Annals of Pure and Applied Logic 160 (2009) 129–153

Contents lists available at ScienceDirect

Annals of Pure and Applied Logic

journal homepage: www.elsevier.com/locate/apal

The predicative Frege hierarchy

Albert Visser

Department of Philosophy, Utrecht University, Heidelberglaan 8, 3584 CS Utrecht, The Netherlands

ARTICLE INFO

Article history:
Received 13 October 2006
Received in revised form 28 January 2009
Accepted 1 February 2009
Available online 3 March 2009
Communicated by S.N. Artemov

MSC:
03A05
03B30
03F25
03F35

Keywords:
Predicative comprehension
Frege
Interpretability

ABSTRACT

In this paper, we characterize the strength of the predicative Frege hierarchy, $\mathsf{P}^{n+1}\mathsf{V}$, introduced by John Burgess in his book [J. Burgess, Fixing frege, in: Princeton Monographs in Philosophy, Princeton University Press, Princeton, 2005]. We show that $\mathsf{P}^{n+1}\mathsf{V}$ and $\mathsf{Q}+\mathsf{con}^n(\mathsf{Q})$ are mutually interpretable. It follows that $\mathsf{PV} := \mathsf{P}^1\mathsf{V}$ is mutually interpretable with Q. This fact was proved earlier by Mihai Ganea in [M. Ganea, Burgess' PV is Robinson's Q, The Journal of Symbolic Logic 72 (2) (2007) 619–624] using a different proof. Another consequence of the our main result is that $\mathsf{P}^2\mathsf{V}$ is mutually interpretable with Kalmar Arithmetic (a.k.a. EA, EFA, $I\Delta_0 + \mathsf{EXP}$, Q3). The fact that $\mathsf{P}^2\mathsf{V}$ interprets EA was proved earlier by Burgess. We provide a different proof.

Each of the theories $\mathsf{P}^{n+1}\mathsf{V}$ is finitely axiomatizable. Our main result implies that the whole hierarchy taken together, $\mathsf{P}^{\omega}\mathsf{V}$, is not finitely axiomatizable. What is more: no theory that is mutually locally interpretable with $\mathsf{P}^{\omega}\mathsf{V}$ is finitely axiomatizable.

1. Introduction

A predicative version of the logicist program is outlined in [2], chapter 2. The idea is to build a hierarchy of stronger and stronger systems obtained by adding at each next stage (i) predicative second order comprehension over the previous system and (ii) the full principle V for the newly added concepts.[1] More precisely, the hierarchy is defined as follows.

- $\mathsf{PV} := \mathsf{P}^1\mathsf{V}$ is the system we obtain by adding second order variables $X^0, Y^0, \ldots$ and a function symbol $\ddagger^0$ to the predicate logic of pure identity, plus the following axioms.

$\mathsf{P}^1 1)\ \vdash \exists X^0\, \forall x\, (X^0 x \leftrightarrow A(x, \vec{y}, \vec{Y}^0))$,
where A does not contain X and does not contain bound concept variables of degree 0.
$\mathsf{P}^1 2)\ \vdash \ddagger^0 X^0 = \ddagger^0 Y^0 \leftrightarrow \forall z\, (X^0 z \leftrightarrow Y^0 z)$.

- $\mathsf{P}^{n+2}\mathsf{V}$ is the theory obtained by adding to $\mathsf{P}^{n+1}\mathsf{V}$ new second order variables X^{n+1} and a new function symbol $\ddagger^{n+1}$, plus the following axioms.

$\mathsf{P}^{n+2} 1)\ \vdash \exists X^{n+1}\, \forall x\, (X^{n+1} x \leftrightarrow A(x, \vec{y}, \vec{Y}^0, \ldots, \vec{Y}^{n+1}))$,
where A does not contain X and does not contain bound concept variables of degree $n+1$.
$\mathsf{P}^{n+2} 2)\ \vdash \ddagger^{n+1} X^{n+1} = \ddagger^n Y^n \leftrightarrow \forall z\, (X^{n+1} z \leftrightarrow Y^n z)$.
$\mathsf{P}^{n+2} 3)\ \vdash \ddagger^{n+1} X^{n+1} = \ddagger^{n+1} Y^{n+1} \leftrightarrow \forall z\, (X^{n+1} z \leftrightarrow Y^{n+1} z)$.

We think this approach to predicativity is in many respects attractive. There is the undeniable simplicity and naturality of the chosen axioms and the charm of combining Fregean and Russellian ideas. More importantly, the hierarchy goes way beyond

E-mail address: Albert.Visser@phil.uu.nl.

1 In the Frege style, the denotations of the second order variables are called *concepts*.

0168-0072/$ – see front matter © 2009 Elsevier B.V. All rights reserved.
doi:10.1016/j.apal.2009.02.001

the predicative systems provided by Nelson's approach. See [13]. Nelson developed predicative systems by considering simply what is interpretable in Q. There are two significant objections to Nelson's project. One is that it is unclear how he justifies the use of unbounded quantification. This criticism was voiced in Pudlák's review [15]. A second criticism is that his approach lacks reflexive closure. Specifically, we cannot prove con(Q) in the Nelson systems. In fact, addition of con(Q) would yield a system that *violates* Nelson's philosophy, since $\mathsf{Q} + \mathsf{con}(\mathsf{Q})$ is mutually interpretable with $I\Delta_0 + \mathsf{EXP}$, and the totality of exponentiation is something Nelson denies. See for this criticism: [9]. The present Frege-style approach partly evades this second criticism. As we will see the hierarchy (up to ω), provides consistency statements for each of its stages. On the other hand, we will show that the hierarchy, in a sense, stops at ω. As a consequence, for no ordinal α, the theory $\mathsf{P}^\alpha\mathsf{V}$ will prove the consistency of $\mathsf{P}^\omega\mathsf{V}$. Thus, the hierarchy only evades the criticism, if we are prepared to view it as 'open ended' towards ω.

A further good point about the hierarchy is that it is reasonably stable w.r.t. design choices, like the choice whether or not to begin with just the theory of identity or rather with, say, the theory of pairing.

Our aim in the present paper is technical rather than philosophical. We provide an answer to the question: *how strong are the* $\mathsf{P}^n\mathsf{V}$? We show that, verifiably in $I\Delta_0 + \Omega_1$, the theory $\mathsf{P}^{n+1}\mathsf{V}$ is mutually interpretable with the theory $\mathsf{Q} + \mathsf{con}^n(\mathsf{Q})$. This result generalizes a result of Mihai Ganea, who shows that $\mathsf{PV} := \mathsf{P}^1\mathsf{V}$ is mutually interpretable with Q. See [6]. Ganea's interpretation of PV in Q is simpler than the one we provide. On the other hand, to verify the correctness of the interpretation, he employs a corollary of the Löwenheim–Behmann Theorem, due to Burgess. It is not known whether this corollary can be verified in $I\Delta_0 + \Omega_1$.

One consequence of the available consistency statements is that we have exponentiation available in our hierarchy —in fact already in $\mathsf{P}^2\mathsf{V}$.

We will show that the $\mathsf{P}^n\mathsf{V}$ are all finitely axiomatizable. In contrast their limit, $\mathsf{P}^\omega\mathsf{V}$ is not finitely axiomatizable. What is more: no theory that is locally mutually interpretable with $\mathsf{P}^\omega\mathsf{V}$ is finitely interpretable.

The proof of the finite axiomatizability result uses Burgess' result that PV is finitely axiomatizable, which in turn uses the Löwenheim–Behmann Theorem. Thus, it is unknown whether it can be verified in $I\Delta_0 + \Omega_1$.

The methodology of the paper is what one could call *miniature model theory*. This endeavor falls between proof theory and model theory. As in proof theory, we study syntactical matters, but unlike in proof theory we seldom look at the details of proofs. As in model theory, we employ the intuition of constructing structures. We lack, however, the possibility to quantify over structures. Our 'structures' will in fact be interpretations given by concrete formulas. In model theory we work in a strong metatheory like ZFC. Here, we work in a weak theory like $I\Delta_0 + \Omega_1$. Thus, we lack even the full strength of induction. We compensate for the lack of induction by employing Solovay's methodology of shortening cuts. In effect, we follow the idea *if you can't do what you want to do with the number system you are working with, switch to another one*.

To prove our main result we should realize two directions. We should move from a consistency statement to predicative comprehension and axiom V. To do this we miniaturize the following model theoretic argument. Given that we know that U is consistent, we can use the Henkin construction to build a countable model of U. We can extend this model to a model of predicative comprehension by adding the parametrically first order definable sets over the model. The class of these sets is countable, so there is a mapping of these sets into the object domain of our model. We choose such a mapping to serve as our Frege function. To miniaturize the argument, we build an interpretation rather than a model. This is done using the Henkin–Feferman construction. It turns out that adding the definable sets is as easy as it is in ordinary model theory. This part of the proof is in Section 5. To find the Frege function, we have to do some work. We need to specify the function concretely. To make this possible, U should satisfy some constraints. The one we use is the demand that U interprets a certain theory of two successors. Also we have to switch number systems to obtain some desired effects of induction. All this is realized in Section 6.

In the other direction, we derive consistency from predicative comprehension. Here, we employ a well-known strategy. We are given a predicative extension of U. We use our classes to build a truth predicate for the U-language. Then, we use the truth predicate to prove consistency of U, compensating for the lack of induction by going to a definable cut. This is executed in Section 7.

A remarkable fact, emerging from the argument, is that the presence of the Frege functions only adds metamathematical strength, when we move from the theory of pure identity to $\mathsf{P}^1\mathsf{V}$. In all subsequent steps, the gain in power is achieved by predicative comprehension all by itself!

Finally, in Section 8, we put everything together.

Prerequisites

A good introduction to many of the methods and ideas of the paper is [7].

2. Theories and interpretations

In this section, we introduce basic notions and tools.

2.1. Theories

We consider theories in many-sorted first order predicate logic. The axiomatization of the theories should be sufficiently simple, e.g. Δ_1^{b}. The default is that our theories have finitely many sorts and are of finite signature.[2] It is optional whether a sort has identity or not.

We will sometimes consider *pointed theories*, i.e. theories with a designated sort. We will always assume that the pointed sort has identity. We will write '$U\lceil\mathfrak{a}\rceil$' for: theory U with designated sort $\mathfrak{a}$. We will confuse one-sorted theories with pointed one-sorted theories. Moreover, specific named theories will often have a fixed implicit point, E.g., the $\mathsf{P}^n\mathsf{V}$ will have, as designated sort, the sort of basic objects.

Our notion of theory is *intensional*. We assume some proof system is fixed, so a theory will be given by its signature (including the sorts) plus an arithmetical formula defining the set of (Gödel numbers of) axioms. We will use $\subseteq$ and $=_{\mathsf{ext}}$ for the subset relation and the identity relation *between the theories considered as sets of theorems*. *A finitely axiomatized theory* is always specified with an explicit numerical bound on the size of the Gödel numbers of the axioms. Note that this notion is stronger than a specification of the set of axioms just by a formula that gives us de facto a finite set. On the other hand it is weaker – in the context of $I\Delta_0 + \Omega_1$ as metatheory – than having a numerical code for the finite set of axioms.

We define two important operations on theories that will play an important role in this paper.

- $\Theta_U := \mathsf{Q} + \mathsf{con}(U)$,
- $\Omega_U := I\Delta_0 + \Omega_1 + \mathsf{con}(U)$.

By a result of Wilkie, the theories Θ_U and Ω_U are mutually interpretable. Note that the operations Θ and Ω are essentially intensional. For every consistent U, we can find a V, such that $U =_{\mathsf{ext}} V$ and $\Omega_U \neq_{\mathsf{ext}} \Omega_V$.

2.2. Interpretations

Interpretations play a main role in the present paper. They are both part of our methods of proof – as the tools of miniature model theory – and of the statement of our results. Our main result is stated in terms of *mutual interpretability* which is a very good way of measuring *metamathematical strength* of theories. In contrast, in proof theory, theories are often compared using conservativity w.r.t. some class of sentences like Π_2^0.

Interpretability in this paper will be *one-dimensional, many-sorted, relative interpretability without parameters where identity is not necessarily translated as identity*. We provide a rather extensive treatment, since we are not aware of a good treatment of interpretability between many-sorted theories in the literature. Especially, there is a tendency towards fuzzy thinking about the relationship between many-sorted theories and their one-sorted flattening. Since, there is an important relationship between flattening and Frege functions, it seems good to provide an introduction.

The choice for one-dimensional interpretations without parameters is mainly one of convenience. Developing the full machinery with the parameters and more dimensionality would be more laborious. Moreover, our main result, which states that certain theories are mutually interpretable becomes stronger, when stated for a more restrictive notion of interpretability. Of course, non-interpretability results become weaker for the more restrictive notion. We will briefly meet this phenomenon in Remark 2.1.

We will define interpretations for *relational* languages. To obtain interpretations for languages with functions, we consider them as consisting of two steps. First one translates the given language to a relational one using a standard algorithm. It is well known this can be done in polynomial time. Then, we apply an interpretation as defined below.[3]

To define an interpretation, we first need the notion of translation. Let Σ and Ξ be finite signatures for many-sorted predicate logic with finitely many sorts. We assume that the sorts are specified with the signature. A *relative translation* $\tau : \Sigma \to \Xi$ is given by a triple $\langle \sigma, \delta, F\rangle$. Here σ is a mapping of the Σ-sorts to the Ξ-sorts. The mapping δ assigns to every Σ-sort $\mathfrak{a}$ a Ξ-formula $\delta^{\mathfrak{a}}$ representing the *domain* for sort $\mathfrak{a}$ of the translation. We demand that $\delta^{\mathfrak{a}}$ contains at most a designated variable $v_0^{\sigma\mathfrak{a}}$ free. The mapping F associates to each relation symbol R ofΣ a Ξ-formula $F(R)$. The relation symbol R comes equipped with a sequence $\vec{\mathfrak{a}}$ of sorts. We demand that $F(R)$ has at most the variables $v_i^{\sigma\mathfrak{a}_i}$ free. We translate Σ-formulas to Ξ-formulas as follows:

- $(R(y_0^{\mathfrak{a}_0}, \ldots, y_{n-1}^{\mathfrak{a}_{n-1}}))^\tau := F(R)(y_0^{\sigma\mathfrak{a}_0}, \ldots, y_{n-1}^{\sigma\mathfrak{a}_{n-1}})$;
  (We assume that some mechanism for α-conversion is built into our definition of substitution to avoid variable-clashes.)
- $(\cdot)^\tau$ commutes with the propositional connectives;
- $(\forall y^{\mathfrak{a}} A)^\tau := \forall y^{\sigma\mathfrak{a}} (\delta^{\mathfrak{a}}(y) \to A^\tau)$;
- $(\exists y^{\mathfrak{a}} A)^\tau := \exists y^{\sigma\mathfrak{a}} (\delta^{\mathfrak{a}}(y) \wedge A^\tau)$.

Suppose τ is $\langle \sigma, \delta, F\rangle$. Here are some convenient conventions and notation.

---

2 There will be just one exception considered in the paper: the theory $\mathsf{P}^\omega\mathsf{V}$.

3 Note that if we have U and the corresponding U^{rel}, we have $U \vdash A \Leftrightarrow U^{\mathsf{rel}} \vdash A^{\mathsf{rel}}$. Moreover, there is an inverse $(\cdot)^{\mathsf{fun}}$ of $(\cdot)^{\mathsf{rel}}$, to wit: substitute $f\vec{x} = y$ for, say, $F_f(\vec{x}, y)$. We have: $U \vdash A\vec{z} \leftrightarrow ((A\vec{z})^{\mathsf{rel}})^{\mathsf{fun}}$ and $U^{\mathsf{rel}} \vdash B\vec{z} \leftrightarrow ((B\vec{z})^{\mathsf{fun}})^{\mathsf{rel}}$. So, in a reasonably strong sense, U and U^{rel} are 'the same'.

