A finitary treatment of the closed fragment of Japaridze's provability logic

Lev D. Beklemishev*, Joost J. Joosten and Marco Vervoort

June 2, 2005

Abstract

We study a propositional polymodal provability logic **GLP** introduced by G. Japaridze. The previous treatments of this logic, due to Japaridze and Ignatiev (see [11, 7]), heavily relied on some non-finitary principles such as transfinite induction up to ε_0 or reflection principles. In fact, the closed fragment of **GLP** gives rise to a natural system of ordinal notation for ε_0 that was used in [1] for a proof-theoretic analysis of Peano arithmetic and for constructing simple combinatorial independent statements.

In this paper, we study Ignatiev's universal model for the closed fragment of this logic. Using bisimulation techniques, we show that several basic results on the closed fragment of **GLP**, including the normal form theorem, can be proved by purely finitary means formalizable in elementary arithmetic. As a corollary, the system of ordinal notation for ε_0 based on the closed fragment of **GLP** is shown to be provably isomorphic to the standard system of ordinal notation up to ε_0. We also settle negatively some conjectures by Ignatiev.

1 Introduction

Giorgi Japaridze introduced a polymodal logic of provability **GLP** back in 1986 [12, 13]. The logic was formulated in a propositional language with infinitely many modalities [0], [1], [2], ..., and was supplied with a specific provability interpretation. The first modality denoted the provability predicate for Peano arithmetic PA, and the others denoted the predicates of provability by one, two, three, etc., nested applications of the ω-rule over PA, thus filling in the gap between the set of provable in PA and the set of true arithmetical sentences. Japaridze proved a highly nontrivial arithmetical completeness theorem for **GLP**, which was one of the first genuine extensions of Solovay's theorems [16] to polymodal provability logics.

Japaridze's work was simplified by Konstantin Ignatiev [11] who showed that the arithmetical completeness theorem of Japaridze holds for more general sequences of theories. Ignatiev also studied modal logical properties of **GLP**, in particular, he proved the analogs of Craig's interpolation theorem and of de Jongh–Sambin's fixed point theorem for **GLP**. He also obtained an effective normal form result and constructed a universal model for the closed fragment of **GLP**. This beautiful model will play an important role in the present paper.

The results of Japaridze and Ignatiev, albeit only for a restricted language with just two modalities, were later incorporated into a well-known textbook by George Boolos on the logic of provability [7]. Boolos himself contributed by proving that Japaridze's logic was arithmetically complete for the interpretation of the second modality as the ω-provability in second order arithmetic.

An interest in Japaridze's logic was renewed after the first author of this paper established its connection with a proof-theoretic ordinal analysis of Peano arithmetic à la Gentzen. In [4, 1] a new simple consistency proof for PA by transfinite induction based on Japaridze's logic and its models, called *graded provability algebras*, was given. The closed fragment of **GLP** under the natural ordering relation 'φ implies consistency of ψ' played the role of a suitable ordinal notation system for the ordinal ε_0. In [5], an interesting combinatorial independent principle suggested by Japaridze's logic was studied and was shown to be independent from PA. In [3], an extension of Japaridze's logic defining an ordinal notation system up to a larger proof-theoretic ordinal Γ_0 was introduced. Thus, methods of provability logic provided a new and interesting type of 'natural' ordinal notation systems and contributed to a better understanding of relationships between formal theories and their proof-theoretic ordinals.

This paper is devoted to a question which has become sufficiently important in view of the general methodology of applying provability logic to proof-theoretic analysis of theories such as PA. The system of ordinal notation based on the closed fragment of **GLP** was shown in [1] to be order-isomorphic to the standard system of ordinal notation for ε_0 based on Cantor normal forms. However, a relatively simple proof of that theorem given in [1] used the arithmetical soundness of Japaridze's logic and the validity of all formulas $\neg[n]\bot$ under the arithmetical interpretation. The latter fact is proof-theoretically strong, that is, not formalizable in PA itself (see Lemma 2.2 below). Hence, it was not a priori clear if the two systems of ordinal notation were provably equivalent in PA, and hence if Gentzen's and Beklemishev's results were really the same.

Similar kinds of problems were already known from Ignatiev's work. In fact, as Ignatiev noted in [10], 'all reasoning [in his paper] about the closed fragment of **GLP** uses ε_0-induction and thus *cannot* be formalized in PA, but only in PA $\cup$ {the ε_0-induction schema}. Since this theory has the same theorems as PA$'$,[1] a natural conjecture is that several modal properties of **GLP**, such as the

---

[1] Ignatiev calls PA$'$ the extension of PA by its local reflection principle. Actually, by a result of Kreisel and Lévy [15] the ε_0-induction schema is equivalent to a stronger *uniform* reflection principle, which would be the correct reading of that passage.

Craig interpolation property, cannot be proved in PA. However, the author has no idea how to prove it.'

In this paper we shall examine the situation concerning the closed fragment of **GLP**, which is particularly interesting from the point of view of the applications to proof theory. The problem of provable isomorphism of the two systems of ordinal notation for ε_0 turns out to be equivalent to a purely modal logical question about **GLP**: Is the consistency of the extension of **GLP** by formulas of the form $\neg[n]\bot$, for all n, provable in PA? Notice that we are talking here about consistency of a decidable propositional logic. Surprisingly enough, the two different consistency proofs known so far relied on some principles outside PA. An obvious proof by the arithmetical interpretation of **GLP** used the uniform reflection principle. An alternative proof based on Ignatiev's model — the only known Kripke model for the logic in question — used transfinite induction. In fact, Ignatiev's model is a converse well-founded ordering of height ε_0, and the standard interpretation of Löb's axiom in a Kripke model simply *means* converse well-foundedness.

Despite this fact, we show that no strong assumptions are necessary here, and the consistency of the above logic can be established by purely finitary means formalizable in elementary arithmetic EA (or $I\Delta_0+\exp$ [9]). Main results on the closed fragment of Japaridze's logic necessary for the proof-theoretic analysis of PA can also be obtained by finitary means. Our proof is based on the analysis of Ignatiev's model and a realization that transfinite induction for *definable* subsets of the model is a proof-theoretically much weaker principle. We present a proof based on semantical bisimulation arguments (see also Joosten [14]). (An alternative more syntactic proof based on refinements of a normal form theorem is also possible, but it is more cumbersome.) We conclude that the systems of ordinal notation based on **GLP** and on Cantor normal forms are provably isomorphic, and hence Gentzen's original result is captured by the provability algebraic approach.[2]

2   Preliminaries

The axioms of GLP.   From the point of view of a finitary treatment there are two problematic axiom schemes in **GLP**: Löb's axiom and the montonicity axiom. Therefore we introduce some subsystems of the closed fragment of **GLP**. As our basic system **I** we choose a polymodal version of **K4** given by the following rules and axiom schemes:

Axioms: (i) Boolean tautologies;

(ii) $[x](\varphi \to \psi) \to ([x]\varphi \to [x]\psi)$;

(iii) $[x]\varphi \to [y][x]\varphi$, for $x \leq y$;

---

[2]It was shown in [2] that using some tricks one can avoid the above mentioned problem of equivalence of ordinal notation systems and directly obtain Gentzen's result for the standard ordinal notation system from the treatment of Peano arithmetic based on **GLP**.

(iv) $\langle x\rangle\varphi \to [y]\langle x\rangle\varphi$, for $x < y$.

Rules: modus ponens, $\vdash \varphi \;\Rightarrow\; \vdash [x]\varphi$.

As usual, $\langle x\rangle\varphi$ abbreviates $\neg[x]\neg\varphi$.

The system $\mathbf{GLP}^-$, denoted LN by Ignatiev for a language with N modalities, is obtained from $\mathbf{I}$ by adding *Löb's axiom*:

$$[x]([x]\varphi \to \varphi) \to [x]\varphi.$$

Japaridze's logic $\mathbf{GLP}$ is obtained from $\mathbf{GLP}^-$ by adding the *monotonicity axiom*:

$$[y]\varphi \to [x]\varphi, \quad \text{for } y \leq x.$$

We notice an obvious property of all these systems, which we call *shift invariance*. Let $x \uparrow \varphi$ be the result of replacing in φ every modality $[y]$ by $[x+y]$. Then $\mathbf{GLP} \vdash \varphi$ implies $\mathbf{GLP} \vdash (x \uparrow \varphi)$, for any formula φ. The converse implication also holds, which follows from the work of Ignatiev.

