

$$[\lambda]_X \left(\forall \psi \forall \eta \prec \xi \left(\exists x \langle \eta \rangle_T \neg \psi(\dot{x}) \vee \Diamond_T (\forall x \psi(x) \wedge \phi) \right) \right),$$

and by definition (Lemma 5.11.1 applied under the box) we get $[\lambda]_X \langle \bar{\xi} \rangle_T \bar{\phi}$. \square

We have essentially proven that GLP_\prec is sound for its omega-rule interpretation, but we need the following definition in order to make this claim precise.

Definition 6.6. An arithmetic interpretation is a function $f : \mathbb{P} \rightarrow \mathsf{L}_\forall^2$.

We denote by f_T^\prec the unique extension of f such that $f_T^\prec(p) = f(p)$ for every propositional variable p , $f_T^\prec(\perp) = \perp$, f_T^\prec commutes with Booleans and $f_T^\prec([\lambda]\phi) = [\lambda]_T f_T^\prec(\phi)$.

Theorem 6.7 (Soundness). *If \prec is any recursive well-order on the naturals, U is a sound theory extending ACA₀, T is \prec -introspective and representable and $\text{GLP}_{\prec} \vdash \phi$ then $U^{\prec} \vdash f_T^{\prec}(\phi)$ for every arithmetic interpretation f .*

Proof. By an easy induction on the length of a GLP_{\prec} -proof of ϕ , using the fact that each of the axioms is derivable. Necessitation is just Corollary 4.3. \square

Now that we have proven that GLP_{\prec} is sound, our main objective will be to prove the converse of Theorem 6.7 which is hyper-arithmetical completeness of GLP_{\prec} . For this, let us first review the modal logic J .

7 The logic J

It is well-known that GLP_{Λ} has no non-trivial Kripke frames for $\Lambda > 1$. In order to remedy for this situation, we pass to a weaker logic, Beklemishev's J . The logic J is as GLP_{ω} where we replace the monotonicity axiom of GLP_{ω} by the two axioms

6. $[n] \rightarrow [m][n]\phi$, for $n \leq m$ and
7. $[n] \rightarrow [n][m]\phi$, for $n < m$.

The logic J is proven in [3] to be sound and complete for the class of finite Kripke models $\langle W, \langle >_n \rangle_{n < N}, [\cdot] \rangle$ such that

1. the relations $<_n$ are transitive and well-founded,
2. if $n < m$ and $w <_m v$ then $<_n(w) = <_n(v)$ (where $<_n(w) = \{u : u <_n w\}$) and,
3. if $n < m$ then $w <_m v <_n u$ implies that $w <_n u$.

It will also be convenient to define some auxiliary relations. Say:

- $w \ll_n v$ if for some $m \geq n$, $w <_m v$ and,
- $w \lll_n v$ if $w \ll_n v$ or there is $u \in W$ such that $w \ll_n u$ and $v \ll_{n+1} u$.

By the above frame conditions it is easy to see that \ll_n is transitive and well-founded.

We will also use \ll_n and \lll_n to denote the respective reflexive closures. Let \approx_n denote the symmetric, reflexive, transitive closure of \ll_n and let $[w]_n$ denote the equivalence class of w under \approx_n . Write $[w]_{n+1} <_n [v]_{n+1}$ if there exist $w' \in [w]_{n+1}$, $v' \in [v]_{n+1}$ such that $w' <_n v'$.

A J -frame W is said to be *stratified* if whenever $[w]_{n+1} <_n [v]_{n+1}$, it follows that $w <_n v$. Note that the property of being stratified in particular entails the modally inexpressible frame condition that $w <_n v$ and $w <_m u$ implies $u <_n v$ whenever $m > n$. With this we may state the following completeness result also from [3]:

Lemma 7.1. *Any \mathbf{J} -consistent formula can be satisfied on a finite, stratified \mathbf{J} -frame.*

Thus if we can reduce \mathbf{GLP}_ω to \mathbf{J} , we will be able to work with finite well-behaved Kripke models. For this, given a formula ϕ whose maximal modality is N , define

$$M(\phi) = \bigwedge_{\substack{[n]\psi \in \text{sub}(\phi) \\ n < m \leq N}} [n]\psi \rightarrow [m]\psi.$$

Then we set $M^+(\phi) = M(\phi) \wedge \bigwedge_{n \leq N} [n]M(\phi)$.

The following is also proven in [3]:

Lemma 7.2. *For any formula $\phi \in \mathcal{L}_\omega$, $\mathbf{GLP}_\omega \vdash \phi$ if and only if*

$$\mathbf{J} \vdash M^+(\phi) \rightarrow \phi.$$

We shall use these results in the next section to prove arithmetical completeness by “piggybacking” from the completeness of \mathbf{J} for finite frames.

8 Completeness

In this section we want to prove that \mathbf{GLP}_\prec is complete for its ω -rule interpretation. This means that, given a consistent formula ϕ , there is an arithmetic interpretation f such that $\neg f_T^\prec(\phi)$ is not derivable in T (we will make this claim precise in Theorem 8.2).

There are many proofs of completeness of \mathbf{GL} and \mathbf{GLP}_ω , and it is possible to go back to an existing proof and adjust it to prove completeness in our setting. Because of this, we should say a few words about our choice of including a full proof in this paper. There are essentially two reasons.

The first is that, while our result follows to a certain degree from known *proofs*, it does not follow from known *results*; even then, there would be several technical issues in adjusting known arguments to our setting, as they make assumptions that are not available to us.

The second is that the argument we propose carries some simplifications over previous proofs that could also be applied to standard interpretations of \mathbf{GLP}_ω , thus contributing to an ongoing effort to find simpler arguments for this celebrated result.

To be more precise, there are at least six proofs in the literature:

1. Solovay originally constructed a function h with domain ω of a self-referential nature and used statements about h to prove the completeness theorem for the unimodal \mathbf{GLP}_1 [12]. The proof used the recursion theorem.
2. De Jongh, Jumelet and Montagna introduced a modification using the fixpoint theorem instead of the recursion theorem [7], where the function h is simulated via finite sequences that represent computations [7]. This approach is presented in greater detail in [6].