- We write σ_τ for σ, δ_τ for δ and F_τ for F.
- We write R_τ for $F_\tau(R)$.
- We will always use '$=^{\mathfrak{a}}$' for the (optional) identity of a theory for sort $\mathfrak{a}$. In the context of translating, we will however switch to '$E^{\mathfrak{a}}$'.
- We write $\vec{x} : \delta^{\vec{\mathfrak{a}}}$ for: $\delta^{\mathfrak{a}_0}(x_0^{\sigma\mathfrak{a}_0}) \wedge \cdots \wedge \delta^{\mathfrak{a}_{n-1}}(x^{\sigma\mathfrak{a}_{n-1}})$.
- We write $\forall\vec{x} : \delta^{\vec{\mathfrak{a}}} A$ for: $\forall x_0^{\sigma\mathfrak{a}_0} \ldots \forall x_{n-1}^{\sigma\mathfrak{a}_{n-1}}\ (\vec{x}{:}\delta^{\vec{\mathfrak{a}}} \to A)$.
  Similarly for the existential case.

A special translation on a signature Σ is the identity translation id_Σ. The first component σ of this translation sends all sorts of Σ to themselves. The second component δ sends each sort $\mathfrak{a}$ to $\top$. The third component F sends each predicate symbol P to $P\vec{v}^{\vec{\mathfrak{a}}}$. We can compose relative translations as follows:

- $\delta^{\mathfrak{a}}_{\tau\nu}(v^{\sigma_\nu\sigma_\tau\mathfrak{a}}) := (\delta_\nu^{\sigma_\tau\mathfrak{a}}(v^{\sigma_\nu\sigma_\tau\mathfrak{a}}) \wedge (\delta^{\mathfrak{a}}_\tau(v^{\sigma_\tau\mathfrak{a}}))^\nu)$,
- $R_{\tau\nu}\vec{v}^{\,\sigma_\nu\sigma_\tau\vec{\mathfrak{a}}} = (R_\tau\vec{v}^{\,\sigma_\tau\vec{\mathfrak{a}}})^\nu$.

We write $\nu \circ \tau := \tau\nu$.

A translation τ supports a *relative interpretation* of a theory U in a theory V, if, for all U-sentences A, we have $U \vdash A \Rightarrow V \vdash A^\tau$.[4] Thus, an interpretation has the form: $K = \langle U, \tau, V\rangle$.

Note that the definition automatically takes care of the theory of identity. Moreover, it follows that $V \vdash \exists v_0\ \delta^{\mathfrak{a}}_\tau$.

We write $K : U \to V$, $K : U \lhd V$ or $K : V \rhd U$, for: K is an interpretation of the form $\langle U, \tau, V\rangle$. The notation $K : U \to V$ is used when we are thinking of theories and interpretations as objects and morphisms in a category. The notation $K : U \lhd V$ is used when we are thinking of $\lhd$ as a preorder. Moreover, the notation $\rhd$ is intended to suggest that interpretability is a generalization of provability.

Par abus de langage, we write 'δ_K' for: δ_{τ_K}; 'P_K' for: P_{τ_K}; 'A^K' for: A^{τ_K}, etc.
Suppose T has signature Σ and $K : U \to V$, $M : V \to W$. We define:

- $\mathsf{id}_T : T \to T$ is $\langle T, \mathsf{id}_\Sigma, T\rangle$,
- $M \circ K : U \to W$ is $\langle U, \tau_M \circ \tau_K, W\rangle$.

We identify two interpretations $K, K' : U \to V$ if, $\sigma_K = \sigma_{K'}$ and:

- for all U-sorts $\mathfrak{a}$, $V \vdash \forall v^{\sigma_K\mathfrak{a}}\ (\delta^{\mathfrak{a}}_K v \leftrightarrow \delta^{\mathfrak{a}}_{K'} v)$,
- $V \vdash \forall\vec{v} : \delta^{\vec{\mathfrak{a}}}\ (R^K\vec{v} \leftrightarrow R^{K'}\vec{v})$, where $\vec{a}$ is the sequence associated with R.

One can show that modulo this identification, the above operations give rise to a category of interpretations that we call $\mathsf{INT}^{\mathsf{ms}}$. If we just consider one-sorted theories, we call the resulting category INT. Isomorphism in $\mathsf{INT}^{\mathsf{ms}}$ is called *synonymy* or *definitional equivalence*.

2.3. Isomorphisms between interpretations

Consider $K, M : U \to V$. An isomorphism $G : K \Rightarrow M$ is a V-definable, V-provable isomorphism from K to M considered as 'parametrized internal models'. Specifically, this means that an isomorphism from K to M is given as a triple $\langle K, G, M\rangle$, where G assigns to each U-sort $\mathfrak{a}$ a formula $G^{\mathfrak{a}}$ with the following properties.

- The free variables of $G^{\mathfrak{a}}$ are among $v_0^{\sigma_K\mathfrak{a}}, v_1^{\sigma_M\mathfrak{a}}$.
  We write $G^{\mathfrak{a}}(x, y)$ or $xG^{\mathfrak{a}}y$, for: $G^{\mathfrak{a}}[v_0 := x, v_1 := y]$.
- $V \vdash xG^{\mathfrak{a}}y \to (x : \delta^{\mathfrak{a}}_K \wedge y : \delta^{\mathfrak{a}}_M)$.
- $V \vdash \forall x : \delta^{\mathfrak{a}}_K\ \exists y : \delta^{\mathfrak{a}}_M\ xG^{\mathfrak{a}}y$.
- $V \vdash \forall y : \delta^{\mathfrak{a}}_M\ \exists x : \delta^{\mathfrak{a}}_K\ xG^{\mathfrak{a}}y$.
- $V \vdash \vec{x}G^{\vec{\mathfrak{a}}}\vec{y} \to (R_K\vec{x} \leftrightarrow R_M\vec{y})$.
  Here '$\vec{x}G^{\vec{\mathfrak{a}}}\vec{y}$' abbreviates $x_0G^{\mathfrak{a}_0}y_0 \wedge \cdots \wedge x_{n-1}G^{\mathfrak{a}_{n-1}}y_{n-1}$, for $\vec{\mathfrak{a}}$ corresponding to R.[5]

4 If we have Σ-collection available in our metatheory, this definition coincides with the one where we just demand that for all U-axioms A, $V \vdash A^\tau$. Since, we will be interested in verifiability in $I\Delta_0 + \Omega_1$, which lacks Σ^0_1-collection, we need the notion involving *theorems*. Otherwise, e.g. the transitivity of interpretability cannot be verified. Note that, if $I\Delta_0 + \Omega_1$ proves that, for all axioms A of U, $V \vdash A^\tau$, then $I\Delta_0 + \Omega_1$ proves also that, for all U-sentences B, if $U \vdash B$, then $V \vdash B^\tau$. This is because $I\Delta_0 + \Omega_1$ will supply p-time bounds on the V-proofs of the A^τ, by a theorem of Wilkie and Paris in their [22]. See further [18].

5 Note that this covers the case of the functionality and the injectivity of $G^{\mathfrak{a}}$ in case the sort $\mathfrak{a}$ has identity. In case the sort does not have identity, we consider the question of functionality and injectivity to be vacuous.

By induction on A, we can show that, for the appropriate $\vec{\mathfrak{a}}$:

$$V \vdash \vec{x} G^{\vec{\mathfrak{a}}} \vec{y} \to (A^K \vec{x} \leftrightarrow A^M \vec{y}). \tag{1}$$

We may divide out isomorphisms of interpretations in the category $\mathsf{INT}^{\mathsf{ms}}$. One can show that in this way we obtain a new category $\mathsf{hINT}^{\mathsf{ms}}$. (See [20] for a treatment of the one-sorted case.) Isomorphism of theories in this category is called: *bi-interpretability*.

2.4. Flattening

Consider any many-sorted theory U of signature Σ. Let sort be the set of sorts of Σ. We associate to U a one-sorted theory $U^\flat$ as follows. We take as language of $U^\flat$ a one-sorted language with the predicate symbols of U plus, for each $\mathfrak{a}$ in sort, a new unary predicate symbol $\Delta^{\mathfrak{a}}$. If $\vec{\mathfrak{a}}$ is the sequence associated to R in Σ, we associate to R a sequence of the same length consisting of the single sort in the new signature. Viewed differently, we give as arity to R in the flat environment the length of $\vec{\mathfrak{a}}$. We define a translation $\eta := \langle \sigma, \delta, F \rangle$ from the language of U to the language of $U^\flat$.

- σ sends all sorts of U to the single sort of $U^\flat$.
- $\delta^{\mathfrak{a}}(v) :\leftrightarrow \Delta^{\mathfrak{a}}(v)$.
- $F(R)(\vec{v}) :\leftrightarrow R(\vec{v})$.

Here are the axioms of $U^\flat$.

$\flat$1) $\vdash \forall v \; \bigvee_{\mathfrak{a} \in \mathsf{sort}} \Delta^{\mathfrak{a}}(v)$.
$\flat$2) $\vdash R(\vec{v}, w, \vec{z}) \to \Delta^{\mathfrak{a}}(w)$, where $\mathfrak{a}$ is the sort corresponding to the location of w in $R(\vec{v}, w, \vec{z})$ according to Σ.
$\flat$3) $\vdash A^\eta$, where A is an axiom of U.

Clearly, there is an interpretation based on η of U in $U^\flat$. Par abus de langage, we call this interpretation also η. The mapping $(\cdot)^\flat$ has all kinds of good properties, as is pointed out in the remark below, but these will pay no further role in this paper.

Remark 2.1. Suppose $K : U \to V$ and V is one-sorted. Then we can easily show that there is a unique $K^\star : U^\flat \to V$, such that $K = K^\star \circ \eta_U$. Thus, (by [10], p81, Theorem 2(ii)) it follows that $(\cdot)^\flat$ is a functor from $\mathsf{INT}^{\mathsf{ms}} \to \mathsf{INT}$, that is left adjoint to the embedding functor of INT into $\mathsf{INT}^{\mathsf{ms}}$.

We do *not* have, generally, that $U^\flat$ is definitionally equivalent or even bi-interpretable with U. In fact, $U^\flat$ need not be mutually interpretable with U. E.g., consider a two-sorted theory W with identity for both sorts and no further predicate symbols. The theory's axioms say that the first sort contains precisely two elements and the second sort precisely three. It is easily seen that W does not interpret $W^\flat$. Note that definitional equivalence implies the existence of a bijection between the sets of sorts. So, a more-than-one-sorted theory can never be definitionally equivalent to a one-sorted one.

Our discussion depends on the precise choice of our notion of interpretation. If we allow multi-dimensional interpretations, we can make an interpretation of $U^\flat$ in U, assuming that U has identity and proves the existence of at least two objects in one of its sorts.

Along another line, we can also establish a close connection between U and $U^\flat$. By a model theoretic argument, we can easily show that: $U \vdash A \Leftrightarrow U^\flat \vdash A^\eta$.

2.5. Interpretability

We define partial preorders on many-sorted theories.

- $K : U \rhd V :\Leftrightarrow K : V \lhd U$.
- $U \rhd V :\Leftrightarrow V \lhd U :\Leftrightarrow \exists K \; K : V \lhd U$.
  We read $V \lhd U$ as: V is interpretable in U. We read $U \rhd V$ as: U interprets V.
- We also want interpretability between pointed theories.
  $U\lceil \mathfrak{a} \rceil \rhd V\lceil \mathfrak{b} \rceil :\Leftrightarrow V\lceil \mathfrak{b} \rceil \lhd U\lceil \mathfrak{a} \rceil :\Leftrightarrow \exists K \; (K : V \lhd U \wedge \sigma_K(\mathfrak{b}) = \mathfrak{a})$.
- A finite subtheory $V \restriction n$ of V is a theory with axioms defined by $\alpha_V(x) \wedge x < \underline{n}$. We define:
  $U \rhd_{\mathsf{loc}} V :\Leftrightarrow V \lhd_{\mathsf{loc}} U :\Leftrightarrow \forall n \; (V \restriction n) \lhd U$.
  We read $V \lhd_{\mathsf{loc}} U$ as: V is locally interpretable in U.
- We define $U \equiv V :\Leftrightarrow (U \lhd V \wedge V \lhd U)$.
  We say: U and V are mutually interpretable. Similarly for $U \equiv_{\mathsf{loc}} V$.

3. Addition of principles as a functor

In this section, we study basic constructions used to build predicative systems.

3.1. The functor PC

We study the operation of adding predicative comprehension to a theory. We show that this operation gives us a functor PC. The functor PC works on pointed theories. Let a pointed theory $U\lceil\mathfrak{a}\rceil$ be given. We extend the language of U with an extra sort $\mathfrak{c}$ for concepts. We write the variables of the new sort as capitals. We add a new predicate app with associated sequence $\mathfrak{c}\mathfrak{a}$. We write 'Xx' for $\mathsf{app}(X, x)$. We add the axiom scheme of predicative comprehension:

$$\vdash \exists X\, \forall x^{\mathfrak{a}}\ (Xx \leftrightarrow A(x, \vec{y}, \vec{Y})).$$

Here A does not contain any concept quantifiers. Moreover, A does not contain X. The sequence $\vec{y}$ may contain variables of any sort of U.

Remark 3.1. Note that in the official language this could have been written as:

$$\vdash \exists x^{\mathfrak{c}}\, \forall y^{\mathfrak{a}}\ (\mathsf{app}(x, y) \leftrightarrow A(y, \vec{z})).$$

Here A does not contain quantifiers of sort $\mathfrak{c}$. Moreover, x does not occur in A. The sequence $\vec{z}$ consists of variables of all sorts.

The resulting pointed theory is $\mathsf{PC}(U\lceil\mathfrak{a}\rceil)$, also written as $(U\lceil\mathfrak{a}\rceil)^{\mathsf{pc}}$. We take as new point simply the old point $\mathfrak{a}$. We will write $\wr x \mid A(x, \vec{y}, \vec{Y})\wr$ for a concept provided by comprehension. Note that we should be careful in using comprehension terms, since the theory does not guarantee uniqueness. The comprehension terms are only unique modulo definable extensional equality.

The following theorem tells us that the mapping PC is a functor w.r.t. the interpretability preorder.

Theorem 3.2. *We have:* $U\lceil\mathfrak{a}\rceil \rhd V\lceil\mathfrak{b}\rceil \Rightarrow (U\lceil\mathfrak{a}\rceil)^{\mathsf{pc}} \rhd (V\lceil\mathfrak{b}\rceil)^{\mathsf{pc}}$. *This fact is verifiable in* $I\Delta_0 + \Omega_1$.

Proof. Suppose $K : U\lceil\mathfrak{a}\rceil \rhd V\lceil\mathfrak{b}\rceil$. We define K^{pc} as follows.

- $\sigma_{K^{\mathsf{pc}}}$ restricted to the sorts of V is σ_K; $\sigma_{K^{\mathsf{pc}}}(\mathfrak{c}) = \mathfrak{c}$;
- $\delta_{K^{\mathsf{pc}}}$ restricted to the sorts of V is δ_K;
  $\delta^{\mathfrak{c}}_{K^{\mathsf{pc}}}(X) :\leftrightarrow \forall x^{\mathfrak{a}}\ (Xx \rightarrow \delta^{\mathfrak{b}}_K(x)) \wedge \forall x^{\mathfrak{a}}, y^{\mathfrak{a}}\ ((Xx \wedge xE^{\mathfrak{b}}_K y) \rightarrow Xy)$;
  (Remember our assumption that designated sorts have identity.)
- $F_{K^{\mathsf{pc}}}$ restricted to the predicates of U is F_K; $(Xx)_{K^{\mathsf{pc}}} :\leftrightarrow Xx$.