Arithmetical interpretation. Fix an r.e. theory T in the language of Peano arithmetic containing a sufficiently strong fragment of PA such as Kalmar elementary arithmetic EA. Let $[0]_T$ denote a Σ_1-provability predicate for T, and $[n]_T$ a Σ_{n+1}-provability predicate for the extension of T by all true arithmetical Π_n-sentences. A standard arithmetical interpretation of the polymodal language, somewhat different from the original interpretation by Japaridze, reads $[n]$ as $[n]_T$, for each n. An arithmetical realization is a finite function f mapping propositional variables to arithmetical sentences. Let $f_T(\varphi)$ denote the result of substituting arithmetical sentences for the variables of φ according to f and translating the modalities as above. It is well known (and a matter of routine checking) that all axioms of $\mathbf{GLP}$ are valid under this interpretation.

Lemma 2.1 $\mathbf{GLP} \vdash \varphi$ *implies* $\mathsf{EA} \vdash f_T(\varphi)$, *for any arithmetical realization f of the variables of φ.*

Japaridze [12] and Ignatiev [11] showed that the converse implication also holds provided T is arithmetically sound, that is, if all theorems of T are true in the standard model.

We note that the condition of soundness of T is equivalent to the truth of all formulas of the form $\langle n\rangle_T\top$. Indeed, if T proves a false Σ_n-sentence, for some n, then the extension of T by all true Π_n-sentences is contradictory. Hence, $\langle n\rangle_T\top$ is false. On the other hand, if T is sound, then the extension of T by all true Π_n-sentences is also sound, and hence it is consistent.

Soundness of T is formally expressed by the *uniform reflection principle* for T, that is, the schema

$$\mathsf{RFN}(T): \qquad \forall x\, ([0]_T\varphi(\dot{x}) \to \varphi(x)),$$

for all formulas φ. The previous argument can be formalized in EA and yields the following lemma (see [1]).

4

Lemma 2.2 *Over* EA, *the schema* $\{\langle n \rangle_T \top : n < \omega\}$ *is equivalent to* RFN(T).

Hence, $\{\langle n \rangle_{\mathsf{EA}} \top : n < \omega\}$ is equivalent to the uniform reflection principle for EA. Kreisel and Lévy [15] have shown that EA + RFN(EA) is equivalent to PA.

Let $\mathbf{GLP}^*$ denote an extension of $\mathbf{GLP}$ by all axioms of the form $\{\langle n \rangle \top : n < \omega\}$ and with modus ponens as the only rule of inference.

Proposition 2.3 $\mathbf{GLP}^*$ *is consistent.*

Proof. By induction on the length of a $\mathbf{GLP}^*$-derivation it is easy to show that, if $\mathbf{GLP}^* \vdash \varphi$, then PA $\vdash f_{\mathsf{EA}}(\varphi)$, for any realization f of the variables of φ. Indeed, all formulas of the form $f_{\mathsf{EA}}(\langle n \rangle \top)$ are provable in PA, by Lemma 2.2, and Lemma 2.1 accounts for the axioms of $\mathbf{GLP}$. Since PA is consistent, $\bot$ is not derivable in $\mathbf{GLP}^*$. ⊠

Of course, this simple argument uses a strong assumption of consistency of PA. In the present paper we shall give a different proof of this fact which is formalizable in EA.

The closed fragment of GLP. Let S denote the set of all words over the alphabet $\mathbb{N}$, including the empty word Λ. S_x will denote the set of words over the alphabet $\{y \in \mathbb{N} : x \leq y\}$. To each element $\alpha = x_1 x_2 \ldots x_k$ of S we associate its *modal interpretation*, that is, the closed modal formula

$$\langle x_1 \rangle \langle x_2 \rangle \cdots \langle x_k \rangle \top, \tag{1}$$

We do not distinguish between the word α and formula (1). We also identify Λ with $\top$.

For each $x \in \mathbb{N}$ there is a natural ordering $<_x$ on S_x defined by

$$\alpha <_x \beta \iff \mathbf{GLP} \vdash \beta \to \langle x \rangle \alpha.$$

It is immediately seen that $<_x$ is transitive. For any x, we have the following lemma (provably in EA).

Lemma 2.4 $<_x$ *is irreflexive if and only if* $\mathbf{GLP}^*$ *is consistent.*

Proof. In view of Löb's axiom, $\alpha <_x \alpha$ iff $\mathbf{GLP} \vdash \alpha \to \langle x \rangle \alpha$ iff $\mathbf{GLP} \vdash \neg \alpha$. Let n be greater than any letter occurring in α. We verify by induction on the length of α that

$$\mathbf{GLP} \vdash \langle n \rangle \top \to \langle n \rangle \alpha.$$

Indeed, the statement is obvious if $\alpha = \top$. If $\alpha = \langle m \rangle \beta$, then, by the induction hypothesis,

$$
\begin{aligned}
\mathbf{GLP} \vdash \langle n \rangle \top \quad &\to \quad \langle n \rangle \beta \\
&\to \quad \langle n \rangle \top \wedge \langle m \rangle \beta, \quad \text{by the monotonicity axiom} \\
&\to \quad \langle n \rangle \top \wedge [n] \langle m \rangle \beta, \quad \text{by Axiom (iv)} \\
&\to \quad \langle n \rangle \langle m \rangle \beta, \quad \text{by Axiom (ii).}
\end{aligned}
$$

If $\mathbf{GLP} \vdash \neg\alpha$, then $\mathbf{GLP} \vdash [n]\neg\alpha$ and hence $\mathbf{GLP} \vdash [n]\bot$. This yields the inconsistency of $\mathbf{GLP}^*$. Conversely, if $\mathbf{GLP}^*$ is inconsistent, then by the deduction theorem $\mathbf{GLP} \vdash [n]\bot$, for some n. Then for $\alpha = \langle n \rangle \top$ one has $\alpha <_x \alpha$, for any x, hence $<_x$ is not irreflexive. $\boxtimes$

We conclude that the orderings $<_x$ are indeed irreflexive, but our current proof of this fact relies on the assumption of consistency of PA, in view of Proposition 2.3.

We formulate a normal form result for the closed fragment of $\mathbf{GLP}$. We first define a subset $NF \subset S$ of *normal forms*. Let $w(\alpha)$, *width* of α, denote the number of different letters occurring in α. Normal forms are defined by recursion on their width.

- Λ and any word of width 1 belongs to NF.

- Assume $w(\alpha) > 1$ and let $x = \min(\alpha)$. Then graphically $\alpha = \alpha_1 x \cdots x \alpha_k$, where all α_i do not contain x and hence $w(\alpha_i) < w(\alpha)$ for $1 \leq i \leq k$. Then $\alpha \in NF$ iff all $\alpha_i \in NF$ and, for all $1 \leq i < k$, $\alpha_{i+1} \not<_{x+1} \alpha_i$. (Note that $\alpha_i \in S_{x+1}$.)

The following propositions are an alternative version of Ignatiev's normal form theorem (see [1, 3]).

Proposition 2.5 *Every word is equivalent in $\mathbf{GLP}$ to a unique word in a normal form.*

Proposition 2.6 *Every closed formula is equivalent in $\mathbf{GLP}$ to a boolean combination of words.*

Proposition 2.7 below (see [1]) implies that the ordering $(S, <_0)$ is well-founded and isomorphic to ε_0 modulo equivalence in $\mathbf{GLP}$. We define a natural function o mapping words to ordinals below ε_0.

Let $o(0^k) = k$. If $\alpha = \alpha_1 0 \alpha_2 0 \cdots 0 \alpha_n$, where all $\alpha_i \in S_1$ and not all of them empty, then recursively define

$$o(\alpha) = \omega^{o(\alpha_n^-)} + \cdots + \omega^{o(\alpha_1^-)},$$

where β^- is obtained from $\beta \in S_1$ by replacing every letter $m+1$ by m.