3. A more elementary construction using the simultaneous fixpoint theorem is also given in [7].
4. Japaridze proved completeness for GLP_ω with essentially the ω -rule interpretation we are presenting here [10].
5. Ignatiev generalized this result to a large family of “strong provability predicates” [9].
6. Beklemishev gave a simplified argument using the logic J , which is very well-behaved. However, this proof still considers a family of N Solovay functions h_n with domain ω , where N is the number of modal operators appearing in our “target formula” ϕ .

Despite these strong provability predicates being quite general, they do not apply to our interpretation, as for example it is assumed that they are of increasing logical complexity whereas our iterated provability operators are all given by a single Π_1^1 formula. The argument we present here, aside from being the first that considers arbitrary recursive well-orders, combines ideas from [7] and [4] by considering finite paths over a polymodal J -frame. We do so by introducing an additional trick, which is to work with all modalities simultaneously, where our path makes a λ_n -step whenever appropriate. Readers familiar with known proofs might find it surprising that this is not problematic, but indeed it isn’t and otherwise the argument proceeds as in other settings. As always, we will mimic a Kripke structure using arithmetic formulas and define our arithmetic interpretation based on them.

Since GLP_\prec is Kripke incomplete, we will resort to J -models instead. These models are related to GLP_ω as described in the previous section. The step from GLP_\prec to GLP_ω is provided by the following easy lemma which is also given in [5].

Lemma 8.1. *Let ϕ be a GLP_\prec formula whose occurring modalities in increasing order are $\{\lambda_0, \dots, \lambda_N\}$. By $\bar{\phi}$ we denote the condensation of ϕ that arises by simultaneously replacing each occurrence of $[\lambda_i]$ by $[i]$. It now holds that*

$$\text{GLP}_\prec \not\vdash \phi \implies \text{GLP}_\omega \not\vdash \bar{\phi}.$$

Proof. Arguing by contrapositive, given a GLP_ω -derivation d of ϕ we may replace every occurrence of $[n]$ in d by $[\lambda_n]$, thus obtaining a derivation of ϕ . \square

With this easy lemma at hand we may give an outline of the proof of our completeness theorem, which reads as follows.

Theorem 8.2. *If \prec is recursive, T is any sound, representable, \prec -introspective theory extending ACA_0 and proving $\text{wo}(\prec)$ and ϕ is any $L_{[\cdot]}$ -formula, $\text{GLP}_\prec \vdash \phi$ if and only if, for every arithmetical interpretation f , $T^\prec \vdash f_T^*(\phi)$.*

Proof sketch. One direction is soundness and has already been established.

For the other, if $\text{GLP}_\prec \not\vdash \phi$ then by our above Lemma 8.1 combined with Lemma 7.2, $M^+(\overline{\phi}) \wedge \neg\overline{\phi}$ is not provable in J .

Thus by Lemma 7.1, $M^+(\overline{\phi}) \wedge \neg\overline{\phi}$ can be satisfied on a world w_* of some J -model $\mathfrak{W}' = \langle W', \langle >_n \rangle_{n < N}, \llbracket \cdot \rrbracket \rangle$ where $W' = [1, M]$ for some $M \geq 1$. We construct a new model \mathfrak{W} which is as W' only that now a new $<_0$ -maximal root 0. The valuation of propositional letters on 0 is chosen arbitrarily and is irrelevant.

The next ingredient is to assign to each $w \in W$ an arithmetic sentence σ_w so that the formulas σ are a ‘‘snapshot’’ of \mathfrak{W} . We will make this precise in Definition 8.4, but let us outline the essential properties that we need from σ .

First, we need for the arithmetic interpretation f that sends a propositional variable p to $f(p) := \bigvee_{w \in \llbracket p \rrbracket} \sigma_w$ to have the property that

$$\mathfrak{W}, w \Vdash \overline{\psi} \iff T \vdash \sigma_w \rightarrow f_T^\prec(\psi)$$

for each $w \in W'$ and each subformula ψ of ϕ . In particular we have $T \vdash \sigma_{w_*} \rightarrow \neg f_T^\prec(\phi)$ from which we obtain

$$T \vdash \Diamond_T \sigma_1 \rightarrow \neg \Box_T f_T^\prec(\phi). \quad (8)$$

Our desired result will follow if the formulas σ satisfy two more properties: the second is that

$$T \vdash \sigma_0 \rightarrow \Diamond_T \sigma_1,$$

and the third, that $\mathbb{N} \models \sigma_0$. By the assumption that T is sound we conclude that $\mathbb{N} \models \neg \Box_T f_T^\prec(\phi)$. Hence, $f_T^\prec(\phi)$ is not provable in T which is what was to be shown. \square

Before we proceed to give the details needed to complete the proof we state as an easy consequence of our arithmetic completeness theorem the following lemma which was also proven by purely modal means in [5].

Corollary 8.3. *Given a recursive well-order \prec and an $L_{[\cdot]}$ -formula ϕ we have that*

$$\text{GLP}_\prec \vdash \phi \iff \text{GLP}_\omega \vdash \overline{\phi}.$$

Proof. One direction is Lemma 8.1. For the other direction, suppose $\text{GLP}_\omega \not\vdash \overline{\phi}$. By the proof of Theorem 8.2 we find an arithmetical interpretation f so that $\overline{\text{ACA}_0} \not\vdash f_T^\prec(\phi)$. By the soundness theorem (Thm. 6.7) we conclude that $\text{GLP}_\prec \not\vdash \phi$. \square

Before entering into further detail, we first say what it means that a collection of sentences $\sigma = \{\sigma_0, \dots, \sigma_k\}$ is a snapshot of a Kripke structure with nodes $\{0, \dots, k\}$ inside a theory. Most importantly, this means that each world w will be associated with an arithmetic sentence σ_w so that this sentence carries all the important information in terms of accessible worlds.