To verify that we do interpret comprehension, we have to show:

$$U \vdash (\vec{y} : \delta^{\vec{\mathfrak{s}}}_{K^{\mathsf{pc}}} \wedge \vec{Y} : \delta^{\mathfrak{c}}_{K^{\mathsf{pc}}}) \rightarrow (\exists X\, \forall x^{\mathfrak{b}}\ (Xx \leftrightarrow A(x, \vec{y}, \vec{Y})))^{K^{\mathsf{pc}}}.$$

Here A does not contain bound concept variables, A does not contain X, and $\vec{\mathfrak{s}}$ is a sequence of sorts of U.

So, in U, under the presupposition of the antecedent, we have to provide an X in $\delta^{\mathfrak{c}}_{K^{\mathsf{pc}}}$ such that, for all x in $\delta^{\mathfrak{a}}_{K^{\mathsf{pc}}}$, Xx iff $A^{K^{\mathsf{pc}}}(x, \vec{y}, \vec{Y})$. Noting that our translation does not introduce new bound concept variables, we see that $X := \wr x \mid \delta^{\mathfrak{a}}_{K^{\mathsf{pc}}}(x) \wedge A^{K^{\mathsf{pc}}}(x, \vec{y}, \vec{Y})\wr$, provides a desired solution.

We easily see that the U-proofs that verify comprehension inside K^{pc} are p-time in the code of A. The main thing to verify is the closure of X as defined under $E^{\mathfrak{b}}_K$. This is an induction on A, where one provides a p-time estimate in A of the U-proof. □

3.2. Finite axiomatizability I

For one-sorted theories with pairing, predicative comprehension can given by finitely many axioms. Here is a statement of the theorem.

Theorem 3.3. *Suppose U is a one-sorted theory with identity and pairing. Then,* U^{pc} *can be finitely axiomatized* over U.

We give a proof of Theorem 3.3 in Appendix A. If we consider one-sorted sequential theories everything gets much simpler. Moreover, the question where the theorem can be verified is more perspicuous. Since, we only need the theorem for the sequential case, we will treat the sequential case here.

A one-sorted theory is *sequential* if it has a good notion of sequence of objects that works for *all* objects of the domain. This means that the theory interprets a weak arithmetic, say Q, via an interpretation, say N.[6]

Further the theory defines a domain of sequences with projection functions w.r.t. the N-numbers. It verifies principles stating that we have an empty sequence and that can always move from σ to $\sigma * \langle x\rangle$. For details, see [7]. We can always improve our theory of sequences, by shortening N. First, we can strengthen the theory of numbers that is interpreted to, say, $I\Delta_0 + \Omega_1$. Secondly we can add all kinds of desirable operations on sequences like concatenation.

Theorem 3.4. *Suppose U is a one-sorted, sequential theory. Then,* U^{pc} *can be finitely axiomatized* over U.

6 Pudlák asks that the interpretation N preserves identity. I prefer to define sequentiality without this demand. In the present context the distinction is irrelevant, since we can work with an identity preserving N. See also Remark 4.6.

Proof. We may assume that the numbers of U satisfy $I\Delta_0 + \Omega_1$. Say $N : U \rhd I\Delta_0 + \Omega_1$ is the relevant interpretation. As is well known, we can choose N is such a way that we have attractive extra properties for our sequences like closure under concatenation and the presence of projection functions. We code our syntax in N. We code assignments as finite sequences of pairs of variables and objects, satisfying the uniqueness condition. If a variable does not occur in the sequence, it is assigned a default value, say 0. We use:

- v : var, for: v is (a code of) a variable.
- ν : varseq$_n$, for: ν is a sequence of (codes of) variables of length n.
- σ : ass, for: σ is (a code of) an assignment.
- X : val, for: X is a concept consisting of assignments.
- $\sigma[v := x]$, for: the result of resetting v in σ to x. In other words,

$$\tau = \sigma[v := x] \ :\leftrightarrow \exists y, \rho, \rho' \ (\sigma = \rho * \langle\langle v, y\rangle\rangle * \rho' \wedge \tau = \rho * \langle\langle v, x\rangle\rangle * \rho') \\ \vee (\forall y, \rho, \rho' \ \sigma \neq \rho * \langle\langle v, y\rangle\rangle * \rho' \wedge \tau = \sigma * \langle\langle v, x\rangle\rangle).$$

Here are the axioms.

F1. $\forall \nu : \mathsf{varseq}_n \, \exists X \, \forall \sigma \ (X\sigma \leftrightarrow (\sigma : \mathsf{ass} \wedge P(\sigma\nu_0, \ldots, \sigma\nu_{n-1})))$,
for any P and n, where $\mathsf{ar}_U(P) = n$.
F2. $\forall X \, \forall \nu : \mathsf{varseq}_1 \, \exists Y \, \forall \sigma \ (Y\sigma \leftrightarrow (\sigma : \mathsf{ass} \wedge X(\sigma\nu_0)))$,
F3. $\forall X : \mathsf{val} \, \forall \sigma : \mathsf{ass} \, \forall v : \mathsf{var} \, \exists Y \, \forall x \ (Yx \leftrightarrow X(\sigma[v := x]))$.
F4. $\forall X \, \exists Y \, \forall x \ (Yx \leftrightarrow \neg Xx)$.
F5. $\forall X, Y \, \exists Z \, \forall x \ (Zx \leftrightarrow (Xx \wedge Yx))$.
F6. $\forall X : \mathsf{val} \, \forall v : \mathsf{var} \, \exists Y \, \forall \sigma \ (Y\sigma \leftrightarrow (\sigma : \mathsf{ass} \wedge \exists x \, X(\sigma[v := x])))$.

It is easy to see that these axioms follow from Predicative Comprehension. Conversely, we can obtain instances of Predicative Comprehension in the obvious way. Instead of specifying the procedure, let's just consider an example. We want of produce, for any z, an X such that,

$$\forall x \ (Xx \leftrightarrow \exists y \ ((Yy \wedge Pxy) \wedge Pyz)).$$

We pick variables v_0, v_1, v_2. We apply Axiom F2 to Y and $\langle v_1\rangle$ to obtain X_0. We apply Axiom F1 to P and $\langle v_0, v_1\rangle$ to obtain X_1. We apply Axiom F1 to P and $\langle v_1, v_2\rangle$ to obtain X_2. Next we apply Axiom F5 to X_0 and X_1 to obtain X_3. We apply Axiom F5 to X_3 and X_2 to obtain X_4. By applying Axiom F6 to X_4 and v_1, we obtain X_5. Finally, we apply Axiom F3 to X_5 and $\langle\langle v_2, z\rangle\rangle$ and v_0 to obtain X. We now unpack the definition of X to obtain the desired instance of Comprehension. We put $\sigma_0 := \langle\langle v_2, z\rangle, \langle v_0, x\rangle, \langle v_1, y\rangle\rangle$. We have:

$$\begin{aligned} Xx &\leftrightarrow X_5(\langle\langle v_2, z\rangle\rangle[v_0 := x]) \\ &\leftrightarrow X_5\langle\langle v_2, z\rangle, \langle v_0, x\rangle\rangle \\ &\leftrightarrow \langle\langle v_2, z\rangle, \langle v_0, x\rangle\rangle : \mathsf{ass} \wedge \exists y \, X_4(\langle\langle v_2, z\rangle, \langle v_0, x\rangle\rangle[v_1 := y]) \\ &\leftrightarrow \exists y \, X_4\langle\langle v_2, z\rangle, \langle v_0, x\rangle, \langle v_1, y\rangle\rangle \\ &\leftrightarrow \exists y \ (X_3\sigma_0 \wedge X_2\sigma_0) \\ &\leftrightarrow \exists y \ (X_0\sigma_0 \wedge X_1\sigma_0 \wedge X_2\sigma_0) \\ &\leftrightarrow \exists y \ (Y(\sigma_0 v_1) \wedge P(\sigma_0 v_0, \sigma_0 v_1) \wedge P(\sigma_0 v_1, \sigma_0 v_2)) \\ &\leftrightarrow \exists y \ (Yy \wedge P(x, y) \wedge P(y, z)). \end{aligned}$$

We turn to the issue of verifiability in $I\Delta_0 + \Omega_1$. Suppose we want to verify Predicative Comprehension for A. Note that, for any A, the number of natural big 'steps' in the proof will be of order $c|A|$, for standard c. Here $|A| := \mathsf{entier}({}^2\mathsf{log}(A + 1))$. The indices of the variables and the coded variables involved can be bounded by $||A||$. Each big step has a fixed form, where the specific parameters and parts of A or variables representing parts of A are plugged in. Moreover the number of small substeps, like computations of σv, will be of order $d|A|$, for standard d. So the size of the big step will be of order $e|A|^3$, for standard e. So the total size of the proof will be of order $f|A|^4$, for standard f. Thus, we can bound the verification π_A of comprehension for A by a p-time function. □

3.3. *The Frege function and direct interpretations*

The functor PCF acts on pointed theories. We write $(U\lceil\mathfrak{a}\rceil)^{\mathsf{pcf}}$ for $\mathsf{PCF}(U\lceil\mathfrak{a}\rceil)$. We define $\mathsf{PCF}(U\lceil\mathfrak{a}\rceil)$ as the result of adding the symbol for the Frege function ‡ and the axiom V to $\mathsf{PC}(U\lceil\mathfrak{a}\rceil)$. Here axiom V is:

$$\vdash \ddagger X =^{\mathfrak{a}} \ddagger Y \leftrightarrow \forall z^{\mathfrak{a}}(Xz \leftrightarrow Yz).$$

We define, for $U\lceil\mathfrak{a}\rceil$ and $V\lceil\mathfrak{b}\rceil$:

- $U\lceil\mathfrak{a}\rceil \rhd_{\mathrm{dir}} V\lceil\mathfrak{b}\rceil$ iff $K : U\lceil\mathfrak{a}\rceil \rhd_{\mathrm{dir}} V\lceil\mathfrak{b}\rceil$, for some K that is *direct*, i.e. K sends identity for sort $\mathfrak{b}$ to identity for sort $\mathfrak{a}$ and $\delta_K^{\mathfrak{b}}(v) :\leftrightarrow v =^{\mathfrak{a}} v$.

Lemma 3.1. *Consider pointed theories* $U\lceil\mathfrak{a}\rceil$ *and* $V\lceil\mathfrak{b}\rceil$. *Suppose* $U\lceil\mathfrak{a}\rceil \rhd_{\mathrm{dir}} V\lceil\mathfrak{b}\rceil$. *Then,* $(U\lceil\mathfrak{a}\rceil)^{\mathsf{pcf}} \rhd_{\mathrm{dir}} (V\lceil\mathfrak{a}\rceil)^{\mathsf{pcf}}$. *This fact is verifiable in* $I\Delta_0 + \Omega_1$.

The proof of the lemma is easy.

4. The hierarchy $\mathsf{P}^n\mathsf{V}$, a first round

In this section, we provide some basic insights concerning the hierarchy $\mathsf{P}^n\mathsf{V}$. We call the sort of objects 0 and we call the sorts of concepts 1, 2, ... We make 0 the designated sort. We write x for x^0 and X^n for x^{n+1}. Here are some definitions.

- $x \in^{j+1} y :\leftrightarrow \exists Y^j\, (Y^j x \wedge y = \ddagger^j Y)$,
- $\{x \mid Ax\}^{j+1} := \ddagger^j \wr x \mid Ax \wr^{j+1}$.[7]

We remind the reader that $\mathsf{P}^{n+2}\mathsf{V}$ is $(\mathsf{P}^{n+1}\mathsf{V})^{\mathsf{pcf}}$ plus the following variant of V:

$$\vdash \ddagger^{n+1} X^{n+1} = \ddagger^n Y^n \leftrightarrow \forall z\, (Xz \leftrightarrow Yz).$$

4.1. Dropping a variant of V

In this subsection, we show that the variant of V can be omitted.

Theorem 4.1. *We have:* $\mathsf{P}^{n+2}\mathsf{V} \equiv_{\mathrm{dir}} (\mathsf{P}^{n+1}\mathsf{V})^{\mathsf{pcf}}$. *This fact is verifiable in* $I\Delta_0 + \Omega_1$.

Proof. In one direction this is trivial. In the other direction, we have two problems to solve. The first is that the formulas for which PCF provides predicative comprehension do not contain the new Frege function. The second is the extra variant of V we have in $\mathsf{P}^{n+2}\mathsf{V}$.

The first problem is solved by an observation of Allen Hazen. Consider an instance of predicative comprehension for degree $n+1$ involving the formula $A(x, \vec{y}, \vec{Y})$, which does contain Frege terms of the form $\ddagger^{n+1}Z$. Since the concept variables occurring in these Frege terms are free in A, for some B, we have $A = B(x, \vec{y}, \vec{z}, \vec{Y}_0)[\vec{z} := \ddagger^{n+1}\vec{Y}_1]$. Here B does not contain Frege terms of the from $\ddagger^{n+1}Z$. The intersection of Y_0 and Y_1 is allowed to be non-empty. It is immediate that comprehension for A follows from comprehension for B by universal instantiation.

We turn to the second problem. We let the 'Julius Caesar indeterminacy' of Frege's system work in our favor. We define $M : (\mathsf{P}^{n+1}\mathsf{V})^{\mathsf{pcf}} \rhd \mathsf{P}^{n+2}\mathsf{V}$ as follows. The interpretation M is the identity interpretation except for the interpretation of the Frege functions. Note that in $\mathsf{P}^1\mathsf{V}$ we have pairing and two distinct objects $0 := \emptyset^1$ and $1 := \{\emptyset\}^1$. We set:

- for $j \leq n$, $\mathsf{freg}^j_M(X^j, u) :\leftrightarrow u = \langle 0, \ddagger^j X^j\rangle$;
- $\mathsf{freg}^{n+1}_M(X^{n+1}, u) :\leftrightarrow \exists Y^n\, (\forall z\, (X^{n+1}z \leftrightarrow Y^n z) \wedge u = \langle 0, \ddagger^n Y^n\rangle)$
  $\vee (\neg \exists Y^n\, \forall z\, (X^{n+1}z \leftrightarrow Y^n z) \wedge u = \langle 1, \ddagger^{n+1} X^{n+1}\rangle)$.

We easily see that this works. Verifiability in $I\Delta_0 + \Omega_1$ is clear. □

By Theorem 4.1 and Lemma 3.1, we may conclude:

Corollary 4.2. $\mathsf{P}^{n+1}\mathsf{V} \equiv_{\mathrm{dir}} \mathsf{PCF}^{n+1}(\mathsf{ID})$, *where* ID *is one-sorted predicate logic with only the identity predicate and with single designated sort* 0.

Prima facie, we need exponentiation to verify Corollary 4.2, since for each increase from $n+1$ to $n+2$, we have a binary splitting. We think that by being slightly more careful the corollary should be verifiable in $I\Delta_0 + \Omega_1$, but we did not work out the proof.

---

7 Par abus de langage, we also use accolades for virtual classes.

4.2. Nothing new beyond ω

A trivial, but still surprising insight is that hierarchies like the PV-hierarchy stop at ω modulo local interpretability. We formulate this insight in a somewhat frivolous form. Let's momentarily drop our restriction to finite signatures and simple axiom sets. Let $\mathcal{P}$ be any partial ordering (of whatever cardinality). We define $\mathsf{P}^{\mathcal{P}}\mathsf{V}$ as follows. It is a many-sorted theory with sorts $\mathfrak{o}$ (of objects) and p of concepts, for any p in $\mathcal{P}$. We have identity for sort $\mathfrak{o}$, application predicates app^p with characteristic sequence $p\mathfrak{o}$, and Frege functions $\ddagger^p$ with characteristic sequence $p\mathfrak{o}$. We write $Y^p x$ for $\mathsf{app}^p(y^p, x^{\mathfrak{o}})$. Our theory is axiomatized as follows.