Proposition 2.7 *For all $\alpha, \beta \in S$,*

$$\mathbf{GLP} \vdash \alpha \leftrightarrow \beta \quad iff \quad o(\alpha) = o(\beta);$$
$$\mathbf{GLP} \vdash \beta \to \Diamond\alpha \quad iff \quad o(\alpha) < o(\beta).$$

Notice that by the shift invariance of $\mathbf{GLP}$

$$\alpha <_0 \beta \quad \Rightarrow \quad x \uparrow \alpha <_x x \uparrow \beta. \tag{2}$$

The converse implication also holds, and in fact $x \uparrow \cdot$ is an isomorphism of $(S_0, <_0)$ onto $(S_x, <_x)$ (see Corollary 7 in [3]).

Ignatiev's model.   One of the aims of this paper is to popularize a beautiful Kripke model $\mathcal{U}$ for the closed fragment of **GLP** constructed by Ignatiev. The model is an infinite structure equipped with infinitely many irreflexive ordering relations R_0, R_1, R_2, ..., which has a remarkable self-similar (or 'fractal') kind of symmetry.

First, we introduce some notation. If $\mathcal{K}$ is a structure with the above signature, let $\mathcal{K}^+$ denote the structure with the same universe whose relation R_0 is empty and whose relation R_{n+1} coincides with R_n on $\mathcal{K}$, for each $n < \omega$.

Every ordinal $\alpha > 0$ can be written in the form $\alpha = \beta + \omega^\gamma$. By Cantor's normal form theorem, γ is uniquely defined. We denote γ by $d(\alpha)$ and stipulate $d(0) = 0$.

We define $\mathcal{U}$ as a direct limit of structures $\mathcal{U}_\alpha$ constructed by stages $\alpha < \varepsilon_0$. Thus, we simultaneously define $\mathcal{U}_\alpha$ and the canonical embeddings $i_{\beta\alpha} : \mathcal{U}_\beta \to \mathcal{U}_\alpha$, for all $\beta < \alpha$, by induction on α.

$\mathcal{U}_0$ consists of a single point 0 with empty relations. Assume all $\mathcal{U}_\beta$, for $\beta < \alpha$, are constructed. Let $\mathcal{V}_\alpha := \varinjlim \{ \mathcal{U}_\beta : \beta < \alpha \}$ and let $\mathcal{W}_\alpha := \mathcal{U}_{d(\alpha)}^+$. Then $\mathcal{U}_\alpha$ is defined to be the disjoint union of $\mathcal{V}_\alpha$ and $\mathcal{W}_\alpha$, with all the inherited ordering relations. Besides, R_0 on $\mathcal{U}_\alpha$ is extended by putting any element of $\mathcal{W}_\alpha$ below any element of $\mathcal{V}_\alpha$, that is, xR_0y whenever $x \in \mathcal{W}_\alpha$ and $y \in \mathcal{V}_\alpha$.

The embeddings $i_{\beta\alpha}$, for $\beta < \alpha$, are defined in an intuitively obvious way as follows. Let $j_\alpha : \mathcal{V}_\alpha \hookrightarrow \mathcal{U}_\alpha$ and $h_\alpha : \mathcal{W}_\alpha \hookrightarrow \mathcal{U}_\alpha$ denote the canonical embeddings associated with the disjoint union, and let $j_{\beta\alpha} : \mathcal{U}_\beta \to \mathcal{V}_\alpha$ be the embedding associated with the direct limit. Then $i_{\beta\alpha} = j_\alpha \circ j_{\beta\alpha}$ is the required embedding of $\mathcal{U}_\beta$ into $\mathcal{U}_\alpha$. This completes the inductive definition of $\mathcal{U}_\alpha$ and of $\mathcal{U} := \varinjlim \{ \mathcal{U}_\alpha : \alpha < \varepsilon_0 \}$.

From now on we can think of $\mathcal{U}_\alpha$ as being a substructure of $\mathcal{U}$; direct limits then simply become unions. The nodes of $\mathcal{W}_\alpha$, that is, the nodes from $\mathcal{U}_\alpha \setminus \bigcup_{\beta < \alpha} \mathcal{U}_\beta$ will be called the nodes *introduced at stage* α. Since $\mathcal{W}_\alpha$ is isomorphic to $\mathcal{U}_{d(\alpha)}^+$, there is a function $p : \mathcal{U} \to \mathcal{U}$ mapping a node x introduced at stage α to the corresponding point in $\mathcal{U}_{d(\alpha)}$; we let $p(0) = 0$.

We also inductively define the *root* $\hat{\alpha}$ of $\mathcal{U}_\alpha$. By definition, $\hat{0} = 0$, and $\hat{\alpha}$ is the node of $\mathcal{W}_\alpha$ corresponding to the root of $\mathcal{U}_{d(\alpha)}$, in other words, the unique node $x \in \mathcal{W}_\alpha$ such that $p(x) = \widehat{d(\alpha)}$. The model $\mathcal{U}$ has no root.

The *main axis* $M \subset \mathcal{U}$ is the set of all nodes of the form $\hat{\alpha}$, for $\alpha < \varepsilon_0$. This set plays a special role in the treatment of $\mathcal{U}$. It is linearly ordered by R_0 and has order type ε_0^* (the converse of ε_0).

$\mathcal{U}$ can be considered as a Kripke frame for the language of **GLP**. The validity of formulas is defined as usual, in particular,

$$\mathcal{U}, u \Vdash [n]\varphi \iff \forall v \in \mathcal{U} \, (uR_nv \Rightarrow \mathcal{U}, v \Vdash \varphi),$$

for any φ. It is easy to see that the axioms of **GLP**$^-$ are valid in $\mathcal{U}$. Indeed, R_n are converse well-founded ordering relations satisfying the additional condition

$$\forall x, y \, (xR_ny \Rightarrow \forall z \, (xR_kz \Leftrightarrow yR_kz)), \quad \text{if } k < n.$$
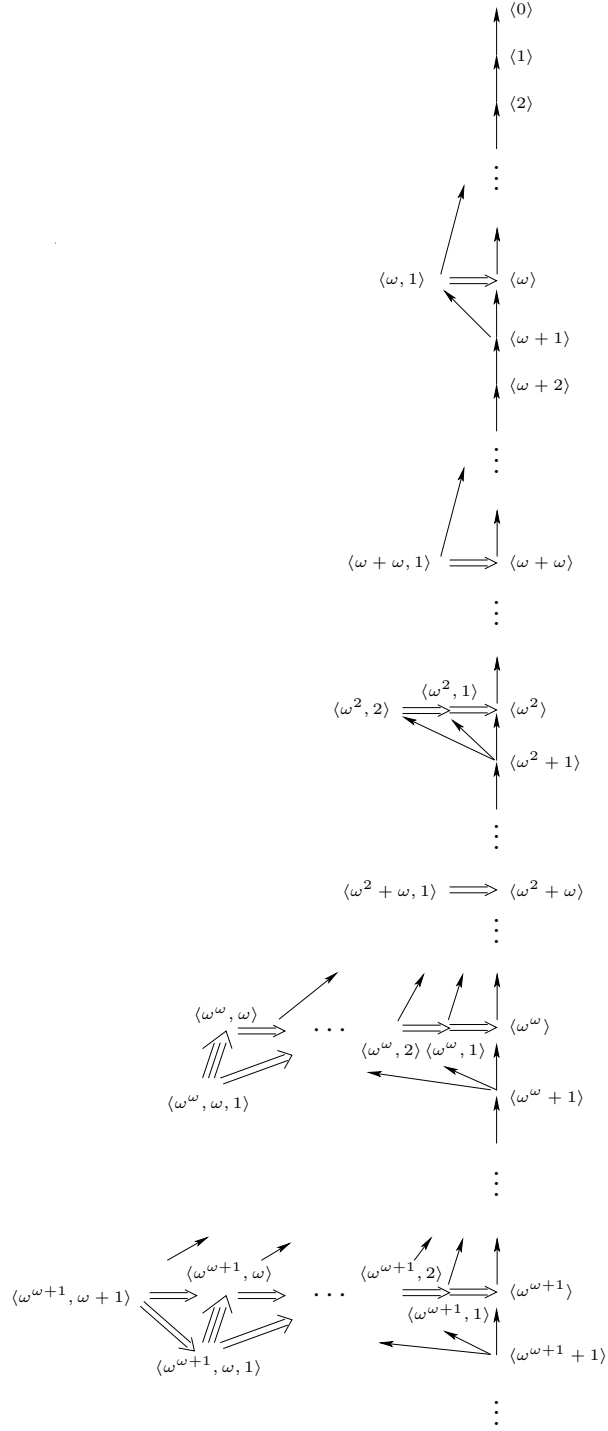
7

Figure 1: The universal model for **GLP**$_0$.