Definition 8.4. Given a sequence

$$\boldsymbol{\lambda} = \lambda_0 \prec \lambda_1 \prec \dots \prec \lambda_{N-1},$$

a finite J-model $\mathfrak{W} = \langle W, \langle <_n \rangle_{n < N}, \llbracket \cdot \rrbracket \rangle$ with root 0, and a formal theory T , a family of formulas $\{\sigma_w : w \in W\}$ is a $\boldsymbol{\lambda}$ -snapshot of \mathfrak{W} in T if

1. $T \vdash \bigwedge_{w \neq v \in W} \neg(\sigma_w \wedge \sigma_v),$
2. $T + \sigma_w \vdash \langle \bar{\lambda}_n \rangle_T^\prec \sigma_v$ for all $n < N$ and $v <_n w$,
3. for all $n < N$ and for each world $w \neq 0$,

$$T + \sigma_w \vdash [\bar{\lambda}_n]_T^\prec \bigvee_{v \lll_n w} \sigma_v$$

4. $\mathbb{N} \models \sigma_0.$

If $\mathfrak{W}, \boldsymbol{\sigma}, \boldsymbol{\lambda}, T$ are as above we will write $\boldsymbol{\sigma} : \mathfrak{W} \xrightarrow{\boldsymbol{\lambda}} T$.

Lemma 8.5. Suppose that $\boldsymbol{\sigma} : \mathfrak{W} \xrightarrow{\boldsymbol{\lambda}} T$, ϕ is an $L_{[\cdot]}$ -formula with modalities amongst $\boldsymbol{\lambda}$ such that $\mathfrak{W} \models M^+(\bar{\phi})$, and $f(p) := \bigvee_{w \in \llbracket p \rrbracket} \sigma_w$. Then, for all $0 \neq w \in W$ and every subformula ψ of ϕ ,

1. if $w \in \llbracket \bar{\psi} \rrbracket$ then $T + \sigma_w \vdash f_T^\prec(\psi)$
2. if $w \notin \llbracket \bar{\psi} \rrbracket$ then $T + \sigma_w \vdash \neg f_T^\prec(\psi)$

Proof. By an easy induction on the complexity of ψ . □

In the remainder of this section, we shall mainly see how to produce snapshots of a given Kripke model \mathfrak{W} in a theory T . We define the corresponding sentences σ_w for $w \in W$ in a standard way as limit statements of certain computable Solovay functions.

One important notion in all known Solovay-style proofs, including our own, is the notion of a “code for a ξ -derivation of ϕ .”

Definition 8.6. Let X be a set of natural numbers. A code of a 0-proof of ϕ over X is any natural number m satisfying $\text{Proof}_T(m, \phi)$. We assume that every derivation proves a unique formula in our coding, and that this fact is derivable; we also assume that every derivable formula has arbitrarily large derivations².

For $\xi > 0$, a triple $\langle \zeta, n, m \rangle$ codes a ξ -derivation of ϕ over X if $\xi > \zeta$ and

1. $n = \lceil \psi \rceil$ for some ψ such that, given $k < \omega$, $[\zeta]_X \psi(\bar{k})$ and

²Usually, derivations are represented as sequences of formulas, and any given derivation can be considered to *only* prove its last formula. Similarly, if derivations are represented as trees, then only the root is considered to be proven. Moreover, in standard proof systems, given a derivation d , there are ways to produce a longer derivation with the same end-formula; for example, one may add many redundant copies of an axiom at the beginning of d .

2. m is a code of a 0-proof of $\forall x\psi \rightarrow \phi$.

Let $\text{Proof}_X(x, \xi, \phi)$ be a formula stating that $x = \langle \zeta, \psi, d \rangle$ codes a ξ -derivation of ϕ over X .

Of course in Proof_X the intention is for X to be an iterated provability operator, and we may define

$$\text{Proof}_T^\prec = \forall X (\text{IPC}^\prec(X) \rightarrow \text{Proof}_X).$$

Note that formula Proof_T^\prec is of rather high complexity (Π_1^1) which is moreover independent of ξ . However, the behavior of $\text{Proof}_T^\prec(x, \xi, \phi)$ is simple in the eyes of λ provability whenever $\lambda \succcurlyeq \xi$, as is expressed in the next lemma.

Lemma 8.7. *Let T be a theory extending ACA₀. For $\xi \preccurlyeq \lambda$ we have*

1. $\overline{T}^\prec \vdash \forall x \forall \xi \forall \phi \text{Proof}_T^\prec(x, \xi, \phi) \rightarrow [\lambda]_T^\prec \text{Proof}_T^\prec(\dot{x}, \dot{\xi}, \dot{\phi})$,
2. $\overline{T}^\prec \vdash \forall x \forall \xi \forall \phi \neg \text{Proof}_T^\prec(x, \xi, \phi) \rightarrow [\lambda]_T^\prec \neg \text{Proof}_T^\prec(\dot{x}, \dot{\xi}, \dot{\phi})$.

Proof. We prove the first item. We reason in \overline{T}^\prec and assume $\text{Proof}_T^\prec(x, \xi, \phi)$. If $\xi = 0$, $\text{Proof}_T(d, \phi)$ is $\hat{\Delta}_1^0$ so $[0]_X \text{Proof}_T(\overline{d}, \overline{\phi})$. So we assume $\xi \succ 0$ whence $\text{Proof}_T(d, \forall x\psi(x) \rightarrow \phi)$ is equivalent to

$$x = \langle \psi, \mu, d \rangle \wedge \mu \prec \xi \wedge \forall n [\mu]_T^\prec \psi(\overline{n}) \wedge \text{Proof}_T(d, \forall x\psi(x) \rightarrow \phi). \quad (9)$$

All conjuncts other than $\forall n [\mu]_T^\prec \psi(\overline{n})$ are of complexity $\hat{\Delta}_1^0$ so they—as their negations—all are 0-provable whence certainly λ -provable. By Lemma 6.4 we obtain $\forall n [\mu]_T^\prec \psi(\dot{n}) \rightarrow \forall n [\mu]_T^\prec [\dot{\mu}]_T^\prec \psi(\dot{n})$, and since $\mu \prec \lambda$ we use one application of the ω -rule to see that

$$\forall n [\mu]_T^\prec \psi(\dot{n}) \rightarrow [\lambda]_T^\prec \forall n [\dot{\mu}]_T^\prec \psi(\dot{n}).$$

Since the $[\lambda]_T^\prec$ predicate is closed under conjunction we have the entire conjunction (9) under the scope of the $[\lambda]_T^\prec$ predicate which was to be shown.