$\mathcal{P}1)$ $\vdash \exists X^p\, \forall x\, (X^p x \leftrightarrow A(x, \vec{y}, \vec{Y}))$,
where A does not contain free occurrences of X, where A does only contain concept variables Y^q, for $q \leq p$, where A does only contain bound concept variables Z^q, for $q < p$.
$\mathcal{P}2)$ $\vdash \ddagger^p X^p = \ddagger^q Y^q \leftrightarrow \forall z\, (X^p z \leftrightarrow Y^q z)$.

Note that, modulo some notational divergence, $\mathsf{P}^n\mathsf{V}$ under the old and under the new definition are the same (if we view n as a finite ordinal, modeled as the standard ordering on its predecessors). We write $[\![\mathcal{P}]\!]$ for the supremum of the lengths of all finite ascending sequences in $\mathcal{P}$. Note that $[\![\mathcal{P}]\!] \in \omega + 1$. We have:

Theorem 4.3. *We have:* $\mathsf{P}^{\mathcal{P}}\mathsf{V} \equiv_{\mathsf{loc}} \mathsf{P}^{[\![\mathcal{P}]\!]}\mathsf{V}$.

Proof. We use, locally, the notational conventions of $\mathsf{P}^{\mathcal{P}}\mathsf{V}$ for $\mathsf{P}^n\mathsf{V}$ and $\mathsf{P}^{\omega}\mathsf{V}$.

Consider any finitely axiomatized subtheory U of $\mathsf{P}^{\mathcal{P}}\mathsf{V}$. Let $\mathcal{P}_0$ be the subordering of $\mathcal{P}$ generated by the p occurring in the axioms of U. We define ν from $\mathcal{P}_0$ to ω by: $\nu(p) := \sup\{\nu(q) + 1 \mid q < p\}$. For p in $\mathcal{P}_0$, we send sort p to sort $\nu(p)$, app^p to $\mathsf{app}^{\nu(p)}$, etc. For all other sorts, the translation is don't care. It is easily seen that this translation yields the desired interpretation of U in $\mathsf{P}^{[\![\mathcal{P}]\!]}\mathsf{V}$. The other direction is similar. $\Box$

Specifically, we get for all infinite ordinals α: $\mathsf{P}^{\alpha}\mathsf{V} \equiv_{\mathsf{loc}} \mathsf{P}^{\omega}\mathsf{V}$.

4.3. Predicative Frege set theory

This subsection is devoted to explicating something that is in a sense known to everybody who studied these matters: there is a natural one-sorted version of the predicative Frege hierarchy that is in some sense 'the same'. However, what is sameness here? This subsection is devoted to (i) a precise description of the flat hierarchy and (ii) establishing that the relevant notion of sameness is bi-interpretability. The Frege functions will play the role of *isomorphism between interpretations*. It is certainly remarkable to see how 'thin' Frege's strict distinction between concept and object is from the technical point of view.

Because of their elementary character, all considerations of this subsection are verifiable in $I\Delta_0 + \Omega_1$.

We define the flat Frege hierarchy FST^n as follows.[8] The theory FST^0 is simply ID, the one-sorted predicate logic of pure identity. The language of FST^{n+1} adds to the language of FST^n a unary predicate symbol set^{n+1} and binary predicate symbol $\in^{n+1}$. A formula A is in $\mathcal{A}^x_{n+1}$, iff, whenever a variable y occurs in a context $\mathsf{set}^{n+1}(y)$ or $z \in^{n+1} y$ or $y \in^{n+1} y$, then y is not syntactically equal to x and the occurrence under consideration is free in A. The theory FST^{n+1} is FST^n plus the following axioms.

- $\vdash x \in^{n+1} y \to \mathsf{set}^{n+1}(y)$.
- For any A in $\mathcal{A}^x_{n+1}$, such that y is not free in A, we have:
  $\vdash \exists y : \mathsf{set}^{n+1}\, \forall x\, (x \in^{n+1} y \leftrightarrow Ax\vec{z})$.
- $\vdash (x : \mathsf{set}^{n+1} \wedge y : \mathsf{set}^n) \to (x = y \leftrightarrow \forall z\, (z \in^{n+1} x \leftrightarrow z \in^n y))$.
  (We only have this axiom if $n > 0$.)
- $\vdash x, y : \mathsf{set}^{n+1} \to (x = y \leftrightarrow \forall z\, (z \in^{n+1} x \leftrightarrow z \in^{n+1} y))$.

One can show that FST^n is, *modulo definitional equivalence*, the extension of $(\mathsf{P}^n\mathsf{V})^\flat$ with an axiom stating that everything is an object: $\vdash \forall x\, \Delta^0(x)$, plus axioms saying that the Frege functions are identity functions:

$$\vdash \forall x : \mathsf{set}^{j+1}\, \forall y\, (\mathsf{freg}^j(x, y) \leftrightarrow x = y).$$

The interpretations of sets in classes is $\mathsf{SC} : \mathsf{P}^{n+1}\mathsf{V} \rhd \mathsf{FST}^{n+1}$. The interpretation of classes in sets is $\mathsf{CS} : \mathsf{FST}^{n+1} \rhd \mathsf{P}^{n+1}\mathsf{V}$. Here is the specification of SC.

- $\delta_{\mathsf{SC}}(v) :\leftrightarrow v = v$,
- $vE_{\mathsf{SC}}w :\leftrightarrow v = w$,

8 FST for: Frege Set Theory.

- $\mathsf{set}^{j+1}_{\mathsf{SC}}(v) :\leftrightarrow \exists X^j\ v = \ddagger^j X^j$ $(j = 0, \ldots, n)$,
- $v \in^{j+1}_{\mathsf{SC}} w :\leftrightarrow \exists Y^j\ (w = \ddagger^j Y^j \wedge Y^j v)$ $(j = 0, \ldots, n)$.

We can show that SC is indeed an interpretation of the flat theory in its sorted counterpart. We treat the comprehension axiom. Suppose $Ax\vec{z}$ is in $\mathcal{A}^x_{n+1}$. Consider $Bx\vec{z} := (Ax\vec{z})^{\mathsf{SC}}$. It is clearly sufficient to produce, in $\mathsf{P}^{n+1}\mathsf{V}$, for all $\vec{z}$, an Y such that $\forall x\ (Yx \leftrightarrow Bx\vec{z})$. Let's say that $C\vec{z}$ is a *context* for $Bx\vec{z}$, if C is a conjunction of formulas C_i, where C_i is of one of the forms $(\mathsf{set}^{n+1}(z_i))^{\mathsf{SC}}$ or $\neg\,(\mathsf{set}^{n+1}(z_i))^{\mathsf{SC}}$, for each z_i in $\vec{z}$. Since, the disjunction of all contexts is a tautology, it is sufficient to produce Y under the assumption of some context, say C. We work in $\mathsf{P}^{n+1}\mathsf{V} + C$.

Any subformula of the form $(u \in^{j+1} v)^{\mathsf{SC}}$ or $(\mathsf{set}^{j+1}(v))^{\mathsf{SC}}$, for $j < n$, only contains bound concept variables V^j.

Consider any subformula occurrence $(u \in^{n+1} v)^{\mathsf{SC}}$ or $(\mathsf{set}^{n+1}(v))^{\mathsf{SC}}$. Here, v will be free in B and unequal to x. Thus, either $(\mathsf{set}^{n+1}(v))^{\mathsf{SC}}$ or $\neg\,(\mathsf{set}^{n+1}(v))^{\mathsf{SC}}$. will be in the context. In the first case, we have a variable V^n such that $\ddagger^n V = v$. It follows that $(\mathsf{set}^{n+1}(v))^{\mathsf{SC}}$ may be replaced, modulo provable equivalence, by $\top$, and that $(u \in^{n+1} v)^{\mathsf{SC}}$ may be replaced by Vu. Both replacements do not contain quantifiers over variables of degree n. In the second case we may replace both $(u \in^{n+1} v)^{\mathsf{SC}}$ and $(\mathsf{set}^{n+1}(v))^{\mathsf{SC}}$ by $\bot$. So in all cases B reduces to a formula without quantifiers over variables of the form V^n. We may now apply predicative comprehension in $\mathsf{P}^{n+1}\mathsf{V}$, to obtain Y. (Note that we implicitly use $\exists$-elimination.)

Next we specify CS.

- $\delta^0_{\mathsf{CS}}(v) :\leftrightarrow v = v$, $\delta^{j+1}_{\mathsf{CS}}(v) :\leftrightarrow \mathsf{set}^{j+1}(v)$ $(j = 0, \ldots, n)$,
- $vE^0_{\mathsf{CS}}w :\leftrightarrow v = w$,
- $\mathsf{app}^{j+1}_{\mathsf{CS}}(w, v) :\leftrightarrow v \in^{j+1} w$,
- $\mathsf{freg}^{j}_{\mathsf{CS}}(v, w) :\leftrightarrow v = w$.

We easily check that CS is indeed an interpretation of the sorted theory in its flat companion. We show that $\mathsf{CS} \circ \mathsf{SC}$ is the identity interpretation modulo FST^{n+1}-provable equivalence. Thus, $\mathsf{CS} \circ \mathsf{SC}$ is equal to the identity interpretation for FST^{n+1} in $\mathsf{INT}^{\mathsf{ms}}$. This tells us that FST^{n+1} is a *retract* of $\mathsf{P}^{n+1}\mathsf{V}$ in $\mathsf{INT}^{\mathsf{ms}}$. We treat the cases of δ and $\in$. We have in FST^{n+1}:

$$\begin{aligned}
\delta_{\mathsf{CS}\circ\mathsf{SC}}(x) &\leftrightarrow \delta^0_{\mathsf{CS}}(x) \wedge (\delta_{\mathsf{SC}}(x))^{\mathsf{CS}} \\
&\leftrightarrow x = x. \\
x \in^{j+1}_{\mathsf{CS}\circ\mathsf{SC}} y &\leftrightarrow (x \in^{j+1}_{\mathsf{SC}} y)^{\mathsf{CS}} \\
&\leftrightarrow (\exists Y^j\ (y = \ddagger^j Y^j \wedge Y^j x))^{\mathsf{CS}} \\
&\leftrightarrow \exists z : \mathsf{set}^{j+1}\ (y = z \wedge x \in^{j+1} z) \\
&\leftrightarrow x \in^{j+1} y.
\end{aligned}$$

Let $\nabla := \mathsf{SC} \circ \mathsf{CS}$. We show that ∇ is isomorphic to the identity interpretation id on $\mathsf{P}^{n+1}\mathsf{V}$. The isomorphism from id to ∇ is specified as follows:

- $xG^0y :\leftrightarrow x = y$,
- $X^jG^{j+1}y :\leftrightarrow \ddagger^n(X^j) = y$.

We have e.g. in $\mathsf{P}^{n+1}\mathsf{V}$:

$$\begin{aligned}
\delta^{j+1}_{\nabla}(\ddagger^j X^j) &\leftrightarrow \exists Y^j\ (\ddagger^j X^j = \ddagger Y^j) \\
&\leftrightarrow \top. \\
\mathsf{app}^{j+1}_{\nabla}(\ddagger^j X^j, y) &\leftrightarrow \exists Z^j\ (\ddagger^j X^j = \ddagger^j Z^j \wedge Z^j y) \\
&\leftrightarrow X^j y. \\
\mathsf{freg}^{j}_{\nabla}(\ddagger X^j, y) &\leftrightarrow \ddagger X^j = y.
\end{aligned}$$

We may conclude that $\mathsf{P}^{n+1}\mathsf{V}$ and FST^{n+1} are isomorphic in $\mathsf{hINT}^{\mathsf{ms}}$. In other words, these theories are bi-interpretable.

Open Question 4.4. Is $\mathsf{P}^{\omega}\mathsf{V}$ bi-interpretable with a theory of finite signature?

We end this subsection, with a useful insight concerning the FST^{n+1}.

Theorem 4.5. *Each theory FST^{n+1} is sequential. This fact is verifiable in $I\Delta_0 + \Omega_1$.*

Proof. It is sufficient to show that FST^1 is sequential. Burgess shows how to interpret Q in $\mathsf{P}^1\mathsf{V}$. This interpretation preserves identity. We transfer this interpretation to FST^1 via CS. We can improve the resulting interpretation by shortening it in order to have the principles stating that $<$ is a linear ordering. We define sequences in the way that is usual in set theory: as functions from the numbers below a number n to arbitrary objects. Here functions are modeled as sets of ordered pairs. It is easy to verify that we have the desired properties.

Verifiability in $I\Delta_0 + \Omega_1$ is evident, since we only have to show direct interpretability of a standardly finite number of principles. □

Remark 4.6. In stead of using Burgess result in the proof of Theorem 4.5, we could also have used the result from [3] or from [11].

We note that the interpretation of Q can be given in the system adjunctive class theory with Frege relation, acf, which is modulo a direct interpretation a subsystem of $\mathsf{P}^1\mathsf{V}$. The language of acf has two sorts: one for objects and one for concepts. We have identity for objects and two further binary relations η from objects to concepts and $\digamma$ from concepts to objects. We have the following axioms.

acf1. $\vdash \exists X\ \forall x\ \neg\, x\eta X$.
acf2. $\vdash \forall X, x\, \exists Y\ \forall y\ (y\eta Y \leftrightarrow (y\eta X \vee y = x))$.
acf3. $\vdash \forall X\ \exists x\ X \digamma x$.
acf4. $\vdash \forall X, Y, z\ ((X \digamma z \wedge Y \digamma z) \to \forall u\ (u\eta X \leftrightarrow u\eta Y))$.

So, roughly, acf is $\mathsf{P}^1\mathsf{V}$, with a weaker comprehension principle and with axiom V minus the uniqueness condition of the Frege function and without extensionality. We can now use the ideas sketched in Appendix III of [12] to prove that (the flattening of) this theory is sequential. Since we lack extensionality, the interpretation N of Q will not preserve identity.

As is easily seen, if we combine axioms W3 and W4 with full comprehension, we still get the Russell paradox.

4.4. Finite axiomatizability II

Burgess shows that the theory $\mathsf{P}^1\mathsf{V}$ is finitely axiomatizable: the concepts definable are generated from empty concept, singleton concept using complement and intersection. See [2], pp. 89, 90. Since $\mathsf{P}^1\mathsf{V}$ has a pairing operation, we would like to apply Theorem 3.3 to conclude that all the $\mathsf{P}^{n+1}\mathsf{V}$ are finitely axiomatizable. However, Theorem 3.3 is only formulated for one-sorted theories. We show how to work around this problem. First we need two lemmas.

Lemma 4.1. *Finite axiomatizability is preserved over retractions in* $\mathsf{hINT}^{\mathsf{ms}}$. *It follows that bi-interpretability preserves finite axiomatizability. This fact is verifiable in* $I\Delta_0 + \Omega_1$.

Proof. Suppose U is a retract of V in $\mathsf{hINT}^{\mathsf{ms}}$ and that V is finitely axiomatized. We can now axiomatize U by the translations of the axioms of V, plus axioms stating that the $G^{\mathfrak{a}}$ form an isomorphism between the identity interpretation and the composition of co-retraction and retraction. □

Lemma 4.2. *Suppose* $U\lceil\mathfrak{a}\rceil$ *and* $V\lceil\mathfrak{b}\rceil$ *are bi-interpretable via direct interpretations. Then* $(U\lceil\mathfrak{a}\rceil)^{\mathsf{pc}}$ *and* $(V\lceil\mathfrak{b}\rceil)^{\mathsf{pc}}$ *are bi-interpretable via direct interpretations, i.e. interpretations preserving the domain and the identity of the designated sorts. This theorem can be verified in* $I\Delta_0 + \Omega_1$.