8

This condition guarantees the validity of axioms (iii) and (iv) of $\mathbf{GLP}^-$. However, it is notably *not* the case that $R_n \subseteq R_k$ for $k < n$. Hence, the monotonicity axiom fails in $\mathcal{U}$ for a suitable evaluation of variables.

Ignatiev has established (using strong assumptions) that $\mathcal{U}$ is not only a model but also a universal model for the *closed fragment* of $\mathbf{GLP}$.

Proposition 2.8 *If φ is a closed modal formula, then*

$$\mathbf{GLP} \vdash \varphi \iff \mathcal{U}, u \Vdash \varphi, \quad \text{for all } u \in \mathcal{U}.$$

In fact, Ignatiev shows that any closed formula unprovable in $\mathbf{GLP}$ is false at some node of the main axis of $\mathcal{U}$. A strong completeness of $\mathcal{U}$, also established by Ignatiev, is that there is a canonical bijection between nodes of $\mathcal{U}$ and maximal $\mathbf{GLP}$-consistent sets of closed formulas. With a node $x \in \mathcal{U}$ we associate the set of all closed formulas valid at x. Thus, $\mathcal{U}$ can be seen as a canonical model for the closed fragment of $\mathbf{GLP}$.

To give a recursive presentation of $\mathcal{U}$ we associate with every node $x \in \mathcal{U}$ an infinite sequence $\vec{\alpha} = (\alpha_0, \alpha_1, \ldots)$ of ordinals $< \varepsilon_0$ as follows.

Let α_0 be the first ordinal α such that $x \in \mathcal{U}_\alpha$. Since α_0 is the first one, $x \notin \bigcup_{\beta < \alpha_0} \mathcal{U}_\beta$, therefore $x \in \mathcal{W}_\alpha$. Let $x_1 = p(x)$ and let α_1 be the first ordinal α such that $x_1 \in \mathcal{U}_\alpha$. The process can be continued. We notice that, by the construction of $\mathcal{U}$,

$$\alpha_1 \leq d(\alpha_0), \quad \alpha_2 \leq d(\alpha_1), \quad \ldots \tag{3}$$

Hence, the chain of ordinals strictly decreases until 0 is reached. If $\alpha_m = 0$, then $\alpha_n = 0$, for all $n \geq m$. Hence, the sequence $(\alpha_0, \alpha_1, \ldots)$ ends with an infinite tail of zeros.

It is not difficult to see that such a sequence $\vec{\alpha}$ uniquely characterizes an element $x \in \mathcal{U}$. Indeed, any element $x \in \mathcal{U}$ is characterized by the least α such that $x \in \mathcal{U}_\alpha$ and by the position of x in the model $\mathcal{W}_\alpha \simeq \mathcal{U}_{d(\alpha_0)}^+$. This position is, by the induction hypothesis, uniquely characterized by the subsequence $(\alpha_1, \alpha_2, \ldots)$. It is also easy to see that to every sequence satisfying (3) one can find a corresponding element of the model.

Notice that, for any $\alpha < \varepsilon_0$, the sequence $\hat{\alpha} = (\alpha, d(\alpha), d(d(\alpha)), \ldots)$ represents the root of $\mathcal{U}_\alpha$. The set of all such sequences constitutes the main axis of $\mathcal{U}$.

See Figure 1 for a drawing of an upper part of $\mathcal{U}$. On this picture, R_0 is the transitive closure of the relation represented by single arrows. Double and triple arrows represent R_1 and R_2, in a similar way. All sequences assigned to the nodes end up with an infinite tail of zeros which is not shown. It is useful to think of this picture as being embedded into a three-dimensional space in which 0-arrows connect the 'planes' of points connected by 1- and 2-arrows, and in every such plain 1-arrows connect the 'lines' of points connected by 2-arrows, etc. In general, the whole of $\mathcal{U}$ can be thought of as being embedded into an ω-dimensional space, in which the infinite sequences of ordinals associated with the points represent the coordinates.

9

The accessibility relations can be easily expressed in terms of such co-ordinates. Assume $x, y \in \mathcal{U}$ with the corresponding sequences $\vec{\alpha}$ and $\vec{\beta} = (\beta_0, \beta_1, \ldots)$.

Lemma 2.9 xR_ky *holds iff* $\forall i < k \; \alpha_i = \beta_i$ *and* $\alpha_k > \beta_k$.

Proof. We prove the statement by induction on k. Consider $k = 0$. Nodes of $\mathcal{U}$ introduced at the same stage are never connected by R_0. Therefore, xR_0y iff y has been introduced at an earlier stage than x. Hence, xR_0y iff $\alpha_0 > \beta_0$.

Consider $k > 0$. xR_ky holds iff x and y have been introduced at the same stage $\gamma = \alpha_0 = \beta_0$ and $x_1R_{k-1}y_1$, where $x_1 = p(x)$ and $y_1 = p(y)$. The sequences associated with x_1 and y_1 are $(\alpha_1, \alpha_2, \ldots)$ and $(\beta_1, \beta_2, \ldots)$, respectively. By the induction hypothesis, $x_1R_{k-1}y_1$ holds iff $\forall i \, (0 < i < k \;\Rightarrow\; \alpha_i = \beta_i)$ and $\alpha_k > \beta_k$. Putting this together with $\alpha_0 = \beta_0$ yields the result. $\boxtimes$

Thus, $\mathcal{U}$ can be given the following isomorphic presentation. Let Ω denote the set of all ω-sequences $\vec{\alpha} = (\alpha_0, \alpha_1, \ldots)$ of ordinals below ε_0.

$$\begin{aligned} \mathcal{U}_\alpha &= \{\vec{\alpha} \in \Omega : \alpha_0 = \alpha \text{ and } \forall k < \omega \; \alpha_{k+1} \leq d(\alpha_k)\}; \\ xR_ky &\Leftrightarrow (\forall i < k \; \alpha_i = \beta_i \text{ and } \alpha_k > \beta_k). \end{aligned}$$

This presentation of $\mathcal{U}$ is, in fact, elementary recursive. An infinite sequence of ordinals satisfying (3) is uniquely determined by its maximal non-zero piece. Hence, the elements of $\mathcal{U}$ can also be represented by finite sequences of non-zero ordinals $< \varepsilon_0$ satisfying condition (3). Since the ordinals below ε_0 can be coded in an elementary way by their Cantor normal forms, and the coding of sequences is available in EA, the predicates $x \in \mathcal{U}$, $x \in \mathcal{U}_\alpha$, xR_ny can be defined in a natural way by bounded formulas in EA.

Despite the fact that $\mathcal{U}$ has an elementary presentation, we remark that the validity of Löb's axiom in $\mathcal{U}$ is equivalent to the converse well-foundedness of an ordering of height ε_0. Hence, it cannot be established in (a conservative second-order extension of) Peano arithmetic. This is precisely the source of difficulty with the proof of consistency of $\mathbf{GLP}^*$ via Ignatiev's model. A way out is to realize that one actually only needs to prove the validity of the set of *closed instances* of Löb's axiom in $\mathcal{U}$. This fact is much weaker proof-theoretically, but it requires some analysis of definable sets of nodes in $\mathcal{U}$.

3 Bisimulations and a soundness proof

In this section we give a proof of the consistency of $\mathbf{GLP}^*$. Our proof will make an extensive use of the standard notion of n-bisimilarity (see e.g. [6]). Let a Kripke model $\mathcal{A}$ (in the language of $\mathbf{GLP}$) be given. We define n-bisimilarity equivalence relations $\simeq_n$ on $\mathcal{A}$, for each $n < \omega$, by induction on n.