The second item goes analogously now using Lemma 6.5 instead of 6.4. \square

8.1 Solovay sequences

Let us define a *Solovay sequence* or *path*; these sequences are given by a recursion based on provability operators which depends on a parameter ϕ . Later we will choose an appropriate value of ϕ via a fixpoint construction. We shall use the following notation: $\text{Seq}(x)$ is a $\hat{\Delta}_0^0$ formula stating that x codes a sequence, $\text{last}(x)$ is a term that picks out the last element of x , $x \sqsubseteq y$ is a $\hat{\Delta}_0^0$ formula that states that x is an initial segment of y , $|x|$ gives the length of x and x_y a term which picks the y -coordinate of x . As in previous sections, it is not necessary to have these terms available in our language, as we can define their graphs and replace them by pseudo-terms, but we shall write them as such for simplicity of exposition.

We will also define a (pseudo) term Lim which gives a formula stating that the paths satisfying ϕ “converge” to w :

Definition 8.8. Define

$$\text{Lim}(\phi, w) := \exists s \left(\phi(\dot{s}) \wedge \forall s' \sqsupseteq s \ \phi(\dot{s}') \rightarrow \text{last}(s') = \bar{w} \right).$$

We shall use these formulas to define our recursive paths.

Definition 8.9. Let $\mathfrak{W} = \langle W, \langle <_n \rangle_{n < N}, [\cdot] \rangle$ be a Kripke frame and let $\lambda = \langle \lambda_n \rangle_{n < N}$ a finite sequence. We define a formula $s : T \xrightarrow{\lambda} \mathfrak{W} \mid \phi$ by

$$\begin{aligned} s : T \xrightarrow{\lambda} \mathfrak{W} \mid \phi := \\ \text{Seq}(s) \wedge \text{last}(s) \neq 0 \\ \wedge \forall x < |s| - 1 \ \bigwedge_{w \in W} \left(s_x = \bar{w} \rightarrow \right. \\ \left(\bigwedge_{n < N} \bigwedge_{v <_n w} \neg \text{Proof}_X(x, \lambda_n, \neg \text{Lim}(\phi, v)) \right) \rightarrow s_{x+1} = \bar{w} \\ \left. \wedge \left(\bigvee_{n < N} \bigvee_{v <_n w} \text{Proof}_X(x, \lambda_n, \neg \text{Lim}(\phi, v)) \right) \rightarrow s_{x+1} = \bar{v} \right). \end{aligned}$$

We then set $s : T \xrightarrow{\lambda} \mathfrak{W} \mid \phi = \forall X \left(\text{IPC}_T^<(X) \rightarrow s : T \xrightarrow{\lambda} \mathfrak{W} \mid \phi \right)$.

We should remark that s, X, ϕ are variables and λ, \mathfrak{W} are external parameters so that we are in fact defining a family of formulas. With this we can say what it means to be a Solovay path.

Definition 8.10 (Solovay path). We define a Solovay path to be any natural number s satisfying the formula $s : T \xrightarrow{\lambda} \mathfrak{W}$ defined using the fixpoint theorem on the parameter ϕ in $s : T \xrightarrow{\lambda} \mathfrak{W} \mid \phi$, so that

$$\text{ACA}_0 \vdash s : T \xrightarrow{\lambda} \mathfrak{W} \leftrightarrow (\dot{s} : T \xrightarrow{\lambda} \mathfrak{W} \mid \Gamma x : T \xrightarrow{\lambda} \mathfrak{W}^\Gamma).$$

Further, we say w is a Solovay value at i if

$$w \simeq \{T \xrightarrow{\lambda} \mathfrak{W}\}_i := \exists s \ (s : T \xrightarrow{\lambda} \mathfrak{W} \wedge |s| > i \wedge s_i = w)$$

holds, and w is a limit Solovay value if it satisfies

$$w \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\} := \text{Lim}(\Gamma x : T \xrightarrow{\lambda} \mathfrak{W}^\Gamma, w).$$

The following shows that Solovay values in fact define a function.

Lemma 8.11. If U extends ACA_0 , T is any representable theory, \mathfrak{W} is a J-frame and λ a \prec -increasing sequence, it is derivable in U that

1. $\forall s \forall s' (s : T \xrightarrow{\lambda} \mathfrak{W} \wedge s' : T \xrightarrow{\lambda} \mathfrak{W} \rightarrow s \sqsubseteq s' \vee s' \sqsubseteq s)$

2. $\forall I \exists s : T \xrightarrow{\lambda} \mathfrak{W} \wedge |s| > I$
3. $\forall i \exists ! w : w \simeq \{T \xrightarrow{\lambda} \mathfrak{W}\}_i$ and
4. $\forall s \forall j < i : \bar{v} \simeq \{T \xrightarrow{\lambda} \mathfrak{W}\}_j \wedge \bar{w} \simeq \{T \xrightarrow{\lambda} \mathfrak{W}\}_i \rightarrow \bar{w} \leqslant_0 \bar{v}$.

Proof.

1 Clearly it suffices to prove that

$$\left((s : T \xrightarrow{\lambda} \mathfrak{W} \mid \phi) \wedge (s' : T \xrightarrow{\lambda} \mathfrak{W} \mid \phi) \wedge i < |s| \wedge i < |s'| \right) \rightarrow s_i = s'_i,$$

for then if s, s' are any two paths and, say, $|s| \leq |s'|$, it follows that $s_i = s'_i$ for all $i < |s|$ and thus $s \sqsubseteq s'$. Moreover, this formula is arithmetic and hence we may proceed by induction on i .