Proof. Suppose the witnessing interpretations are $M : U\lceil\mathfrak{a}\rceil \to V\lceil\mathfrak{b}\rceil$ and $N : V\lceil\mathfrak{b}\rceil \to U\lceil\mathfrak{a}\rceil$. We lift M and N to M^{pc} and N^{pc} as in Theorem 3.2. Note that the directness causes the lifted interpretations to act identically on the second order vocabulary. We can now lift the isomorphisms between $M \circ N$ and id_V and between $N \circ M$ and id_U. E.g. if G is the isomorphism from id_U to $N \circ M$, then we set: $XG^{\mathfrak{c}}Y :\leftrightarrow \forall x, y\ (xG^{\mathfrak{a}}y \to (Xx \leftrightarrow Yy))$. □

We are now ready to prove finite axiomatizability of the $\mathsf{P}^{n+1}\mathsf{V}$.

Theorem 4.7. *Each theory* $\mathsf{P}^{n+1}\mathsf{V}$ *is finitely axiomatizable.*

Proof. We have already seen that $\mathsf{P}^1\mathsf{V}$ is finitely axiomatizable. For the induction step, suppose $\mathsf{P}^{n+1}\mathsf{V}$ is finitely axiomatizable. Since, $\mathsf{P}^{n+1}\mathsf{V}$ is bi-interpretable with FST^{n+1}, we find that FST^{n+1} is finitely axiomatizable. This is a one-sorted sequential theory. By Theorem 3.4, $(\mathsf{FST}^{n+1})^{\mathsf{pc}}$ is finitely axiomatizable. By Lemma 4.2, $(\mathsf{FST}^{n+1})^{\mathsf{pc}}$ is bi-interpretable with $(\mathsf{P}^{n+1}\mathsf{V})^{\mathsf{pc}}$. It follows that $(\mathsf{P}^{n+1}\mathsf{V})^{\mathsf{pc}}$ is finitely axiomatizable. Since $\mathsf{P}^{n+2}\mathsf{V}$ is a finite extension of $(\mathsf{P}^{n+1}\mathsf{V})^{\mathsf{pc}}$, we are done. □

Open Question 4.8. We do not know whether the argument for the finite axiomatizability of $\mathsf{P}^1\mathsf{V}$ can be formalized in $I\Delta_0 + \Omega_1$.

1. Can $I\Delta_0 + \Omega_1$ verify the finite axiomatizability of $\mathsf{P}^1\mathsf{V}$?
2. If not, is the finite axiomatizability of $\mathsf{P}^{n+1}\mathsf{V}$, for $n \geq 1$, verifiable in $I\Delta_0 + \Omega_1$?

The finite axiomatizations of the stages provided by the proof of Theorem 4.7 are not optimal since we are going back and forth using the interpretations SC and CS. Here is the finite axiomatization of $\mathsf{P}^{n+1}\mathsf{V}$ after some simplifications. The axioms Fi are taken from the proof of Theorem 3.4.

- The axioms of identity for the object sort.
- The variants of axiom V for all concept sorts occurring in $\mathsf{P}^{n+1}\mathsf{V}$.
- The finite axiomatization of comprehension of $\mathsf{P}^1\mathsf{V}$.
- The axioms F1 for $\in^j$ and set^j, for $0 < j \leq n$, and the concept variables X^k, for $j \leq k \leq n$. Here $\in^j$ and set^j are treated as abbreviations.
- The axioms F2 to F6 for concept variables X^j, for $0 < j \leq n$.

We call the theories with the above axiomatization $\mathsf{P}^{n+1}\mathsf{V}_{\mathsf{fa}}$. So Theorem 4.7 tells us that $\mathsf{P}^{n+1}\mathsf{V}_{\mathsf{fa}} =_{\mathsf{ext}} \mathsf{P}^{n+1}\mathsf{V}$. We also have:

Theorem 4.9. *The theory $I\Delta_0 + \Omega_1$ verifies, that, for all n:*

- $\mathsf{P}^{n+2}\mathsf{V}_{\mathsf{fa}} =_{\mathsf{ext}} (\mathsf{P}^{n+1}\mathsf{V}_{\mathsf{fa}})^{\mathsf{pcf}}$,
- $\mathsf{P}^{n+1}\mathsf{V}_{\mathsf{fa}} \subseteq \mathsf{P}^{n+1}\mathsf{V}$.

5. From consistency to comprehension

In this section, we treat the Henkin–Feferman construction and show how it can be extended to an interpretation of predicative comprehension.

5.1. The Henkin–Feferman construction

We briefly and informally discuss the Henkin–Feferman construction. Since, we will use some details of the construction in some of our proofs, it is good to have, at least in outline, a sketch in mind of how the proof works. The Henkin–Feferman construction is a 'syntactification' of the Henkin construction of a model from a consistent theory. Here, we do not construct a model but an interpretation. To complicate things we execute the construction in the context of a weak theory, so that apparently there is not enough induction around. The lack of induction is compensated using Solovay's methodology of shortening cuts. Details can be found in [18].[9]

Remark 5.1. The early history of the Henkin–Feferman construction is discussed in [5].

In their book [8], in 1939, Hilbert and Bernays gave a formalization of Gödel's Completeness Theorem, formalizing Gödel's own construction. The result was extended, in 1951, by Hao Wang in his paper [21]. One could say this gave us the Gödel-(Hilbert+Bernays)-Wang construction.

Then, in 1960, in his classical paper [4], Solomon Feferman further improved the result, using an arithmetical construction based on the Henkin construction. This gave us the Henkin–Feferman construction.

Feferman's result employed Δ^0_2-induction. This can be improved using Solovay's method of shortening definable cuts. Solovay found his method in 1976. It is reported in the unpublished note [16]. For an exposition, see e.g. [7], V.5. This improvement leads to the insight that the construction can be done when we have Robinson's Q available. I guess, this is optimal: whenever the result can be meaningfully formulated, we have it. The construction in weak theories was known as folklore to the specialists. The first detailed exposition of the construction in the context of weak theories is [17]. This exposition was improved in [18].

There are further variants of the construction involving cut free consistency, restricted consistency, non-standard proof predicates and the like, that are outside the scope of this paper.

The development of the constructions in this subsection can be executed in $I\Delta_0 + \Omega_1$ as metatheory. Remember that $\Omega_U := I\Delta_0 + \Omega_1 + \mathsf{con}(U)$.

Consider any theory U. We want to construct an interpretation $H : \Omega_U \rhd U$.

We extend the language of U with Henkin constants in an inductive way: whenever we have a sentence $\exists x^{\mathfrak{a}}\, Ax$ in the language, we add a constant $c[\exists x^{\mathfrak{a}}\, Ax]$ of sort $\mathfrak{a}$. The formula B is *a direct subformula in the extended sense* of A, if B is a direct subformula of A in the usual sense or if $B = \exists x^{\mathfrak{a}}\, B_0 x$ and $c[B]$ occurs in A. We stipulate: *subformula in the extended sense* is the reflexive transitive closure of *subformula in the extended sense*.

We arrange that the language is coded in such a way that all syntactic operations are p-time and that $I\Delta_0 + \Omega_1$ can verify all elementary facts. The coding will also satisfy monotonicity in the sense that, if B is a proper subformula in the extended sense of A, then the code of B is smaller than the code of A. To simplify inessentially we will assume that our only official quantifier is $\exists$.

Reason in Ω_U. A *definable cut* will be a class of numbers given by a formula, such that, verifiably, this class is closed under 0, S, $+$, $\times$ and ω_1, and is downwards closed w.r.t. $<$. We construct a complete Henkin theory in steps.

step 0 The theory $\mathcal{H}_0$ is U.

step $n+1$ In case n is not the Gödelnumber of a sentence of the extended language, we take $\mathcal{H}_{n+1} := \mathcal{H}_n$. Suppose n is the Gödelnumber of A. Suppose $\mathcal{H}_n + A$ is consistent. If A is not of the form $\exists x^{\mathfrak{a}}\, Bx$, we take $\mathcal{H}_{n+1} := \mathcal{H}_n + A$. If A is of the form $\exists x^{\mathfrak{a}}\, Bx$, we take $\mathcal{H}_{n+1} := \mathcal{H}_n + A + B(c[A])$. If $\mathcal{H}_n + A$ is not consistent, we take $\mathcal{H}_{n+1} := \mathcal{H}_n$.

We can easily show that the set X of n, such that $\mathcal{H}_n$ is defined and consistent, is closed under 0 and successor. Let $\mathfrak{I}$, or, more explicitly, $\mathfrak{I}_U$, be a definable cut that is a shortening of X. We have that $\mathcal{H} := \mathcal{H}_{\mathfrak{I}} := \bigcup_{i\in\mathfrak{I}} \mathcal{H}_i$ is a complete Henkin theory in the language $\mathcal{S}$, of all sentences of the extended language of U restricted to $\mathfrak{I}$. We define an interpretation H, or more explicitly H_U, as follows.

[9] In fact, only the one-sorted case is treated there. However, the many-sorted case only asks for very minor adaptations.

- $\delta^{\mathfrak{a}}_H$ is the set of all Henkin constants of sort $\mathfrak{a}$ in $\Im$. We will distinguish $x^{\mathfrak{a}}$ qua domain object from $c_{x^{\mathfrak{a}}}$, which is $x^{\mathfrak{a}}$ in its role of Henkin constant, even if strictly speaking $x^{\mathfrak{a}} = c_{x^{\mathfrak{a}}}$.
- $R_H(\vec{x}) :\leftrightarrow \mathcal{H}(R\vec{c}_x)$; $xE^{\mathfrak{a}}_H y :\leftrightarrow \mathcal{H}(c_x =^{\mathfrak{a}} c_y)$, if we have identity for sort $\mathfrak{a}$.

We have the following theorem.

Theorem 5.2. 1. *The theory Ω_U proves the Tarski clauses, w.r.t. $\mathcal{S}$, for $\mathcal{H}_U$ as a truth predicate of the interpretation H_U. For example, in Ω_U, we have:*
- *for a predicate R with associated sequence $\vec{\mathfrak{a}}$ and for $\vec{x} : \delta^{\vec{\mathfrak{a}}}_H$, we have: $R_H(\vec{x}) :\leftrightarrow \mathcal{H}(R(\vec{c}_x))$,*
- *for all B, C in $\mathcal{S}$, we have: $\mathcal{H}(B \wedge C) \leftrightarrow (\mathcal{H}(B) \wedge \mathcal{H}(C))$,*
- *for all B in $\mathcal{S}$ and all variables $v^{\mathfrak{a}}$ of the appropriate sort (coded) in $\Im$, we have: $\mathcal{H}(\forall v^{\mathfrak{a}}\, Bv) \leftrightarrow \forall x : \delta^{\mathfrak{a}}_H\ \mathcal{H}(Bc_x)$.*

Note that it follows, that, for any A in the language of U,
$\Omega_U \vdash \forall \vec{x} : \delta^{\vec{\mathfrak{a}}}_H\ (\mathcal{H}(A\vec{c}_x) \leftrightarrow A^H\vec{x})$.

2. *Let $\mathcal{S}_0$ be the set of U-sentences in $\Im$. We have:*
$\Omega_U \vdash \forall A{\in}\mathcal{S}_0\ (\Box_U A \to \mathcal{H}(A))$.

The Henkin interpretation as defined here has a remarkable property, that becomes immediately evident, when we reflect on the fact that any cut shortening the set of acceptable stages X would have done the trick. This property is stated in the theorem below.

Consider any interpretation $K : T \rhd Z$. We assume that σ_K maps all Z-sorts to a single T-sort, say $\mathfrak{b}$. Consider any T-formula $Av^{\mathfrak{b}}$, having only v free, Suppose $T \vdash \delta^{\mathfrak{a}}_K \cap A \neq \emptyset$, for all Z-sorts $\mathfrak{a}$. We define $K \restriction A$ as the interpretation we obtain by restricting the domains $\delta^{\mathfrak{a}}_K$ of K to $\delta^{\mathfrak{a}}_K \cap A$.

Lemma 5.1. *Consider any theory U. Let J be any Ω_U-cut. We have:*

$$\Omega_U \vdash \forall \vec{x} : \delta^{\vec{\mathfrak{a}}}_H \cap J\ (A^{H\restriction J}\vec{x} \leftrightarrow A^H\vec{x}).$$

(The notation '$\delta^{\vec{\mathfrak{a}}}_H \cap J$' is intended to convey that each $\delta^{\mathfrak{a}}_K$ *is intersected with J.) In other words $H \restriction J$ is elementarily equivalent in Ω_U to H.*

Proof. Clearly, the elements of the $\delta^{\mathfrak{a}}_H \cap J$ are precisely the Henkin constants of sort $\mathfrak{a}$ for the complete Henkin theory $\mathcal{H} \cap J$. We have, in Ω_U, for $\vec{x} : \delta^{\vec{\mathfrak{a}}}_H \cap J$:

$$\begin{aligned} A^{H\restriction J}\vec{x} &\leftrightarrow (\mathcal{H} \cap J)(A\vec{c}_x) \\ &\leftrightarrow \mathcal{H}(A\vec{c}_x) \\ &\leftrightarrow A^H\vec{x}. \end{aligned}$$

Note that, since A is standard, $A\vec{c}_x$ will be automatically in J, whenever $\vec{x}$ in J. □

Note that we can easily prove a stronger fact.

Fact 5.3. *Suppose $M : Z \rhd \Omega_U$. Let J be a Z-definable cut in the M-numbers. Then, $(M \circ H_U) \restriction J$ is elementarily equivalent in Z to $M \circ H_U$.*

5.2. Henkin–Feferman meets comprehension

To give an optimal formulation of a number of our results, it is better to widen the framework a little bit. Consider a many-sorted predicate logical language and a designated sort $\mathfrak{b}$. We extend the language by adding new unary second order variables for concepts of objects of the sort $\mathfrak{b}$. Officially: we add a new sort $\mathfrak{c}$, plus a predicate app with associated sequence $\mathfrak{cb}$. An *s-theory* is a theory in this extended language *without quantifiers over the new sort* $\mathfrak{c}$. We map an s-theory U to a theory U^{sch} in the original language, by replacing all axioms $A\vec{X}$ by $A[\vec{B}]$, for all $\vec{B}$, where the $\vec{B}$ are formulas of the original language, and where $A[\vec{B}]$ is the result of replacing the $X_i x$ in $A\vec{X}$, by $B_i[v := x]$, where v is a fixed designated free variable of sort $\mathfrak{b}$, while no occurrences, from the free variable occurrences in $B_i[v := x]$, distinct from the occurrences of x are bound in $A[\vec{B}]$.

The second order theory U^{pc} is obtained from U by reading U second order and by adding the predicative comprehension scheme:

$$\vdash \exists X\, \forall x^{\mathfrak{b}}\ (Xx \leftrightarrow A(x, \vec{y}, \vec{Y})),$$

where A does not contain bound second order variables and X does not occur in A.

Theorem 5.4. *Suppose U is an s-theory with designated sort $\mathfrak{b}$. Then, we have, verifiably in $I\Delta_0 + \Omega_1$, that* $(\mathsf{Q}{+}\mathsf{con}(U^{\mathsf{sch}})) \rhd U^{\mathsf{pc}}$.