Definition 3.1 (n-bisimilarity) *Let $x, x' \in \mathcal{A}$.*

- $x \simeq_0 x'$ *if x and x' force the same variables.*

- x, x' *are* $n + 1$-*bisimilar if*

  - $x \simeq_n x'$;
  - $\forall i {<} \omega \, \forall y \; (x R_i y \Rightarrow \exists y' \; (x' R_i y' \; \& \; y \simeq_n y'))$;
  - $\forall i {<} \omega \, \forall y' \; (x' R_i y' \Rightarrow \exists y \; (x R_i y \; \& \; y' \simeq_n y))$.

Let $dp(\varphi)$ denote the modality depth of φ, that is, the maximal number of nested modalities in φ. The following lemma is standard.

Lemma 3.2 *For any* $x, y \in \mathcal{A}$, *if* $x \simeq_n y$ *then*

$$x \Vdash \varphi \Leftrightarrow y \Vdash \varphi,$$

for every φ *with* $dp(\varphi) \leq n$.

Proof. By an easy induction on n with a subsidiary induction on the length of φ. $\boxtimes$

We are going to study the n-bisimilarity relation on Ignatiev's model $\mathcal{U}$ for the closed fragment of **GLP**. Notice that, for a language without variables, all nodes of $\mathcal{U}$ are 0-bisimilar.

Lemma 3.3 *For any ordinals* $\alpha, \beta < \varepsilon_0$,

$$(i) \quad \beta < \alpha \; \& \; \gamma < d(\alpha) \Rightarrow \beta + \omega^\gamma < \alpha;$$
$$(ii) \quad \beta < \alpha \; \& \; \gamma \leq d(\alpha) \Rightarrow \beta + \omega^\gamma \leq \alpha.$$

Proof. By elementary observations on the Cantor normal forms of α and β. $\boxtimes$

Definition 3.4 *An ordinal* α *has* nested width *at most* n, *we write* $nw(\alpha) \leq n$, *if the Cantor normal form of* α *contains at most* n *terms, and each exponent of each term has nested width at most* n.

We note that for each α there is an n with $nw(\alpha) \leq n$, and that $nw(\alpha) \leq n$ implies $nw(\alpha) \leq n + 1$. As all our ordinals are below ε_0, we see that, for any p and α, there are only finitely many $\beta < \alpha$ with $nw(\beta) \leq p$. Hence, we can define alternative fundamental sequences as follows:

$$\alpha\langle p \rangle := \max\{\xi < \alpha \mid nw(\xi) \leq p\}.$$

We also stipulate $0\langle p \rangle = -1$, where -1 is smaller than any ordinal. It is immediate that $\forall p \; \alpha\langle p \rangle < \alpha$, for each α. If α is a limit ordinal, we have $\sup_{p<\omega} \alpha\langle p \rangle = \alpha$. We also note that $\alpha\langle p \rangle$ is monotone in both α and p.

The following is our main technical lemma.

Lemma 3.5 *Let* $\vec{\alpha}, \vec{\beta} \in \mathcal{U}$ *be such that, for all* i,

$$\alpha_i > \beta_i\langle p \rangle \quad and \quad \beta_i > \alpha_i\langle p \rangle.$$

Then, $\vec{\alpha} \simeq_p \vec{\beta}$.

Proof. By induction on p. The case $p = 0$ is trivial, so let us consider the induction step. We assume that for all i, $\alpha_i > \beta_i\langle p+1\rangle$ and $\beta_i > \alpha_i\langle p+1\rangle$. We have to see that we can reply any R_n-step from $\vec{\alpha}$ to $\vec{\alpha'}$ with an R_n-step from $\vec{\beta}$ to a $\vec{\beta'}$ so that $\vec{\alpha'} \simeq_p \vec{\beta'}$ and vice versa. By symmetry, it is sufficient to only consider one step.

So, we assume $\vec{\alpha}R_n\vec{\alpha'}$. Then, ignoring the tails of 0s, $\vec{\alpha}$ and $\vec{\alpha'}$ have the form

$$
\begin{aligned}
\vec{\alpha} &= \alpha_0, \cdots, \alpha_{n-1}, \alpha_n, \cdots && \text{and} \\
\vec{\alpha'} &= \alpha_0, \cdots, \alpha_{n-1}, \alpha'_n, \cdots, \alpha'_m.
\end{aligned}
$$

We now reply this step in $\vec{\beta}$ by going to a $\vec{\beta'}$ with the same length as $\vec{\alpha'}$ defined as follows.

$$
\begin{aligned}
\beta'_m &:= \alpha'_m\langle p\rangle + 1, \\
\beta'_k &:= \alpha'_k\langle p\rangle + \omega^{\beta'_{k+1}}, && \text{for } n \le k < m; \\
\beta'_k &:= \beta_k, && \text{for } 0 \le k < n.
\end{aligned}
$$

We now have to check two things. First, we have to see that $\vec{\beta'} \in \mathcal{U}$ and $\vec{\beta}R_n\vec{\beta'}$. Second, we should see that indeed $\vec{\alpha'} \simeq_p \vec{\beta'}$.

Thus, we should check that, for all i, $\beta'_{i+1} \le d(\beta'_i)$ and, moreover, $\beta'_n < \beta_n$. Now, $\beta'_{i+1} \le d(\beta'_i)$ is easy to see once we know that, indeed, $\beta'_n < \beta_n$. In order to check that $\beta'_n < \beta_n$, we reason as follows. By induction and using Lemma 3.3, we see that

$$
\alpha'_i \ge \beta'_i, \quad \text{for} \quad n \le i \le m. \tag{$*$}
$$

Combining this with $\vec{\alpha}R_n\vec{\alpha'}$ we obtain that $\beta'_n \le \alpha'_n < \alpha_n$. By an easy induction we see that $nw(\beta'_i) \le p + 1$, for $n \le i \le m$. Thus, $\alpha_n\langle p+1\rangle \ge \beta'_n$. Combining this with our assumption that $\beta_n > \alpha_n\langle p+1\rangle$, we conclude $\beta_n > \alpha_n\langle p+1\rangle \ge \beta'_n$.

Now we prove $\vec{\alpha'} \simeq_p \vec{\beta'}$. In view of the induction hypothesis, it is sufficient to show that, for all i,

$$
\begin{aligned}
(a) &\quad \alpha'_i > \beta'_i\langle p\rangle, \\
(b) &\quad \beta'_i > \alpha'_i\langle p\rangle.
\end{aligned}
$$

Let us first prove (a). By $(*)$ we see that $\alpha'_i \ge \beta'_i > \beta'_i\langle p\rangle$, for $n \le i \le m$. For $i < n$, we have $\alpha'_i = \alpha_i > \beta_i\langle p+1\rangle \ge \beta_i\langle p\rangle = \beta'_i\langle p\rangle$.

For $n \le i \le m$, (b) follows by the definition of $\vec{\beta'}$. Clearly, $\beta'_m = \alpha'_m\langle p\rangle + 1 > \alpha'_m\langle p\rangle$. For $n \le i < m$, either $\omega^{\beta'_{i+1}} \le \alpha'_i\langle p\rangle$ whence $\beta'_i = \alpha'_i\langle p\rangle + \omega^{\beta'_{i+1}} > \alpha'_i\langle p\rangle$, or $\omega^{\beta'_{i+1}} > \alpha'_i\langle p\rangle$ whence $\beta'_i = \alpha'_i\langle p\rangle + \omega^{\beta'_{i+1}} \ge \omega^{\beta'_{i+1}} > \alpha'_i\langle p\rangle$. For $i < n$ we reason as in (a).

The induction hypothesis now yields $\vec{\alpha'} \simeq_p \vec{\beta'}$, and we are done. $\boxtimes$

We remark that this lemma provides a sufficient but not a necessary condition for p-bisimilarity. Clearly, $\hat{0} \simeq_0 \hat{1}$ but not $0 > 1\langle 0\rangle = 0$.

The next lemma is an immediate corollary of the previous one.

Lemma 3.6 *Assume an $\vec{\alpha} \in \mathcal{U}$ and an $n < \omega$ are fixed. Let $\vec{\beta}$ be defined by backwards recursion as follows:*

$$\beta_k := \begin{cases} 0 & \text{if } \alpha_k = 0; \\ \alpha_k \langle p \rangle + \omega^{\beta_{k+1}} & \text{if } n \leq k \text{ and } \alpha_k \neq 0; \\ \alpha_k & \text{if } k < n. \end{cases}$$

Then, $\vec{\alpha} \simeq_p \vec{\beta}$, $\vec{\alpha} R_n \vec{\beta}$ and $\forall k \geq n \; nw(\beta_k) \leq p + 1$.