The base case is trivial since $s_0 = s'_0 = 0$. For the inductive step, we assume $w = s_i = s'_i$. Then, we must have that either $\text{Proof}_X(i, \lambda_n, \neg \text{Lim}, \phi, \bar{v})$ holds for some v, n , or it does not. If it does, then the value of v is uniquely determined (as i may be the code of a derivation of only one formula) and thus $s_{i+1} = s'_{i+1} = v$. Otherwise, $\bigwedge_n \bigwedge_{v <_n w} \neg \text{Proof}_X(x, \lambda_n, \neg \text{Lim}(\phi, \bar{v}))$ holds and $s_{i+1} = s'_{i+1} = w$. Once again the claim follows by introducing universal quantifiers over X and ϕ .

2 The proof follows the above structure; here we observe that if s is a Solovay path, we may always add one additional element to s depending on which condition is met.

3 This is immediate from items 1 and 2.

4 By the recursive definition of a Solovay path, it is always the case that $s_{j+1} \leqslant_0 s_j$. Since \leqslant_0 is transitive, this implies inductively that $s_i \leqslant_0 s_j$ whenever $j < i$, and this induction can be easily formalized in U . \square

Lemma 8.12. *Let \mathfrak{W} be a finite J-frame and λ a \prec -increasing sequence. Suppose further that U extends ACA₀, \prec is recursive and T is representable and \prec -introspective, $w \in W$ and $n \leq N$. Then,*

$$\begin{aligned} U^\prec \vdash \bar{w} &\simeq \{T \xrightarrow{\lambda} \mathfrak{W}\}_k \\ &\rightarrow [\lambda_n]_T^\prec \bigwedge_{v \in W} \bar{v} \simeq \{T \xrightarrow{\lambda} \mathfrak{W}\}_k \rightarrow \bar{v} \approx_{n+1} \bar{w}. \end{aligned}$$

Further,

$$\begin{aligned} U^\prec \vdash m &\leq n \wedge \bar{u} >_m \bar{w} \\ &\wedge \bar{u} \simeq \{T \xrightarrow{\lambda} \mathfrak{W}\}_k \wedge \bar{w} \simeq \{T \xrightarrow{\lambda} \mathfrak{W}\}_{k+1} \\ &\rightarrow [\lambda_n]_T^\prec \bigwedge_{v \in W} \bar{w} \simeq \{T \xrightarrow{\lambda} \mathfrak{W}\}_k. \end{aligned}$$

Proof. Reasoning within U , we will prove both claims simultaneously. To be precise, we show by induction on k that

$$\begin{aligned} \text{IPC}^\prec(X) \wedge s : T \xrightarrow{\lambda} \mathfrak{W} \mid \Gamma x : T \xrightarrow{\lambda} \mathfrak{W} \wedge |s| > k \rightarrow \\ \bigwedge_{w \in W} s_k = w \rightarrow [\lambda_n]_X \forall x \quad x : T \xrightarrow{\lambda} \mathfrak{W} \wedge |x| > k \rightarrow x_k \approx_{n+1} \bar{w} \\ \wedge \bigwedge_{m < n} s_k >_m w \rightarrow [\lambda_n]_X \forall x \quad x : T \xrightarrow{\lambda} \mathfrak{W} \wedge |x| > i \rightarrow x_{i+1} = \bar{w}. \end{aligned}$$

Case 1. Suppose that $w <_m s_k$ for some $m \leq n$. Then, k codes a λ_m -derivation of $\bar{w} \not\simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$, and by Lemma 8.7.1,

$$[\lambda_n]_X \text{Proof}_T^\prec(\bar{k}, \bar{\lambda}_m, \bar{w} \not\simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}).$$

Meanwhile, for $v = s_k$, by our induction hypothesis

$$[\lambda_n]_X(x : T \xrightarrow{\lambda} \mathfrak{W} \wedge |x| > k \rightarrow x_k \approx_{n+1} \bar{v}),$$

but $x_k \approx_{n+1} v <_m w$ provably implies that $x_k <_m w$ by the J-frame conditions and thus

$$[\lambda_n]_X \forall x \left(x : T \xrightarrow{\lambda} \mathfrak{W} \wedge |x| > \bar{k} + 1 \rightarrow x_{\bar{k}+1} = \bar{w} \right).$$

The claim follows by quantifying over all X .

Case 2. Suppose that for no $m \leq n$ do we have that $s_{k+1} <_m s_k$; then as in Case 1 we have that $w <_m x_k$ if and only if $w <_m s_k$ and by Lemma 8.7.2 we have for each such w and m that

$$[\lambda_n]_T^\prec \neg \text{Proof}_T^\prec(\bar{k}, \bar{\lambda}_m, \bar{w} \not\simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}).$$

Thus in view of Lemma 8.11.4 and the previous case we must have that

$$x_{k+1} \ll_{n+1} x_k \approx_{n+1} s_k \gg_{n+1} s_{k+1},$$

which by definition implies that $x_{k+1} \approx_{n+1} s_{k+1}$. Formalizing this reasoning within T , it follows that

$$[\lambda_n]_X \forall x \left(x : T \xrightarrow{\lambda} \mathfrak{W} \wedge |x| > \bar{k} + 1 \rightarrow x_{\bar{k}+1} \approx_{n+1} \bar{w} \right),$$

and once again we conclude the original claim by quantifying over X . \square

From here on, it remains to show that the formulas $\bar{w} \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$ give a snapshot of our Kripke model.