Proof. We first move to a more convenient environment to work in. Remember that $(\mathsf{Q} + \mathsf{con}(U^{\mathsf{sch}})) = \Theta_{U^{\mathsf{sch}}} \rhd \Omega_{U^{\mathsf{sch}}}$. So, it is sufficient to show: $\Omega_{U^{\mathsf{sch}}} \rhd U^{\mathsf{pc}}$. We specify our interpretation $H^\star := H^\star_U : \Omega_{U^{\mathsf{sch}}} \rhd U^{\mathsf{pc}}$ as follows. For all vocabulary of the language without $\mathfrak{c}$, $H^\star$ coincides with H. The domain of concepts, $\delta^{\mathfrak{c}}_{H^\star}$, is the class of formulas, in $\Im_{U^{\mathsf{sch}}}$, of the language of U^{sch} extended with Henkin constants, that have just one designated variable v of sort $\mathfrak{b}$ free. We interpret, for X in $\delta^{\mathfrak{c}}_{H^\star}$, say X is Bv, Xx, by $\mathcal{H}(Bc_x)$.

When we consider a sequence $\vec{X}$ of elements of $\delta^{\mathfrak{c}}_{H^\star}$, we will call these objects in their role of formulas: $\vec{B}_X$. Consider any U-formula $C(\vec{Z})$. We can show by external induction over C, that:

$$\Omega_{U^{\mathsf{sch}}} \vdash \forall \vec{x} : \delta^{\vec{\mathfrak{a}}}_{H^\star}\, \forall \vec{X} : \delta^{\mathfrak{c}}_{H^\star}\; ((C(\vec{X}, \vec{x}))^{H^\star} \leftrightarrow \mathcal{H}(C[\vec{B}_X](\vec{c}_x))). \tag{2}$$

Consider any s-axiom $A\vec{X}$ of U. Reason in $\Omega_{U^{\mathsf{sch}}}$. It is clear that, for all $\vec{X}$ in $\delta^{\mathfrak{c}}_{H^\star}$, we have $\Box_{U^{\mathsf{sch}}} A[\vec{B}_X]$. Hence, by Theorem 5.2(2), we have $\mathcal{H}(A[\vec{B}_X])$. (Note that $A[\vec{B}_X]$, will be in $\mathcal{S}_0$.) It follows by Eq. (2), that $(A\vec{X})^{H^\star}$. Thus, outside $\Omega_{U^{\mathsf{sch}}}$, we may conclude that $\Omega_{U^{\mathsf{sch}}} \vdash (\forall \vec{X}\; A\vec{X})^{H^\star}$.

Consider any formula $A(v, \vec{z}, \vec{Z})$, where v is of the designated sort $\mathfrak{b}$, without bound $\mathfrak{c}$-variables. We reason in $\Omega_{U^{\mathsf{sch}}}$. Consider $\vec{y}$ in $\delta^{\vec{\mathfrak{a}}}_{H^\star}$ and $\vec{Y}$ in $\delta^{\mathfrak{c}}_{H^\star}$. Clearly, $X := A(v, \vec{c}_y)[\vec{B}_Y]$ is in $\delta^{\mathfrak{c}}_{H^\star}$. We have, for x in $\delta^{\mathfrak{b}}_{H^\star}$:

$$\begin{aligned}(Xx)^{H^\star} &\leftrightarrow \mathcal{H}(A(c_x, \vec{c}_y)[\vec{B}_Y]) \\ &\leftrightarrow (A(x, \vec{y}, \vec{Y}))^{H^\star}.\end{aligned}$$

We may conclude that $H^\star : \Omega_{U^{\mathsf{sch}}} \rhd U^{\mathsf{pc}}$.

To verify the proof in $I\Delta_0 + \Omega_1$, you have to provide p-time bounds for the $\Omega_{U^{\mathsf{sch}}}$-proofs produced in the induction leading to Eq. (2). □

It is easy to see that we may extend Lemma 5.1 to $H^\star$, where we now restrict also the new domain $\delta^{\mathfrak{c}}_{H^\star}$ to J. Thus, we have:

Lemma 5.2. *Let J be a $\Omega_{U^{\mathsf{sch}}}$-cut. We have that $H^\star_{U^{\mathsf{sch}}} \restriction J$ is elementarily equivalent to $H^\star_{U^{\mathsf{sch}}}$ in $\Omega_{U^{\mathsf{sch}}}$. This result is verifiable in $I\Delta_0 + \Omega_1$.*

6. From consistency to principle V

In this section, we study how to extract Frege mappings from the Henkin–Feferman construction.

6.1. The collapse

The development in this subsection can be executed in $I\Delta_0 + \Omega_1$. We give a collapsing lemma.

Lemma 6.1. *Let U be an arithmetical theory. Let Av and Evw be arithmetical formulas such that, U-provably, E defines an equivalence relation on the virtual class $\{v \mid Av\}$. Then, there is a U-cut J, a set of U-numbers $I_0 \subseteq J$, and a formula F such that I_0 is provably downward closed and such that F, U-provably, defines an injective function from $(A \cap J)/E$ to I_0. Moreover, any shortening of J has the same property.*

Proof. Reason in U. We assume that we have, apart from the natural numbers, one extra element $*$ that stands for 'undefined'. (We can implement this by letting $x + 1$ represent x and letting 0 represent $*$.) We say that a sequence σ of length ℓ is *acceptable* iff it satisfies the following condition. For all $i < \ell$, $(i \notin A$ and $\sigma_i = *)$ or $(i \in A$ and $\forall j{<}i\; \neg jEi$ and σ_i is the smallest number not in the range of $\sigma \restriction i)$ or $(i \in A$ and $\exists j{<}i\; (jEi \wedge \sigma_i = \sigma_j))$. We define J_0 as the set of all ℓ such that:

1. there exists an acceptable sequence of length ℓ;
2. there exists at most one acceptable sequence of length ℓ;
3. for any acceptable sequence of length ℓ all initial subsequences are also acceptable;
4. any acceptable sequence of length ℓ codes a bijection between the classes $(A \restriction \ell)/E$ and m, where m is the supremum of

$$\{x + 1 \mid x \neq * \wedge \exists j{<}\ell\; \sigma_j = x\};$$

(So, if the numerical range of σ is non-empty, m will be the maximum of that range plus one, otherwise $m = 0$.)

5. if σ is acceptable of length ℓ, then, for any i, $\sigma_i = *$ or $\sigma_i \leq i$.

We can easily check that J_0 is closed under successor. Let J be a shortening of J_0 to a cut. We set $Fi = x$ for: $i \in (A \cap J)$ and there is an acceptable σ of length $i + 1$ with $\sigma_i = x$. It is easily seen that F is an injection from $A \cap J$ to J with downwards closed range.

Clearly, the proof also works for any shortening of J. □

Corollary 6.1. *Consider any theory V. There is a $K : \Omega_V \rhd V$, such that each of the domains $\delta_K^{\mathfrak{a}}$ of K is a Ω_V-provably downwards closed class $I_0^{\mathfrak{a}}$ and such that if there is an identity relation for sort $\mathfrak{a}$, then $E_K^{\mathfrak{a}}$ is the identity relation $=$ of Ω_V restricted to $I_0^{\mathfrak{a}}$.*

Proof. To simplify the formulation, we will take, for any interpretation M in V, $E_M^{\mathfrak{a}}$ to be $=$, if $\mathfrak{a}$ does not have an identity relation.

Consider the Henkin interpretation $H : \Omega_V \rhd V$. We apply Lemma 6.1 to each of the $\delta_H^{\mathfrak{a}}$ and $E_H^{\mathfrak{a}}$. Thus we obtain, for each $\mathfrak{a}$, $J^{\mathfrak{a}}$ and $F^{\mathfrak{a}}$ with the promised properties. We take J the intersection of the $J^{\mathfrak{a}}$. Consider $K_0 := H \upharpoonright J$. Then, by Lemma 5.1, $K_0 : \Omega_V \rhd V$. Note that, by the same lemma, $E_{K_0}^{\mathfrak{a}}$ will be $E_H^{\mathfrak{a}}$ restricted to J. It follows that each $F^{\mathfrak{a}}$ collapses $\delta_{K_0}^{\mathfrak{a}}/E_{K_0}^{\mathfrak{a}}$ to a downwards closed $I_0^{\mathfrak{a}} \subseteq J$. Using the $F^{\mathfrak{a}}$, we find the desired K in the obvious way. □

6.2. Implementing V

We consider the following theory 2-$\mathsf{SUCC_p}$, the theory of two partial successors. The language of 2-$\mathsf{SUCC_p}$ consists of one unary predicate symbol Z and three binary predicate symbols $=$, $\mathsf{S_a}$ and $\mathsf{S_b}$. The theory is axiomatized as follows.

- The axioms of identity.
- $\vdash \exists! x\ \mathsf{Z}x$.
- $\vdash (\mathsf{S_a}xz \wedge \mathsf{S_a}yz) \to x = y$.
- $\vdash (\mathsf{S_a}xy \wedge \mathsf{S_a}xz) \to y = z$.
- $\vdash \mathsf{S_a}xy \to \neg\, \mathsf{Z}y$.
- $\vdash (\mathsf{S_b}xz \wedge \mathsf{S_b}yz) \to x = y$.
- $\vdash (\mathsf{S_b}xy \wedge \mathsf{S_b}xz) \to y = z$.
- $\vdash \mathsf{S_b}xy \to \neg\, \mathsf{Z}y$.
- $\vdash \neg\, (\mathsf{S_a}xz \wedge \mathsf{S_b}yz)$.

It is more pleasant to formulate 2-$\mathsf{SUCC_p}$ in an unofficial language of partial functions using the constant 0 for Z and treating $\mathsf{S_a}$ and $\mathsf{S_b}$ as partial functions using $=$ for: *both sides are defined and equal*, $\neq$ for: *both sides are defined and unequal*, and $\downarrow$ for *is defined*. In this language our axioms become:

- $\vdash 0 \downarrow$.
- $\vdash \mathsf{S_a}x = \mathsf{S_a}y \to x = y$.
- $\vdash \neg\, \mathsf{S_a}x = 0$.
- $\vdash \mathsf{S_b}x = \mathsf{S_b}y \to x = y$.
- $\vdash \neg\, \mathsf{S_b}x = 0$.
- $\vdash \neg\, \mathsf{S_a}x = \mathsf{S_b}y$.

We define the following mapping from numbers to formulas.

- $\mathsf{str}_0(u) := \mathsf{Z}u$,
- $\mathsf{str}_{2n+1}(u) := \exists v\ (\mathsf{str}_n(v) \wedge \mathsf{S_a}vu)$,
- $\mathsf{str}_{2n+2}(u) := \exists v\ (\mathsf{str}_n(v) \wedge \mathsf{S_b}vu)$.

When we take some care to recycle our variables in the recursion in an efficient way, the mapping $n \mapsto \mathsf{str}_n(u)$ will be p-time. Let 2-$\mathsf{SUCC_p^{\infty}}$ be 2-$\mathsf{SUCC_p}$ plus all axioms $\vdash \exists u\ \mathsf{str}_n(u)$. In our unofficial language of partial terms we could view the $\mathsf{str}_n(u)$ as defining dyadic numerals, say $\widetilde{n}$. In this notation, 2-$\mathsf{SUCC_p^{\infty}}$ is 2-$\mathsf{SUCC_p}$ plus the axioms $\vdash \widetilde{n} \downarrow$.

Now consider any theory U. Suppose $I\Delta_0 + \Omega_1 \vdash M : U \rhd$ 2-$\mathsf{SUCC_p^{\infty}}$. Note that, by a result due to Wilkie & Paris (see [22]) we have, in $I\Delta_0 + \Omega_1$, a p-time bound on the U-proofs of the M-translations of the axioms of 2-$\mathsf{SUCC_p^{\infty}}$. (As is well known, the mapping $A \mapsto A^M$ is p-time.) We have:

$$I\Delta_0 + \Omega_1 \vdash \forall x\ \Box_U (\exists u\ \mathsf{str}_x(u))^M. \tag{3}$$

We claim:

$$I\Delta_0 + \Omega_1 \vdash \forall x\ \Box_U (\forall u, v\ ((\mathsf{str}_x(u) \wedge \mathsf{str}_x(v)) \to u = v))^M. \tag{4}$$

The verification is by Σ_1^b-PIND-Induction on x, which we have available in $I\Delta_0 + \Omega_1$.[10] In the verification, we have to take some care to estimate the U-proofs by a p-time bound in x. Using Eq. (4), we easily show:

$$I\Delta_0 + \Omega_1 \vdash \forall x\ \Box_U (\forall u, v\ ((\mathsf{str}_x(u) \wedge \mathsf{str}_{2x+1}(v)) \to \mathsf{S_a}(u, v)))^M \tag{5}$$

and, similarly:

$$I\Delta_0 + \Omega_1 \vdash \forall x\ \Box_U (\forall u, v\ ((\mathsf{str}_x(u) \wedge \mathsf{str}_{2x+2}(v)) \to \mathsf{S_b}(u, v)))^M. \tag{6}$$

The U-proofs provided in Eqs. (5) and (6) can be given explicit p-time bounds.

10 In fact, we could have used S_2^1 as our base theory in stead of $I\Delta_0 + \Omega_1$.

In $I\Delta_0 + \Omega_1$ (in fact already in S^1_2), we can develop the theory of binary strings or dyadic numerals, defining e.g. concatenation by $x * y := x \cdot \ell(y) + y$, where $\ell(y)$ is the largest power of 2, which is $\leq y + 1$. We use this to prove the following equation.

$$I\Delta_0 + \Omega_1 \vdash \forall x, y\ (x \neq y \to \Box_U (\forall u \neg (\mathsf{str}_x(u) \wedge \mathsf{str}_y(u)))^M). \tag{7}$$

The proof runs as follows. Reason in $I\Delta_0 + \Omega_1$. Consider x and y with $x \neq y$. We view x and y as binary strings. Either x and y are comparable in the final substring ordering or they are not. Suppose first that they are comparable, say $y = x' * x$, where $x' \neq 0$. As is easily seen: $\Box_U(\forall u\ \neg(\mathsf{str}_{x'}(u) \wedge \mathsf{str}_0(u)))^M$. We can bound the U-proof by a p-time function in x'. We write $[x]_i$ for the initial string of x of length i. We now prove by Σ^b_1-LIND-Induction on i, that, for all $i < \mathsf{length}(x)$, $\Box_U(\forall u \neg(\mathsf{str}_{x' * [x]_i}(u) \wedge \mathsf{str}_{[x]_i}(u)))^M$. We have to keep track of a polynomial bound for the U-proofs. Clearly, the consequent of the implication of Eq. (7) is a direct consequence.

Now suppose that x and y are not comparable in the final substring ordering. Let x' be the largest final string they have in common. We have $x = x'' * x'$ and $y = y'' * x'$, where either $x'' = 2x''' + 1$ and $y'' = 2x''' + 2$, or $x'' = 2x''' + 2$ and $y'' = 2y''' + 1$. It is easily seen that:

$$\Box_U(\forall u \neg(\mathsf{str}_{x''}(u) \wedge \mathsf{str}_{y''}(u)))^M.$$

We now prove by Σ^b_1-LIND-Induction on i, that:

$$\forall i{<}\mathsf{length}(x')\ \Box_U(\forall u \neg(\mathsf{str}_{x'' * [x']_i}(u) \wedge \mathsf{str}_{y'' * [x']_i}(u)))^M.$$

As usual we have to keep track of a polynomial bound for the U-proofs. Clearly, the consequent of the implication in Eq. (7) is a direct consequence.