Proof. The sufficient condition from the previous lemma is easy to check by backwards induction, similarly to the argument in the lemma itself. The condition concerning the nested width of ordinals β_k is clearly satisfied. Using Lemma 3.3 it is easy to check that $\beta_k \leq \alpha_k$, for all $k > n$. Hence, $\beta_n < \alpha_n$ and therefore $\vec{\alpha} R_n \vec{\beta}$. ⊠

Corollary 3.7 *Any point $\vec{\alpha} \in \mathcal{U}$ is p-bisimilar to some point $\vec{\beta}$ on the main axis.*

Proof. In the previous lemma, take $n = 0$. ⊠

Turning to a proof of consistency of $\mathbf{GLP}^*$ we first note an obvious lemma.

Lemma 3.8 *If φ is a closed formula that is provable in $\mathbf{GLP}$, then there is a proof of φ in which only closed formulas occur.*

Proof. If such a proof contains propositional variables, we may substitute $\top$ for them and obtain the desired proof. ⊠

We shall denote by $\mathbf{GLP}_0$ the system $\mathbf{GLP}$ where the axiom schemes are restricted to closed formulas. By the above lemma, $\mathbf{GLP}$ is conservative over $\mathbf{GLP}_0$.

The following proposition states that $\mathbf{GLP}_0$ is sound with respect to $\mathcal{U}$. We first give an argument using the assumption of converse well-foundedness of $\mathcal{U}$. Then we shall refine this proof to obtain an argument formalizable in EA.

Proposition 3.9 *If $\mathbf{GLP}_0 \vdash \varphi$, then $\mathcal{U} \vDash \varphi$.*

Proof. By induction on $\mathbf{GLP}_0$-proofs. The validity of Löb's schema follows from the fact that the model is transitive and conversely well-founded. The only axiom that needs some special attention is the monotonicity axiom:

$$\langle n \rangle \varphi \rightarrow \langle m \rangle \varphi, \quad \text{for } m \leq n.$$

It suffices to show that $\mathcal{U} \vDash \langle n + 1 \rangle \varphi \rightarrow \langle n \rangle \varphi$. So, suppose that for some $\vec{\alpha} \in \mathcal{U}$ we have $\vec{\alpha} \Vdash \langle n + 1 \rangle \varphi$. Thus, for some $\vec{\alpha'}$ with $\vec{\alpha} R_{n+1} \vec{\alpha'}$, we have $\vec{\alpha'} \Vdash \varphi$. Let $p := dp(\varphi)$. Given $\vec{\alpha'}$ and n, by Lemma 3.6, we construct a $\vec{\beta}$ such that $\vec{\alpha} R_n \vec{\beta}$ and $\vec{\alpha'} \simeq_p \vec{\beta}$. By Lemma 3.2, $\vec{\beta} \Vdash \varphi$, hence $\vec{\alpha} \Vdash \langle n \rangle \varphi$. ⊠

13

4 Finite approximations

The given proof of Proposition 3.9 uses the fact that $\mathcal{U}$ is a conversely well-founded model. Here, we present a proof of soundness that is based on certain finite Kripke models approximating $\mathcal{U}$ and is formalizable in EA. These are defined as follows.

Definition 4.1 $\mathcal{F}^n_\alpha := \{\vec{\beta} \in \mathcal{U} \mid \beta_0 < \alpha \wedge \forall i \; nw(\beta_i) \leq n\}$. The R_i relations on $\mathcal{F}^n_\alpha$ are defined as before.

Notice that, for any $\alpha < \varepsilon_0$ and any $n \in \omega$, $\mathcal{F}^n_\alpha$ is finite. The following quite general observation explains the role of the models $\mathcal{F}^n_\alpha$.

Let $\mathcal{A}$ be any Kripke model (in the language of **GLP**), and let $\mathcal{B} \subseteq \mathcal{A}$ be a submodel of $\mathcal{A}$, that is, a structure obtained by restricting all the relations in $\mathcal{A}$ to a fixed subset of $\mathcal{A}$ and preserving the forcing of all the variables.

For $n > 0$, we write $\mathcal{B} \prec_n \mathcal{A}$ if, for each $x \in \mathcal{B}$ and $y \in \mathcal{A}$ such that xR_ky, there is a $y' \in \mathcal{B}$ such that $y' \simeq_{n-1} y$ in $\mathcal{A}$ and xR_ky'. In this case, we say that $\mathcal{B}$ is an *n-elementary submodel* of $\mathcal{A}$. We also stipulate $\mathcal{B} \prec_0 \mathcal{A}$.

Lemma 4.2 Suppose $\mathcal{B} \prec_n \mathcal{A}$. Then, $\mathcal{B}, x \Vdash \varphi$ iff $\mathcal{A}, x \Vdash \varphi$, for any $x \in \mathcal{B}$ and any φ with $dp(\varphi) \leq n$.

Proof. For $n = 0$ the statement is obvious. For $n > 0$, it follows by an easy induction on the length of φ.

Assume $\varphi = \langle k \rangle \psi$ and $\mathcal{A}, x \Vdash \varphi$ where $x \in \mathcal{B}$. Then there is a $y \in \mathcal{A}$ such that xR_ky and $\mathcal{A}, y \Vdash \psi$. By the given condition, there is a $y' \in \mathcal{B}$ such that xR_ky' and $\mathcal{A}, y' \Vdash \psi$, since $y' \simeq_{n-1} y$ and $dp(\psi) < n$. Hence, by the induction hypothesis, $\mathcal{B}, y' \Vdash \psi$ and $\mathcal{B}, x \Vdash \langle k \rangle \psi$. The other cases are quite obvious. $\boxtimes$

Lemma 4.3 $\mathcal{F}^p_\gamma \prec_p \mathcal{U}$, for any γ and p.

Proof. Consider any $\vec{\alpha} \in \mathcal{F}^p_\gamma$. Assume $\vec{\alpha} R_n \vec{\alpha}'$. Construct a $\vec{\beta}$ from $\vec{\alpha}'$ as in Lemma 3.6, taking $p - 1$ for p. Observe that $\vec{\alpha} R_n \vec{\beta}$ and $\vec{\beta} \simeq_{p-1} \vec{\alpha}'$. Since $nw(\beta_k) \leq p$, for $k \geq n$, and $\beta_k = \alpha_k$, for $k < n$, we see that $\vec{\beta} \in \mathcal{F}^p_\gamma$. $\boxtimes$

We shall write $\mathbf{GLP}_0 \vdash_q \varphi$ to indicate that φ is provable in $\mathbf{GLP}_0$ with a proof that uses only formulas of depth at most q.

Theorem 1 If $\mathbf{GLP}_0 \vdash_q \varphi$ then $\mathcal{F}^q_\gamma \vDash \varphi$, for any q, γ and φ. Moreover, this fact is formalizable in EA.

Proof. Clearly, if $dp(\varphi) \leq q$ and $\mathbf{GLP}_0 \vdash \varphi$, then $\mathcal{U} \vDash \varphi$ and hence $\mathcal{F}^q_\gamma \vDash \varphi$, by Lemmas 4.2 and 4.3. However, this argument passes through an infinite structure $\mathcal{U}$. To avoid this problem, we give a direct proof by induction on the length of a derivation in $\mathbf{GLP}_0$ (γ and q are parameters of the induction). We then indicate how to formalize this proof within EA.

First, we observe that Lemmas 3.5 and 3.6 also hold for the model $\mathcal{F}^q_\gamma$ and the associated notion of p-bisimilarity if $p < q$. One easily checks that the

given proofs of these lemmas work in $\mathcal{F}_\gamma^q$, since the constructed points $\vec{\beta}'$ and $\vec{\beta}$ actually belong to $\mathcal{F}_\gamma^q$.[3]

Parallel to the proof of Proposition 3.9, we should now consider any axiom of $\mathbf{GLP}_0$ of depth at most q and see that this axiom holds in $\mathcal{F}_\gamma^q$. Löb's axioms are now valid in $\mathcal{F}_\gamma^q$ simply because this model is finite. Thus, verifying a Löb's axiom in $\mathcal{F}_\gamma^q$ amounts to an application of ordinary induction; the complexity of the associated induction formula is shown below to be elementary.