Lemma 8.13. *If U extends ACA₀, \prec is recursive and T is representable and \prec -introspective, $m < n < N$ and $v <_m w$ then*

$$U \quad + \quad \bar{w} \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\} \quad \vdash \quad \langle \lambda_n \rangle_T^\prec \bar{v} \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}.$$

Proof. Towards a contradiction, suppose that

$$[\lambda_n]_T^\prec \bar{v} \not\simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}.$$

It follows that $[\lambda_n]_T^\prec \bar{v} \not\simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$, and hence there exists some i which satisfies $\text{Proof}_T^\prec(i, \lambda_n, \bar{v} \not\simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\})$. Now, by assumption $w \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$ holds, and hence we may choose s such that $\text{Last}(s') = w$ for all $s' \sqsupseteq s$. Recall that we assumed that every derivable formula has arbitrarily large derivations (see Definition 8.6), and thus we may pick $i > |s|$ such that

$$\text{Proof}_T^\prec(i, \lambda_n, \bar{v} \not\simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\})$$

holds and, in view of Lemma 8.11.2, a Solovay path s' with $|s'| > i + 1$. Then, $s'_i = s'_{i+1} = w$, but by the Solovay recursion we should have $s'_{i+1} = v$, a contradiction. \square

Lemma 8.14. *If U, T extend ACA₀, \prec is recursive and T is representable and \prec -introspective, $w \neq 0$ and $n \leq N$ then*

$$U^\prec + \bar{w} \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\} \vdash [\lambda_n]_T^\prec \bigvee_{v \lll_n w} \bar{v} \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}.$$

Proof. We reason in $U^\prec + \bar{w} \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$. Let s be a Solovay path with $w = \text{Last}(s)$. Let $k_* < |s|$ be the greatest value such that $s_{k_*} <_n s_{k_*-1}$ if there is such a value; otherwise set $k_* = 0$.

By Lemma 8.12,

$$[\lambda_n]_T^\prec \left(\forall x \ x : T \xrightarrow{\lambda} \mathfrak{W} \wedge |x| > \bar{k}_* \rightarrow x_{\bar{k}_*} = \bar{s}_{k_*} \right). \quad (10)$$

Moreover, by Lemma 8.13 we have that $\langle \lambda_m \rangle_T^\prec \bar{v} \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$ so that by Lemma 6.5 we also have

$$[\lambda_n]_T^\prec \bigwedge_{m < n} \bigwedge_{v <_m w} \langle \bar{\lambda}_m \rangle_T^\prec \bar{v} \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\},$$

from which it follows using Lemma 8.11.4 that

$$[\lambda_n]_T^\prec \forall x \ \bar{k}_* < j \wedge x : T \xrightarrow{\lambda} \mathfrak{W} \wedge |x| > j \rightarrow x_j \lll_n x_{\bar{k}_*}. \quad (11)$$

Putting (10) and (11) together, along with the fact that $s_{k_*} \ggg_{n+1} w$ we see that

$$[\lambda_n]_T^\prec \forall x \ x : T \xrightarrow{\lambda} \mathfrak{W} \wedge |x| > \bar{k}_* \rightarrow \text{Last}(x) \lll_n \bar{w}.$$

It remains to disprove the case that $\text{Last}(x) = \bar{w}$. For this, choose the least value of k such that $s_{k+1} = w$; note that this value is well-defined since $s_0 = 0$. It then follows that $\text{Proof}_T^\prec(k, \lambda_n, \bar{w} \not\simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\})$, and in view of Lemma 8.7.1 the latter clearly implies that $[\lambda_n]_T^\prec \bar{w} \not\simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$, as required. \square

We are now ready to prove that $w \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$ provide a snapshot of our J-model.

Lemma 8.15. *Let T be any sound, \prec -introspective theory extending ACA₀. Given a finite J-frame \mathfrak{W} with root 0 and any \prec -increasing sequence λ set*

$$\overrightarrow{\text{Lim}}\{T \xrightarrow{\lambda} \mathfrak{W}\} = \langle \bar{w} \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\} : w \in W \rangle.$$

Then,

$$\overrightarrow{\text{Lim}}\{T \xrightarrow{\lambda} \mathfrak{W}\} : \mathfrak{W} \xrightarrow{\lambda} T.$$

Proof. We must check each of the conditions of Definition 8.4.

1 For the first, suppose that $w \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$ and $v \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$ hold. In view of $w \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$, pick a Solovay path s such that any extension of s has last element w and similarly s' such that any extension of s' has last element v . By Lemma 8.11.1, either $s \sqsupseteq s'$ or $s' \sqsupseteq s$; in either case, it follows by $w \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\} \wedge v \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$ that $w = v$.

2-3 The second condition is Lemma 8.13 and the third, Lemma 8.14.

4 For the fourth, we must use the fact that $[\lambda]_T^\prec$ is sound for all λ and proceed by induction on \ll_0 to show that, if s is a Solovay path and $w \neq 0$, then $s_i \neq 0$ for all i . For indeed, if $s_i = w \neq 0$ for some i , picking the minimal such i we see that i codes a λ_n -derivation of $\bar{w} \not\simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$; hence, by soundness we must have that, for some $s' \sqsupseteq s$, $\text{last}(s') \neq w$. But then, by Lemma 8.11.4, $\text{last}(s') = v \ll_0 w$, but by induction on $v \ll_0 w$ there can be no such path.

We conclude that any Solovay path is identically zero, so that the formula $\bar{0} \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\}$ is true. \square

We may finally prove our main completeness result.

Proof of Theorem 8.2. We have already seen that the logic is sound.

For the other, if ϕ is consistent over GLP_\prec , then by Lemma 7.2, $M^+(\phi) \wedge \phi$ is consistent over J and thus by Lemma 7.1, $M^+(\bar{\phi}) \wedge \bar{\phi}$ can be satisfied on a world w_* of some stratified J-model \mathfrak{W}' . Define \mathfrak{W} by adding 0 as a root to \mathfrak{W}' and let λ be the modalities appearing in ϕ . By Lemma 8.15,

$$\overrightarrow{\text{Lim}}\{T \xrightarrow{\lambda} \mathfrak{W}\} : \mathfrak{W} \xrightarrow{\lambda} T$$

so that by Lemma 8.5.1,

$$T^\prec + w_* \simeq \text{Lim}\{T \xrightarrow{\lambda} \mathfrak{W}\} \vdash f_T^\prec(\phi).$$

Hence, by lemma 8.5.2, $\mathbb{N} \models \Diamond_T f_T^\prec(\phi)$, i.e. $f_T^\prec(\phi)$ is consistent over T . \square

A Alternative provability predicates

In this section we shall briefly discuss some variants of our provability predicates. We do so in an informal setting and in particular shall refer to defining recursions rather than formalizations in second order logic. Moreover, we shall on occasion not be too concerned about the amount of transfinite induction needed in the arguments.