In Ω_U, we have a Henkin interpretation H of U with truth predicate $\mathcal{H}$ and characteristic cut $\mathfrak{I}$. We map $\mathfrak{I}$ into δ_H via $x \mapsto \nu_x := c[(\exists u\ \mathsf{str}_x(u))^M]$. By Theorem 5.2(2) in combination with Eq. (3), we have:

$$\Omega_U \vdash \forall x{\in}\mathfrak{I}\ \mathcal{H}((\mathsf{str}_x(\nu_x))^M). \tag{8}$$

Similarly, by Eq. (7), we have:

$$\Omega_U \vdash \forall x, y : \mathfrak{I}\ (x \neq y \to \mathcal{H}((\nu_x \neq \nu_y)^M)). \tag{9}$$

So it follows that, in Ω_U, we have an injective mapping ν from $\mathfrak{I}$ into $\delta_{H \circ M}$. A fortiori, ν is injective when considered as a mapping from $\mathfrak{I}$ to $\delta^{\mathfrak{a}}_H$, where $\mathfrak{a}$ is $\sigma_M(\mathfrak{o})$ for the single sort $\mathfrak{o}$ of 2-$\mathsf{SUCC}^\infty_\mathsf{p}$. Finally note that our argument also goes through for $H \restriction J$ and $\mathfrak{I} \cap J$, for any Ω_U-cut J. We have proved:

Theorem 6.2. *Suppose, for some M, $I\Delta_0 + \Omega_1 \vdash M : U \rhd$ 2-$\mathsf{SUCC}^\infty_\mathsf{p}$. Then, Ω_U provides an injection from $\mathfrak{I}_U$ into $\delta^{\sigma_M(\mathfrak{o})}_{H_U}$. This fact also holds if, for any Ω_U-cut J, we consider $\mathfrak{I}_U \cap J$ and $H_U \restriction J$.*

Let U be an s-theory with designated sort $\mathfrak{b}$ with identity. We set U^{pcf} for the theory we obtain by extending the language of U^{pc} with a new function symbol ‡ from $\mathfrak{c}$ to $\mathfrak{b}$ and adding the axiom V:

$$\vdash \ddagger X = \ddagger Y \leftrightarrow \forall z^{\mathfrak{b}}\ (Xz \leftrightarrow Yz).$$

Theorem 6.3. *Let U be an s-theory with designated sort $\mathfrak{b}$. Suppose, for some M, $I\Delta_0 + \Omega_1 \vdash M : U^{\mathsf{sch}} \rhd$ 2-$\mathsf{SUCC}^\infty_\mathsf{p}$. Then, we have: $\mathsf{Q} + \mathsf{con}(U^{\mathsf{sch}}) \rhd U^{\mathsf{pcf}}$. This result is verifiable in $I\Delta_0 + \Omega_1$.*

Proof. It is sufficient to show that $\Omega_{U^{\mathsf{sch}}} \rhd U^{\mathsf{pcf}}$. We work in $\Omega_{U^{\mathsf{sch}}}$. Consider $H^\star$ and $\mathfrak{I}$. Let $E^{\mathfrak{c}}$ be the equivalence relation on $\delta^{\mathfrak{c}}_{H^\star}$ given by:

$$XE^{\mathfrak{c}}Y :\leftrightarrow (\forall x^{\mathfrak{b}}(Xx \leftrightarrow Yx))^{H^\star}.$$

By Lemma 6.1, we can find a cut $J \subseteq \mathfrak{I}$ such that we have an F, that defines an injective function from $\delta^{\mathfrak{c}}_{H^\star} \cap J$ modulo $E^{\mathfrak{c}}$ into a downwards closed segment of J. We consider $H^\star \restriction J$. Note that all domains of this interpretation will be coded in J.

By Theorem 6.2, we can find an injection ν from J into $\delta^{\mathfrak{b}}_{H \restriction J}$ modulo $E^{\mathfrak{b}}_H$. So, $G := \nu \circ F$ is a map from $(\delta^{\mathfrak{c}}_{H^\star \restriction J}$ modulo $E^{\mathfrak{c}})$ to $(\delta^{\mathfrak{b}}_{H^\star \restriction J}$ modulo $E^{\mathfrak{b}}_H)$. We use G to interpret the function ‡.

Verifiability in $I\Delta_0 + \Omega_1$ is obvious. □

6.3. *Proving the consistency of* 2-$\mathsf{SUCC}_{\mathsf{p}}^{\infty}$

In this subsection, we verify that Robinson's Arithmetic proves the consistency of 2-$\mathsf{SUCC}_{\mathsf{p}}^{\infty}$ on a definable cut.

Theorem 6.4. *We have:* $\mathsf{Q} \rhd (\mathsf{Q} + \mathsf{con}(2\text{-}\mathsf{SUCC}_{\mathsf{p}}^{\infty}))$.

Proof. Since $\mathsf{Q} \rhd I\Delta_0 + \Omega_2$, it is sufficient to prove the desired consistency statement on a cut in $I\Delta_0 + \Omega_2$. Thus, we work in $I\Delta_0 + \Omega_2$.

An a-assignment σ for a formula A codes a function from the free variables of *A* to the number *a*.[11] *An evaluation tree Θ* for a, p, σ and *A*, where σ is an *a*-assignment for *A*, and where *p* is a polarity $+$ or $-$, is given as follows. Its nodes are sequences of numbers below *a*. The nodes are labeled by triples of assignments, polarities and formulas. The root is the empty sequence, labeled with $\langle \sigma, p, A\rangle$. We give the sample clauses for conjunction, universal quantification and negation.

- Suppose we have a node τ labeled with $\langle \nu, +, (B \wedge C)\rangle$. Then its successor nodes are $\tau 0$ and $\tau 1$, where $\tau 0$ is labeled with $\langle \nu \restriction \mathsf{FV}(B), +, B\rangle$ and $\tau 1$ is labeled with $\langle \nu \restriction \mathsf{FV}(C), +, C\rangle$.
- Suppose we have a node τ labeled with $\langle \nu, -, (B \wedge C)\rangle$. Then its successor node is $\tau 0$, where $\tau 0$ is labeled with either $\langle \nu \restriction \mathsf{FV}(B), +, B\rangle$ or $\langle \nu \restriction \mathsf{FV}(C), +, C\rangle$.
- Suppose we have a node τ labeled with $\langle \nu, +, \forall v\, B\rangle$. In case v is free in *B*, its successor nodes are τx for all $x < a$. where τx is labeled with $\langle \nu[v := x], +, B\rangle$. In case v is not free in *B*, its successor node is $\tau 0$ labeled with $\langle \nu, +, B\rangle$.
- Suppose we have a node τ labeled with $\langle \nu, -, \forall v\, B\rangle$. In case v is free in *B*, its unique successor node is τx for some $x < a$. where τx is labeled with $\langle \nu[v := x], -, B\rangle$. If v is not free in *B*, its successor node is $\tau 0$ with label $\langle \nu, -, B\rangle$.
- Suppose we have a node τ labeled with $\langle \nu, +, \neg B\rangle$. Then its successor node are $\tau 0$, labeled with $\langle \nu, -, B\rangle$.
- Suppose we have a node τ labeled with $\langle \nu, -, \neg B\rangle$. Then its successor node are $\tau 0$, labeled with $\langle \nu, +, B\rangle$.

A tree is *successful* iff, for all leaves τ, we have:

- if the label of τ is $\langle \nu, +, v = w\rangle$, then $\nu(v) = \nu(w)$;
- if the label of τ is $\langle \nu, -, v = w\rangle$, then $\nu(v) \neq \nu(w)$;
- if the label of τ is $\langle \nu, +, \mathsf{S_a} vw\rangle$, then $2 \cdot \nu(v) + 1 = \nu(w)$;
- If the label of τ is $\langle \nu, -, \mathsf{S_a} vw\rangle$, then $2 \cdot \nu(v) + 1 \neq \nu(w)$;
- if the label of τ is $\langle \nu, +, \mathsf{S_b} vw\rangle$, then $2 \cdot \nu(v) + 2 = \nu(w)$;
- If the label of τ is $\langle \nu, -, \mathsf{S_b} vw\rangle$, then $2 \cdot \nu(v) + 2 \neq \nu(w)$.

We can make an estimate of the size of any evaluation tree for *a* and *A*. We will write $|b|$ for the entier of the 2-logarithm of $b + 1$. We write $x\#y := 2^{|x|\cdot|y|}$, $x\#_2 y := 2^{2^{||x||\cdot||y||}}$.

First, consider a node α. Clearly, $|\alpha|$ is bounded by a term of order $c_0\cdot(|a|+c_1)\cdot|A|$, for standard c_0 and c_1. So, α is estimated by $k_0\#a\#A$, for sufficiently large *a*. Here k_0 is fixed and standard. Secondly, the number of nodes is estimated by $k_1\#a\#A$. Finally, the size of a label is estimated by $k_2\#a\#A$. Thus, for a whole tree Θ, we will have: $|\Theta| \leq (k_1\#a\#A)\cdot(c_2 \cdot |a| \cdot |A|) \leq k_3\#a\#A$, for standard k_3 and sufficiently large *a*. Thus, Θ is estimated by $F(a, A) := 2^{k_3\#a\#A} = k_4\#_2 2^a\#_2 2^A$. We will only consider *a* and *A* in the logarithmic cut $\mathsf{log} := \{x \mid 2^x \downarrow\}$, so that *F* will always be defined. Note that, since we are working in $I\Delta_0 + \Omega_2$, the cut log will interpret $I\Delta_0 + \Omega_1$.

There will be some p-time function *G* transforming bounds for trees of *B* to bounds for trees of e.g. $\forall v\, B$. We will assume that we arranged it so that provably $G(a, F(a, B)) \leq F(a, \forall v\, B)$ (and similarly for the other connectives).

We define $a, \sigma \models^p A$ iff there is a successful tree Θ below $F(a, A)$ for a, σ, p, A. We claim that $\models$ satisfies the obvious commutation clauses for *a*, *A* in log. Suppose e.g. that v is free in *B* and, for all $x < a$, we have $a, \sigma[v := x] \models^+ B$. So for all $x < a$, we have a successful tree Θ_x for $a, \sigma[v := x], +, B$. The Θ_x are bounded by $F(a, B)$. By Δ_0-induction we can find, for each such *x*, a smallest such Θ_x, so we can treat the Θ_x as functionally dependent on *x*. We now take as the set of nodes of the new tree all nodes of the form $x\tau$, where $x < a$ and τ is a node of Θ_x. The label of $x\tau$ will be the label of τ in Θ_x. We easily verify that the new tree is successful. We may conclude that $a, \sigma \models^+ \forall v\, B$.

We may verify the validity of predicate logic, on *a* in log, for the language of 2-$\mathsf{SUCC_p}$, where the language is restricted to log. E.g., we may prove, by induction on *B* in log, that, for all $\sigma : \mathsf{FV}(B) \to a$, we have $a, \sigma \models^+ B$ or $a, \sigma \models^- B$. Similarly we may verify the validity of 2-$\mathsf{SUCC_p}$ on *a*.

We claim that $I\Delta_0 + \Omega_2 \vdash \mathsf{con}^{\mathsf{log}}(2\text{-}\mathsf{SUCC}_{\mathsf{p}}^{\infty})$. Reason in $I\Delta_0 + \Omega_2$. Consider a SUCC_p^{∞}-proof π in log. Consider the largest axiom of the form $\exists v\ \mathsf{str}_b(v)$ occurring in π. Clearly, *b* will be in log. Take $a := b + 1$. We now show that all axioms of 2-$\mathsf{SUCC_p}$ are true in *a*. Moreover, we show that for each $c < a$, $\exists v\ \mathsf{str}_c(v)$ is true in *a*. Noting that all formulas occurring in π are in log, we may now show by induction on subproofs that all subconclusions of π are true. (Details will depend on the proof system.) Hence, π cannot be a proof of falsity. □

11 We follow the usual practice to let *a* stand for $\{x \mid x < a\}$.

7. From comprehension to consistency

The proof in this section is a variant of the proofs of the consistency on a definable cut of ZF in GB and of the consistency on a definable cut of PA in ACA_0. The original idea for the proof goes back to Mostowski. We show:

Theorem 7.1. *Suppose U is a finitely axiomatized s-theory. Suppose further that U^{sch} is sequential. Then, $U^{\mathsf{pc}} \rhd (\mathsf{Q} + \mathsf{con}(U^{\mathsf{sch}}))$.*

Proof. Let ν be a measure of complexity that counts *depth of logical constants*. Let $\mathcal{L}_1(x)$ be the set of U-formulas that do not contain concept variables and that are of ν-complexity $\leq x$.

We work in U^{pc}. We let σ range over assignments coded as finite sequences of pairs of variables and objects, satisfying the functionality condition. If a variable does not occur as a first component in the sequence we take its value to be the default 0. We assume syntax to be coded in the natural numbers provided by the fact that U^{sch} is sequential. Suppose U has k predicate symbols. We define (permitting ourselves some abuses of language):

$$\begin{aligned}\mathcal{T}_x(X) \ :\leftrightarrow\ & X \subseteq (\mathsf{ass} \times \mathcal{L}_1(x)) \wedge \forall\sigma\ \forall A{\in}\mathcal{L}_1(x)\ [\\ & ((A = P_0\vec{v}) \to (X\langle\sigma, A\rangle \leftrightarrow P_0(\sigma\vec{v}))) \wedge \\ & \ldots \\ & ((A = P_{k-1}\vec{v}) \to (X\langle\sigma, A\rangle \leftrightarrow P_{k-1}(\sigma\vec{v}))) \wedge \\ & ((A = \neg B) \to (X\langle\sigma, A\rangle \leftrightarrow \neg(X\langle\sigma, B\rangle))) \wedge \\ & ((A = (B \wedge C)) \to (X\langle\sigma, A\rangle \leftrightarrow (X\langle\sigma, B\rangle \wedge X\langle\sigma, C\rangle))) \wedge \\ & ((A = \exists v_i\, B) \to (X\langle\sigma, A\rangle \leftrightarrow \exists y\, X\langle\sigma[v_i := y], B\rangle))\ ].\end{aligned}$$

Let $\mathcal{J}_0$ be the class of all x such that $\exists! X\ \mathcal{T}(X, x)$. We show that $\mathcal{J}_0$ contains 0 and is closed under successor.

For $x = 0$, we take X the concept of all $\langle\sigma, A\rangle$ such that, for some $\vec{v}$, $((A = P_0\vec{v})$ and $P_0(\sigma\vec{v}))$ or … $((A = P_{k-1}\vec{v})$ and $P_{k-1}(\sigma\vec{v}))$. This exists by predicative comprehension. Uniqueness is immediate. Let X_x be the unique concept such that $\mathcal{T}(X_x, x)$. We define X_{x+1} to be the union of X_x with the $\langle\sigma, A\rangle$ such that $\nu(A) = x + 1$ and $((A = \neg B)$ and $\neg X_x\langle\sigma, B\rangle)$ or $((A = (B \wedge C))$ and $X_x\langle\sigma, B\rangle$ and $X_x\langle\sigma, C\rangle)$ or $((A = \exists v_i\, B)$ and $\exists y\ X_x\langle\sigma[v_i := y], B\rangle)$. It is easily seen that $\mathcal{T}(X_{x+1}, x + 1)$. Moreover, if we had $\mathcal{T}(Y, x + 1)$, then we would have $\mathcal{T}(Y \cap \mathcal{L}_1(x), x)$. So, $Y \cap \mathcal{L}_1(x) = X_x$. It now easily follows that $Y = X_{x+1}$.

Let $\mathcal{J}$ be a definable cut of $\mathcal{J}_0$. We will employ the functional notation X_x for x in $\mathcal{J}$. We write $\mathcal{L}_1(\mathcal{J})$ for the class of all formulas in one of the $\mathcal{L}_1(x)$ for $x \in \mathcal{J}$.

Define $\mathsf{Sat}(\sigma, A) :\leftrightarrow \exists x{\in}\mathcal{J}\ X_x\langle\sigma, A\rangle$. It is easily seen that Sat is a satisfaction predicate for $\mathcal{L}_1(\mathcal{J})$ in the strong sense that we can verify the commutation conditions for $\mathcal{L}_1(\mathcal{J})$.