Again, the only really interesting case is to show that

$$\mathcal{F}_\gamma^q \vDash \langle n+1 \rangle \varphi \to \langle n \rangle \varphi.$$

Indeed, consider some $\vec{\alpha} \in \mathcal{F}_\gamma^q$ with $\vec{\alpha} \Vdash \langle n+1 \rangle \varphi$. Consequently, for some $\vec{\alpha}' \in \mathcal{F}_\gamma^q$ with $\vec{\alpha} R_{n+1} \vec{\alpha}'$ we have $\vec{\alpha}' \Vdash \varphi$. Let $p := dp(\varphi)$. As $dp(\langle n+1 \rangle \varphi \to \langle n \rangle \varphi) \le q$, clearly $p < q$. Construct $\vec{\beta}$ from $\vec{\alpha}'$ and n as in Lemma 3.6. By this lemma, $\vec{\alpha} R_n \vec{\beta}$, $\vec{\alpha}' \simeq_p \vec{\beta}$, and $\vec{\beta} \in \mathcal{F}_\gamma^q$, by the condition on the nested width of $\vec{\beta}$. Hence, $\vec{\beta} \Vdash \varphi$ and $\vec{\alpha} \Vdash \langle n \rangle \varphi$.

The above proof is an argument by induction. It would be available in EA if we manage to represent a truth-definition for the model $\mathcal{F}_\gamma^q$ by a multi-exponentially bounded arithmetical formula. It is easy to define an elementary formula (in x, γ and q) expressing $x \in \mathcal{F}_\gamma^q$. In addition, we have to show that the relation $\mathcal{F}_\gamma^q, x \Vdash \varphi$ (in q and the codes of x, φ, and γ) is elementary and that its basic properties are verifiable in EA. Notice that this relation is naturally defined by bounded recursion on φ where the bound is essentially determined by the size of the representation of elements of the model. It is sufficient to only consider models of the form $\mathcal{F}_{\omega_m}^n$, for which we shall estimate the maximal size of the elements under a natural encoding. Here, $\omega_0 = 1$ and $\omega_{n+1} = \omega^{\omega_n}$.

We presuppose that the ordinals are represented by terms using the functions $+$ and ω^α and a constant 0. Let α be such a term. Its *length* is the number of symbols, whereas the *height* $h(\alpha)$ of α is defined by: $h(0) = 0$, $h(\omega^\alpha) = h(\alpha)+1$, and $h(\alpha + \beta) = \max(h(\alpha), h(\beta))$. Clearly, $h(\omega_n) = n + 1$.

Let $f(m,n)$ denote the maximal length of a natural term representation for ordinals of nested width at most n and of height at most m. Clearly, $f(0,n) = 1$ and $f(m+1,n) \le cnf(m,n)$, for some constant c. Thus, we obtain $f(m,n) \le (cn)^m$. The length of the non-zero part of a sequence $\vec{\alpha} \in \mathcal{F}_{\omega_m}^n$ is at most $m+1$. Hence, the length of a natural representation of $\vec{\alpha}$ is bounded by an elementary function of order $O(f(m+1,n)(m+1))$. The Gödel number of $\vec{\alpha}$ will then be bounded by a certain elementary function $b(m,n) = 2^{2^{O(m \log n)}}$.

Using this bound, one can define the forcing relation $\mathcal{F}_\gamma^q, x \Vdash \varphi$ by bounded recursion on the length of φ. For example,

$$\mathcal{F}_{\omega_m}^n, x \Vdash [n]\varphi \quad \text{iff} \quad \forall y \le b(m,n) \; (y \in \mathcal{F}_{\omega_m}^n \wedge x R_n y \to \mathcal{F}_{\omega_m}^n, y \Vdash \varphi).$$

Thus, the forcing relation can be naturally represented in EA. Then it is a routine matter to check that EA proves the usual inductive clauses for the forcing

---

[3]Externally, this is obvious, since p-bisimilarity in $\mathcal{F}_\gamma^q$ coincides with the one induced from $\mathcal{U}$, by Lemma 4.3.

15

relation on $\mathcal{F}^n_{\omega_m}$. To formalize the proof of Theorem 1 one also has to formalize the bounded recursive definition of p-bisimilarity (Definition 3.1) and to prove Lemmas 3.5 and 3.6. This is unproblematic. $\boxtimes$

Corollary 4.4 *Consistency of* $\mathbf{GLP}^*$ *is provable in* EA.

Proof. We reason in EA. Suppose that $\mathbf{GLP}^*$ is inconsistent. Then, for some m, $\mathbf{GLP}_0 \vdash [m]\perp$, by Lemma 3.8. Hence, for some n, we also have $\mathbf{GLP}_0 \vdash_n [m]\perp$. By Theorem 1, $\mathcal{F}^n_{\omega_m} \vDash [m]\perp$, which contradicts the obvious validity $\mathcal{F}^n_{\omega_m}, \widehat{\omega_m} \Vdash \langle m \rangle \top$. $\boxtimes$

Thus, we obtain from Lemma 2.4 the desired property of the orderings $<_x$.

Corollary 4.5 *Provably in* EA, $(S_x, <_x)$ *is an irreflexive ordering.*

Remark 4.6 Having defined the forcing relation uniformly for all the models $\mathcal{F}^n_{\omega_m}$, one can now also define a forcing relation for the whole model $\mathcal{U}$:

$$\mathcal{U}, \vec{\alpha} \Vdash \varphi \iff \exists n, m \, (n \geq \max(dp(\varphi), nw(\vec{\alpha})) \wedge \omega_m > \alpha_0 \wedge \mathcal{F}^n_{\omega_m}, \vec{\alpha} \Vdash \varphi).$$

The formula on the right hand side is bounded, because n and m can actually be bounded by the code of φ and the code of $\vec{\alpha}$. Modulo some work one can verify in EA the inductive clauses for the definition of forcing in $\mathcal{U}$, which allows to speak within EA about validity in $\mathcal{U}$.

5 Further corollaries

Theorem 1 allows us to obtain finitary (formalizable in EA) proofs of some other basic results on $\mathbf{GLP}$. In this section we presuppose familiarity with [1, 3].

First, we deal with the unique normal form theorem for words (Proposition 2.5). We formulate a sharper version of it which will also yield a simpler axiomatization of $\mathbf{GLP}_0$.

Let $\mathbf{Im}$ be $\mathbf{I}$ together with the restricted monotonicity schema

$$\langle x \rangle \alpha \to \langle y \rangle \alpha, \quad \text{for } x \geq y \text{ and } \alpha \text{ a word from } S_y.$$

Let $\mathbf{Iml}$ be $\mathbf{Im}$ plus the restricted Löb's schema

$$[x](\alpha \to \langle x \rangle \alpha) \to [x]\neg\alpha, \quad \text{where } \alpha \in S_x.$$

Proposition 5.1 *Every word is equivalent in* $\mathbf{Im}$ *to a word in a normal form.* $\mathbf{GLP}$*-equivalent normal forms are graphically the same. Moreover, both facts are formalizable in* EA.

Proof (sketch). We inspect the lemmas leading to a proof of Proposition 2.5, which corresponds to Proposition 3 and Corollary 5 in [3]. Locally within the present proof all the references will be to [3]. One has to check that the arguments work in $\mathbf{Im}$ and are formalizable in EA.

Lemma 1 works in **I**. Lemma 2 works in **Im**, hence every word can be brought to a normal form in **Im** (Proposition 3).

Proposition 4 also holds in **Im**, but a word of caution is necessary: the fact that the three cases $\alpha <_x \beta$, $\alpha = \beta$ and $\beta <_x \alpha$ are mutually exclusive presupposes the irreflexivity of the ordering. Hence, the uniqueness of a normal form (Corollary 5) relies on Theorem 1 of this paper.

To formalize these arguments in EA one only has to notice that the length of a word under a normalization procedure can only decrease. The arguments proceed by induction on length, and the lengths of the corresponding **Im**-derivations are polynomially bounded in the length of a given word. Hence, these arguments are formalizable in EA. ⊠

We remark that Corollaries 6, 7, 8 and Lemma 9 of [3] are also formalizable in EA.