We could consider an apparently slightly weaker notion of α provability – let us write $[\alpha]_T^{\prec,w}$ – defined by the following recursion:

$$[\alpha]_T^{\prec,w}\phi : \Leftrightarrow \square_T\phi \vee \exists\psi \exists\beta \prec\alpha (\forall n [\beta]_T^{\prec,w}\psi(\bar{n}) \wedge [\beta]_T^{\prec,w}(\forall x\psi(x) \rightarrow \phi)).$$

However, it is easy to see by transfinite induction that $[\alpha]_T^{\prec,w}\phi \Leftrightarrow [\alpha]_T^\prec\phi$. The \Leftarrow direction is obvious. For the other direction we assume that we can formalize the notion of $[\alpha]_T^{\prec,w}$ just like $[\alpha]_T^\prec$ and prove all the necessary lemmata like monotonicity, distribution axioms, etc. Suppose that $\forall n [\beta]_T^{\prec,w}\psi(\bar{n}) \wedge [\beta]_T^{\prec,w}(\forall x\psi(x) \rightarrow \phi)$ for some formula ψ and ordinal $\beta \prec \alpha$. Then, clearly also $\forall n [\beta]_T^{\prec,w}(\psi(\bar{n}) \wedge (\forall x\psi(x) \rightarrow \phi))$. But, as $[0]_T^\prec((\forall x\psi(x) \wedge (\forall x\psi(x) \rightarrow \phi)) \rightarrow \phi)$ we get

$$\forall n [\beta]_T^{\prec,w}(\psi(\bar{n}) \wedge (\forall x\psi(x) \rightarrow \phi)) \wedge [0]((\forall x\psi(x) \wedge (\forall x\psi(x) \rightarrow \phi)) \rightarrow \phi)$$

which by the induction hypothesis for β is just $[\alpha]_T^\prec\phi$.

Note that in our definition of $[\alpha]_T^\prec$ there is still some uniformity present in that we choose one particular $\beta \prec \alpha$ with $[\beta]_T^\prec\psi(\bar{n})$ for all numbers n . We can make this β also dependent on n . In a sense, this boils down to diagonalizing at limit ordinals. Thus, we define our notion of $[\alpha]_T^{\prec,d}$ as follows:

$$[\alpha]_T^{\prec,d}\phi : \Leftrightarrow \square_T\phi \vee \exists\psi (\forall n \exists\beta_n \prec\alpha ([\beta_n]_T^{\prec,d}\psi(\bar{n}) \wedge \square(\forall x\psi(x) \rightarrow \phi))).$$

By an argument similar as before, we see that, also in this notion we can replace the $\square(\forall x\psi(x) \rightarrow \phi)$ by $\exists\gamma \prec\alpha [\gamma]_T^{\prec,d}(\forall x\psi(x) \rightarrow \phi)$ without losing any strength. This new notion of provability is related to $[\alpha]_T^\prec$ in a simple fashion as is expressed in Lemma A.2 below. Again, we assume that we can formalize the notion $[\alpha]_T^{\prec,d}$ in a suitable way so that the basic properties are provable. We first state a simple but useful observation.

Lemma A.1.

1. $[\alpha + 1]_T^\prec\phi \Leftrightarrow \exists\psi (\forall n [\alpha]_T^\prec\psi(\bar{n}) \wedge \square(\forall x\psi(x) \rightarrow \phi))$
2. $[\alpha + 1]_T^{\prec,d}\phi \Leftrightarrow \exists\psi (\forall n [\alpha]_T^{\prec,d}\psi(\bar{n}) \wedge \square(\forall x\psi(x) \rightarrow \phi))$

Proof. This follows directly from the definition and monotonicity. \square

Lemma A.2.

1. $[n]_T^\prec \phi \Leftrightarrow [n]_T^{\prec,d} \phi$ for $n \in \omega$;
2. $[\alpha + 1]_T^\prec \phi \Leftrightarrow [\alpha]_T^{\prec,d} \phi$ for $\alpha \succ \omega$.

Proof. The proofs proceed by induction on n and α , respectively, and we omit them. \square

As can be seen, there is a fair amount of freedom in defining transfinite iterations of the ω -rule. We chose the current paper's presentation both for the sake of simplicity and because a more refined hierarchy is in general terms more convenient; after all, it is easy to remove intermediate operators later if they are not needed. We also suspect it will be the appropriate notion useful later for a Π_1^0 -ordinal analysis of second-order arithmetics, a goal which now seems well within our reach.

B An afterword on the choice of our base theory

In this paper we have shown sound and completeness of the logic GLP_\prec for the interpretation where each $[\xi]$ modality is interpreted as “provable in ACA_0 using at most ξ nested applications of the omega-rule”. The main applications we have in mind with this result is to provide Π_1^0 ordinal analyses of theories much stronger than PA in the style of Beklemishev ([1]). For the mere soundness of the logic however, there were quite some strong principles needed: the existence of an iterated provability class, plus a certain amount of transfinite induction. We shall discuss here how these principles fit into the intended application of ordinal analyses.

A consistency proof of U in Elementary Arithmetic (EA) plus $\text{TI}(\hat{\Pi}_1^0, \prec)$ (using a natural ordinal notation system for large enough ordinals) is closely related to the Π_1^0 ordinal analysis of U and we shall focus our discussion on such a consistency proof. Such a consistency proof can be seen as a partial realization of Hilbert's program in the sense that over finitsitic mathematics one can prove the consistency of a strong theory with just one additional non-finitist ingredient. As such one could say that U is safeguarded by this method. If one accepts this method of safeguarding it makes philosophically speaking sense that one may use U itself as new base theory to safeguard even stronger theories by the same method. It is in this perspective that having ACA_0 as our base theory is not a bad thing since ACA_0 has already been safeguarded over EA using some amount of transfinite induction. However, for technical reasons it might be desirable to dispense with such an intermediate step.