Proposition 5.2 *The ordering* $(S, <_0)$ *is* EA-*provably isomorphic to the standard system of ordinal notation for* ε_0.

Proof (sketch). We now understand the function o as a mapping from S to the standard ordinal notations for ε_0 and formalize in EA a proof of Proposition 2.7. The most direct proof of this fact follows the one of Proposition 4 from [3].

By induction on the width $w(\alpha\beta)$ one simultaneously proves that, for all $\alpha, \beta \in S$,

$$
\begin{aligned}
o(\alpha) = o(\beta) &\Rightarrow \mathbf{GLP} \vdash \alpha \leftrightarrow \beta; \\
o(\alpha) < o(\beta) &\Rightarrow \mathbf{GLP} \vdash \beta \rightarrow \Diamond\alpha.
\end{aligned}
$$

Notice that, by the irreflexivity of $<_0$, these implications are reversible.

Without loss of generality assume that the smallest letter occurring in $\alpha\beta$ is 0. Write α and β in the form

$$
\alpha = \alpha_k^+ 0 \alpha_{k-1}^+ 0 \cdots 0 \alpha_1^+, \qquad \beta = \beta_m^+ 0 \beta_{m-1}^+ 0 \cdots 0 \beta_1^+.
$$

We may assume that $o(\alpha_i) \geq o(\alpha_{i+1})$, for all i. Otherwise, by the induction hypothesis, α can be brought to an equivalent form satisfying the required condition without changing its ordinal. (This is done as in Proposition 3 from [3] by deleting some of the α_i^+.) The same holds for β.

Assume $o(\alpha) = o(\beta)$. Then, since these ordinals are in normal form, one has $m = k$ and $\forall i \leq m \, o(\alpha_i) = o(\beta_i)$. It follows by the induction hypothesis that $\mathbf{GLP} \vdash \alpha_i \leftrightarrow \beta_i$, for all $i \leq m$. Then it is easy to infer that $\mathbf{GLP} \vdash \alpha \leftrightarrow \beta$.

Assume $o(\alpha) < o(\beta)$. Then, either $\forall i \leq k \, o(\alpha_i) = o(\beta_i)$ and $k < m$, or there is an $s \leq m, k$ such that $o(\alpha_s) < o(\beta_s)$ and $\forall i < s \, o(\alpha_i) = o(\beta_i)$. Following the proof of Proposition 4, one can easily show that $\alpha <_0 \beta$ in either case.

The given proof, being a simple kind of induction, is obviously formalizable in EA, but notice that we also relied on Theorem 1 to inverse the implications. ⊠

The following statement formalizes Proposition 2.6.

Proposition 5.3 *Provably in* EA*, every closed modal formula is* **Iml***-equivalent to a boolean combination of words.*

Proof (sketch). We examine the corresponding lemmas in [3]. A proof of Lemma 10 works in **Iml**. Prima facie, it uses the irreflexivity of $<_x$, however, a small adaptation actually shows that this is not even necessary.

Lemma 11 holds in **Iml** by virtue of Lemma 10. The additional identities used there hold in **I**. An application of Lemma 10 is the only use of Löb's axiom in the whole proof.

Let $w(\varphi)$ denote the width of φ, that is, the number of different indices of modalities in φ. Notice that there are only finitely many different formulas of a fixed width and depth. Also, we notice that, obviously, normalization does not increase the width.

By induction on the depth of φ we prove that there is a boolean combination of words equivalent to φ whose depth does not exceed $dp(\varphi)$. Basis of induction and the cases of boolean connectives are trivial. Suppose φ has the form $\langle x \rangle \psi$ with $dp(\psi) = dp(\varphi) - 1$. There is a boolean combination ψ' such that $dp(\psi') \leq dp(\varphi) - 1$ and ψ' is equivalent to ψ. The transformations in Lemma 10 all can only decrease the depth of $\langle x \rangle \psi'$. This proves the claim.

The number of different words of depth (=length) n and width m is bounded by m^n. Under a natural encoding, the length of any particular boolean combination can be bounded by a function of the form $2^{m^{O(n)}}$. Thus, given a formula φ of depth n we can find an equivalent boolean combination of words whose length is bounded by $2^{m^{O(n)}}$. We can also roughly estimate $m, n \leq |\varphi|$, and hence the length of a normal form of φ is bounded by a multi-exponential function of $|\varphi|$. This bound allows one to formalize the inductive proof of the theorem in EA. ⊠

On the basis of this result one can obtain (and formalize in EA) the following two corollaries, which we mention without proof.

Corollary 5.4 **Iml** *axiomatizes the closed fragment of* **GLP***.*

Corollary 5.5 $\mathbf{GLP}_0$ *is complete with respect to* $\mathcal{U}$*.*

We have shown that many basic results on the closed fragment of **GLP** can be obtained by finitary methods. The first author of this paper has proved the following theorem, which contrasts with these results.

Proposition 5.6 *For each* n*, any extension of* $\mathbf{GLP}^-$ *in a language with* n *modalities by a set of negated words is finitely axiomatizable. However, this statement is unprovable in* ACA_0*, a second-order conservative extension of* PA*.*

A proof of this fact is obtained by showing that the ordering of words over the alphabet $\{0, \ldots, n\}$ by implication in $\mathbf{GLP}^-$ is a well-quasiordering, isomorphic to the so-called Friedman's gap-embedding relation on sequences of numbers. This result is included in a recent paper by L. Carlucci [8].

References

[1] L.D. Beklemishev. Provability algebras and proof-theoretic ordinals, I. *Annals of Pure and Applied Logic*, 128:103–123, 2004.

[2] L.D. Beklemishev. Reflection principles and provability algebras in formal arithmetic. *Uspekhi Matematicheskikh Nauk*, 2005. To appear.

[3] L.D. Beklemishev. Veblen hierarchy in the context of provability algebras. Logic Group Preprint Series 232, University of Utrecht, June 2004. `http://preprints.phil.uu.nl/lgps/`.

[4] L.D. Beklemishev. Provability algebras and proof-theoretic ordinals, I. Logic Group Preprint Series 208, University of Utrecht, March 2001. `http://preprints.phil.uu.nl/lgps/`.

[5] L.D. Beklemishev. The Worm principle. Logic Group Preprint Series 219, University of Utrecht, March 2003. `http://preprints.phil.uu.nl/lgps/`.

[6] P. Blackburn, M. de Rijke, and Y. Venema. *Modal Logic*. Camebridge University Press, 2001.

[7] G. Boolos. *The Logic of Provability*. Cambridge University Press, Cambridge, 1993.

[8] L. Carlucci. Worms, gaps and hydras. To appear in *Mathematical Logic Quarterly*, 2005.

[9] P. Hájek and P. Pudlák. *Metamathematics of First Order Arithmetic*. Springer-Verlag, Berlin, Heidelberg, New York, 1993.

[10] K.N. Ignatiev. The closed fragment of Dzhaparidze's polymodal logic and the logic of Σ_1-conservativity. ITLI Prepublication Series X–92–02, University of Amsterdam, 1992.

[11] K.N. Ignatiev. On strong provability predicates and the associated modal logics. *The Journal of Symbolic Logic*, 58:249–290, 1993.

[12] G.K. Japaridze. The modal logical means of investigation of provability. Thesis in Philosophy, in Russian, Moscow, 1986.

[13] G.K. Japaridze. The polymodal logic of provability. In *Intensional Logics and Logical Structure of Theories: Material from the fourth Soviet–Finnish Symposium on Logic, Telavi, May 20–24, 1985*, pages 16–48. Metsniereba, Tbilisi, 1988. In Russian.

[14] J. J. Joosten. *Interpretability formalized*. PhD thesis, Department of Philosophy, University of Utrecht, 2004.

[15] G. Kreisel and A. Lévy. Reflection principles and their use for establishing the complexity of axiomatic systems. *Zeitschrift f. math. Logik und Grundlagen d. Math.*, 14:97–142, 1968.

[16] R.M. Solovay. Provability interpretations of modal logic. *Israel Journal of Mathematics*, 28:33–71, 1976.