In [11] it is noted that soundness of GLP_\prec suffices to perform a consistency proof and completeness is actually not needed. But also in the soundness proof presented in this paper we needed transfinite induction as well as resorting to the introspective closure of a theory. However, as we have seen in Corollary 5.8, for the sake of consistency-strength it is irrelevant to consider either a theory or its introspective closure as both theories are provably equiconsistent. Thus,

to conclude, let us consider the amount of transfinite induction needed in our soundness proof of GLP_\prec .

First of all, let us note that our soundness proof of GLP_\prec uses at most $\text{TI}(\Pi_1^0, \prec)$. In a sense, this is not bad at all, because it is exactly this ingredient (in parameter free form) that is added EA to perform a consistency proof of our target theory. So, by adding this amount of transfinite induction (with parameters) to EA, we get access to exactly the soundness of GLP_\prec needed to perform this consistency proof, were it not for the case that our base theory was taken to be ACA_0 , not EA. This choice of ACA_0 has been mainly to simplify our exposition and the needed amount of arithmetic can be pushed down a lot further.

We observed that to prove $\text{TI}(\Pi_1^0, \prec)$ over EA^\prec we need Σ_1^0 comprehension. But clearly Σ_1^0 comprehension with second order parameters proves ACA_0 . However, close inspection of the proofs in our paper shows that the only free set-parameters needed are occurrences of iterated provability classes. Thus if we enrich our language with a constant π with an axiom stating that π is an iterated provability class (see Lemma 5.10), we can do with parameter free comprehension since then $\text{TI}(\hat{\Pi}_1^0, \prec)$ suffices.

However, we can do better still in the sense that we need less comprehension by allowing slightly stronger well-ordering assumptions as we shall see in the next lemma.

Let us fix some bijective coding of α on the naturals, and let \prec be a primitive recursive well-order on α . Let $<$ denote the usual ordering on the natural numbers. Using a bijective pairing function we define a new relation

$$\langle \xi, n \rangle \prec' \langle \zeta, m \rangle := \xi \prec \zeta \vee (\xi = \zeta \wedge n < m).$$

Clearly, \prec' provably defines a relation of order type $\omega \cdot \alpha$. With this notation we can now state our lemma.

Lemma B.1. $\hat{\Delta}_0^0\text{-CA} + \text{wo}(\omega \cdot \alpha) \vdash \text{TI}(\hat{\Pi}_1^0, \alpha)$.

Proof. By the usual argument we see that $\hat{\Delta}_0^0\text{-CA} + \text{wo}(\omega \cdot \alpha) \vdash \text{TI}(\hat{\Delta}_0^0, \omega \cdot \alpha)$. Thus, we shall proof $\text{TI}(\hat{\Pi}_1^0, \alpha)$ using $\text{TI}(\hat{\Delta}_0^0, \omega \cdot \alpha)$. Let $\varphi(z, x)$ be some $\hat{\Delta}_0^0$ formula and assume

$$\forall x (\forall y \prec' x \forall z \varphi(z, y) \rightarrow \forall z \varphi(z, x)). \quad (12)$$

If we assume that $\forall y \prec' x \varphi(y_1, y_0)$, using (12) we get $\varphi(x_1, x_0)$ whence by $\text{TI}(\hat{\Delta}_0^0, \omega \cdot \alpha)$ we obtain $\forall x \forall z \varphi(z, x)$. \square

Note that $\omega \cdot \alpha$ is not much larger than α . In particular, if the last term in CNF of α is at least ω^ω we get that $\omega \cdot \alpha = \alpha$. Thus, for natural proof theoretical ordinals we have this equation whence we get the extra induction for free.

References

- [1] L.D. Beklemishev. Provability algebras and proof-theoretic ordinals, I. *Annals of Pure and Applied Logic*, 128:103–124, 2004.
- [2] L.D. Beklemishev. Veblen hierarchy in the context of provability algebras. In P. Hájek, L. Valdés-Villanueva, and D. Westerståhl, editors, *Logic, Methodology and Philosophy of Science, Proceedings of the Twelfth International Congress*. Kings College Publications, 2005.
- [3] L.D. Beklemishev. Kripke semantics for provability logic GLP. *Annals of Pure and Applied Logic*, 161(6):737–744, 2010.
- [4] L.D. Beklemishev. A simplified proof of the arithmetical completeness theorem for the provability logic GLP. *Trudy Matematicheskogo Instituta imeni V.A. Steklova*, 274(3):32–40, 2011. English translation: *Proceedings of the Steklov Institute of Mathematics*, 274(3):25–33, 2011.
- [5] L.D. Beklemishev, D. Fernández-Duque, and J.J. Joosten. On provability logics with linearly ordered modalities. <http://arxiv.org/abs/1210.4809>, 2012.
- [6] G.S. Boolos. *The Logic of Provability*. Cambridge University Press, Cambridge, 1993.
- [7] D. de Jongh, M. Jumelet, and F. Montagna. On the proof of solovay’s theorem. *Studia Logica*, 50:51–69, 1991.
- [8] D. Fernández-Duque and J.J. Joosten. Models of transfinite provability logics. *Journal of Symbolic Logic*, 2013. Accepted for publication, <http://arxiv.org/abs/1204.4837>.
- [9] K.N. Ignatiev. On strong provability predicates and the associated modal logics. *The Journal of Symbolic Logic*, 58:249–290, 1993.
- [10] G. Japaridze. The polymodal provability logic. In *Intensional logics and logical structure of theories: material from the Fourth Soviet-Finnish Symposium on Logic*. Metsniereba, Telavi, 1988. In Russian.
- [11] J.J. Joosten. Π_1^0 -ordinal analysis beyond first-order arithmetic. *Mathematical Communications*, 2013. Accepted for publication, <http://arxiv.org/abs/1212.2395>.
- [12] R.M. Solovay. Provability interpretations of modal logic. *Israel Journal of Mathematics*, 28:33–71, 1976.