Consider any A in $\mathcal{L}_1(\mathcal{J})$, any finite assignment σ and any variable v_i. We claim that:

$$\exists Z\ \forall z\ (Zz \leftrightarrow \mathsf{Sat}(\sigma[v_i := z], A)). \tag{10}$$

Let $\nu(A) =: a$. We can take $Z := \wr z \mid X_a\langle\sigma[v_i := z], A\rangle \int$. We easily verify that Z fulfills the desiderata.

Let $\mathcal{P}$ be the set of all U^{sch}-proofs p, such that all formulas occurring in p are in $\mathcal{L}_1(\mathcal{J})$. Let $\mathcal{I}$ be the set of numbers x such that all $p \in \mathcal{P}$ of length $\leq x$ (length in the sense of *number of steps*) have a conclusion A_p such that, for all σ, $\mathsf{Sat}(\sigma, A_p)$. We show that $0 \in \mathcal{I}$ and that $\mathcal{I}$ is closed under successor.

We first treat the case of 0. Consider any axiom $C\vec{Y}$ of U. Since we are working in U^{pc}, we have $\forall\vec{Y}\ C(\vec{Y})$. Consider $\vec{A}$ in $\mathcal{L}_1(\mathcal{J})$. By Eq. (10), we can find $Z_i := \wr z \mid \mathsf{Sat}(\sigma[v_0 := z], A_i)\int$. We may conclude $C\vec{Z}$. By *external* induction on standard formulas D, we find: $D\vec{Z} \leftrightarrow \mathsf{Sat}(\sigma, D[\vec{A}])$. Thus we may conclude:

$$\forall\vec{A}{\in}\mathcal{L}_1(\mathcal{J})\ \forall\sigma\ \mathsf{Sat}(\sigma, C[\vec{A}]).$$

The case of the induction step is dependent on the proof system. The reader is invited to check that this works for her favorite proof system.

Thus, there can be no p in $\mathcal{P}$ with length in $\mathcal{I}$ with conclusion $\perp$. By Solovay's method of shortening initial segments, we can find a cut $\mathcal{K}$ that is verifiably part of $\mathcal{I}$ and $\mathcal{J}$. Since we use a standard coding, any U^{sch}-proof in $\mathcal{K}$ will only contain formulas D with $\nu(D)$ in $\mathcal{J}$ and will have length in $\mathcal{I}$. Thus, we certainly have $\mathsf{con}^{\mathcal{K}}(U^{\mathsf{sch}})$.[12] □

In the paper, we will only apply this result to ordinary theories, not s-theories. Theorem 9.5 illustrates that the result does not work for theories axiomatized by arbitrary RE axiom sets.

Since $\mathsf{Q} + \mathsf{con}(U^{\mathsf{sch}})$ is finitely axiomatized, the previous theorem is trivially verifiable in $I\Delta_0 + \Omega_1$. However, there is a catch. What is verifiable is of the form "*given a finitely axiomatized s-theory U, we can verify in $I\Delta_0 + \Omega_1$ that* …". So the

12 Note that the result involving $\mathcal{I}$ and $\mathcal{J}$ is stronger than the present one. I know of no application of the extra information.

bound on the axioms of U is external/standard. If, however we want to have the bound internally, we see that the size of the code of our satisfaction predicate is exponential in the number of predicates of the signature of the theory. Thus, to make the proof work in $I\Delta_0 + \Omega_1$, we have to assume that the number of predicates is a logarithmic number n, i.e. 2^n exists. Inspection shows that the size of the satisfaction predicate is the only obstacle. We have:

Theorem 7.2. *The theory $I\Delta_0 + \Omega_1$ proves the following. Suppose that U is a finitely axiomatized s-theory with logarithmic bound. Suppose further that U^{sch} is sequential. Then, we have: $U^{\mathrm{pc}} \rhd (\mathsf{Q} + \mathsf{con}(U^{\mathrm{sch}}))$.*

We end with two corollaries.

Corollary 7.3. *Consider any finitely axiomatized theory U. Suppose U is mutually interpretable with a finitely axiomatized sequential theory V. Then, we have $U^{\mathrm{pc}} \rhd (\mathsf{Q} + \mathsf{con}(U))$.*

Proof. By Theorem 3.2, we have (i) $U^{\mathrm{pc}} \rhd V^{\mathrm{pc}}$. By Theorem 7.1, we find that (ii) $V^{\mathrm{pc}} \rhd (\mathsf{Q}+\mathsf{con}(V))$. We have, by applying $\exists\Sigma_1^{\mathrm{b}}$-completeness to $V \rhd U$, that (†) $I\Delta_0 + \Omega_1 \vdash \mathsf{con}(V) \to \mathsf{con}(U)$. Moreover, $I\Delta_0 + \Omega_1$ is interpretable in Q on a cut. Hence $I\Delta_0 + \Omega_1 + \mathsf{con}(V)$ is interpretable in $\mathsf{Q} + \mathsf{con}(V)$. By (†), we find that $I\Delta_0 + \Omega_1 + \mathsf{con}(U)$ is interpretable in $\mathsf{Q} + \mathsf{con}(V)$. So, a fortiori, (iii) $(\mathsf{Q} + \mathsf{con}(V)) \rhd (\mathsf{Q} + \mathsf{con}(U))$. Combining (i), (ii) and (iii), we find $U^{\mathrm{pc}} \rhd (\mathsf{Q} + \mathsf{con}(U))$. □

By the reasoning leading to Theorem 7.2, we obtain:

Corollary 7.4. *The theory $I\Delta_0 + \Omega_1$ proves the following. Suppose U and V are finitely axiomatized theories with logarithmic bound. Suppose further that $U \equiv V$ and V is sequential. Then, $U \rhd (\mathsf{Q} + \mathsf{con}(U))$.*

8. Putting it all together

In this section, we assemble the building blocks to prove our main result. We define $(U)_\omega := U + \{\mathsf{con}^n(U) \mid n \in \omega\}$.

Theorem 8.1. *For all n, we have, that it is both true and verifiable in $I\Delta_0 + \Omega_1$, that $\mathsf{P}^{n+1}\mathsf{V} \equiv \mathsf{Q} + \mathsf{con}^n(\mathsf{Q})$. It follows that $\mathsf{P}^{\omega}\mathsf{V} \equiv_{\mathrm{loc}} (\mathsf{Q})_\omega$.*[13]

Proof. The proof is by (external) induction on n. Here is the outline of the base case.

$$\begin{aligned} \mathsf{P}^1\mathsf{V} &\rhd \mathsf{Q} && (11)\\ &\rhd \mathsf{Q} + \mathsf{con}(2\text{-}\mathsf{SUCC}^{\infty}_{\mathrm{p}}) && (12)\\ &\rhd (2\text{-}\mathsf{SUCC}^{\infty}_{\mathrm{p}})^{\mathrm{pcf}} && (13)\\ &\supseteq \mathsf{P}^1\mathsf{V}. && (14) \end{aligned}$$

The various steps are justified as follows.

step (11) This result is proved in [2], Section 2.2.
step (12) This step is our Theorem 6.4.
step (13) This step is provided by our Theorem 6.3.

Since each of the steps is verifiable in $I\Delta_0 + \Omega_1$, so is their composition. Next we prove the induction step. We suppose that it is both true and verifiable in $I\Delta_0 + \Omega_1$ that $\mathsf{P}^{n+1}\mathsf{V} \equiv (\mathsf{Q} + \mathsf{con}^n(\mathsf{Q}))$. Let W be finitely axiomatized a sequential theory extending Q, like S^1_2, that is interpretable in Q on a cut. We have:

$$\begin{aligned} \mathsf{P}^{n+2}\mathsf{V} &\supseteq (\mathsf{P}^{n+1}\mathsf{V})^{\mathrm{pc}} && (15)\\ &\rhd (\mathsf{Q} + \mathsf{con}^n(\mathsf{Q}))^{\mathrm{pc}} && (16)\\ &\rhd \mathsf{Q} + \mathsf{con}^{n+1}(\mathsf{Q}) && (17)\\ &\rhd \mathsf{Q} + \mathsf{con}(\mathsf{P}^{n+1}\mathsf{V}) && (18)\\ &\rhd (\mathsf{P}^{n+1}\mathsf{V})^{\mathrm{pcf}} && (19)\\ &\rhd \mathsf{P}^{n+2}\mathsf{V}. && (20) \end{aligned}$$

Here is the justification of the steps.

step (16) This step is justified by applying Theorem 3.2 to the left–right direction of the induction hypothesis.
step (17) This step is an application of Corollary 7.3, using the fact that the theory $\mathsf{Q} + \mathsf{con}^n(\mathsf{Q})$ is mutually interpretable with a sequential theory.

[13] Since Q_ω is reflexive, it follows that $\mathsf{Q}_\omega \rhd \mathsf{P}^{\omega}\mathsf{V}$.

step (18) We reason as follows. By the second conjunct of the induction hypothesis in the right-to-left direction, we have:

$$I\Delta_0 + \Omega_1 \vdash \ (\mathsf{Q} + \mathsf{con}^n(\mathsf{Q})) \rhd \mathsf{P}^{n+1}\mathsf{V}.$$

It follows that:

$$I\Delta_0 + \Omega_1 \vdash \mathsf{con}^{n+1}(\mathsf{Q}) \to \mathsf{con}(\mathsf{P}^{n+1}\mathsf{V}).$$

Since Q interprets $I\Delta_0 + \Omega_1$ on a definable cut, we may conclude that:

$$(\mathsf{Q} + \mathsf{con}^{n+1}(\mathsf{Q})) \rhd (\mathsf{Q} + \mathsf{con}(\mathsf{P}^{n+1}\mathsf{V})).$$

step (19) This is by Theorem 6.3.
step (20) This is by Theorem 4.1.

Since each of the steps is verifiable in $I\Delta_0 + \Omega_1$, the end result is. □

Remark 8.2. A curious observation about our proof is that the presence of the principles V only contributes to consistency strength in the case of $\mathsf{P}^1\mathsf{V}$. For all others, predicative comprehension already does all the work. Of course, consistency strength is but one aspect of a theory, so it could very well be that $\mathsf{P}^{n+2}\mathsf{V}$ has advantages of another kind over $\mathsf{PC}^{n+1}(\mathsf{P}^1\mathsf{V})$. It would be interesting to have a closer look at this matter.

Where is the induction of the proof of Theorem 8.1, verifiable? There is one obstacle. In the induction we iterate the p-time functions used in the steps. So it would seem that our theorem is verifiable in EA. However, we have not verified this in detail. In stead, we give an alternative proof in that uses Löb's Rule.

Theorem 8.3. *The theory $I\Delta_0 + \Omega_1$ proves that, for all logarithmic n:* $\mathsf{P}^{n+1}\mathsf{V} \equiv \mathsf{Q} + \mathsf{con}^n(\mathsf{Q})$.

Proof. We write $x : \mathsf{log}$ for $\exists y\ 2^x = y$. Let:

$$A_\star x := ((\mathsf{P}^{x+1}\mathsf{V} \equiv (\mathsf{Q} + \mathsf{con}^x(\mathsf{Q}))) \wedge (\mathsf{P}^{x+1}\mathsf{V} \equiv \mathsf{P}^{x+1}\mathsf{V}_{\mathsf{fa}})).$$

We show that $I\Delta_0 + \Omega_1 \vdash \forall x : \mathsf{log}\ A_\star x$. We will prove this using Löb's Rule. Thus, it is sufficient to show that:

$$I\Delta_0 + \Omega_1 \vdash \Box_{I\Delta_0+\Omega_1} \forall x : \mathsf{log}\ A_\star x \to \forall x : \mathsf{log}\ A_\star x.$$

We reason in $I\Delta_0 + \Omega_1$. Suppose

$$\Box_{I\Delta_0+\Omega_1} \forall x : \mathsf{log}\ A_\star(x). \tag{21}$$

Note that it follows that:

$$\Box_{I\Delta_0+\Omega_1} \forall x : \mathsf{log}\ (\mathsf{con}(\mathsf{P}^{x+1}\mathsf{V}) \leftrightarrow \mathsf{con}^x(\mathsf{Q})) \tag{22}$$

and:

$$\Box_{I\Delta_0+\Omega_1} \forall x : \mathsf{log}\ (\mathsf{con}(\mathsf{P}^{x+1}\mathsf{V}) \leftrightarrow \mathsf{con}(\mathsf{P}^{x+1}\mathsf{V}_{\mathsf{fa}})). \tag{23}$$

We proceed to prove $A_\star 1$.

$$\begin{aligned} \mathsf{P}^1\mathsf{V} &\supseteq \mathsf{P}^1\mathsf{V}_{\mathsf{fa}} && (24)\\ &\rhd \mathsf{Q} && (25)\\ &\rhd \mathsf{Q} + \mathsf{con}(2\text{-}\mathsf{SUCC}^\infty_\mathsf{p}) && (26)\\ &\rhd (2\text{-}\mathsf{SUCC}^\infty_\mathsf{p})^{\mathsf{pcf}} && (27)\\ &\supseteq \mathsf{P}^1\mathsf{V}. && (28) \end{aligned}$$

The only new step here is Step (24). It is justified since the interpretability of Q in $\mathsf{P}^1\mathsf{V}$ only asks for a standard proof. Hence, Q is also interpretable in $\mathsf{P}^1\mathsf{V}_{\mathsf{fa}}$ by a standard proof. This standard proof is of course available inside $I\Delta_0 + \Omega_1$, where we are at present working. Next we prove the case $A_\star(x+2)$, for $x : \mathsf{log}$. We set $\mathsf{FST}^{x+1}_{\mathsf{fa}}$, for the obvious flat counterpart of $\mathsf{P}^{x+1}\mathsf{V}_{\mathsf{fa}}$. We have:

$$\begin{aligned} \mathsf{P}^{x+2}\mathsf{V} &\supseteq (\mathsf{P}^{x+1}\mathsf{V})^{\mathsf{pc}} && (29)\\ &\supseteq (\mathsf{P}^{x+1}\mathsf{V}_{\mathsf{fa}})^{\mathsf{pc}} && (30)\\ &\rhd (\mathsf{FST}^{x+1}_{\mathsf{fa}})^{\mathsf{pc}} && (31)\\ &\rhd \mathsf{Q} + \mathsf{con}(\mathsf{FST}^{x+1}_{\mathsf{fa}}) && (32)\\ &\rhd \mathsf{Q} + \mathsf{con}(\mathsf{P}^{x+1}\mathsf{V}_{\mathsf{fa}}) && (33)\\ &\rhd \mathsf{Q} + \mathsf{con}^{x+1}(\mathsf{Q}) && (34)\\ &\rhd \mathsf{Q} + \mathsf{con}(\mathsf{P}^{x+1}\mathsf{V}) && (35)\\ &\rhd \mathsf{P}^{x+1}\mathsf{V}^{\mathsf{pcf}} && (36)\\ &\rhd \mathsf{P}^{x+2}\mathsf{V}. && (37) \end{aligned}$$

Here are the justifications of the steps.

step (30) This is immediate by Theorem 4.9.
step (31) This combines Theorem 3.2 with the definition of $\mathsf{FST}^{x+1}_{\mathsf{fa}}$.
step (32) This is by Corollary 7.4, using the fact that x is logarithmic.
step (33) Since $\Box_{I\Delta_0+\Omega_1}(\mathsf{FST}^{x+1}_{\mathsf{fa}} \rhd \mathsf{P}^{x+1}\mathsf{V}_{\mathsf{fa}})$, it follows that:

$$\Box_{I\Delta_0+\Omega_1}(\mathsf{con}(\mathsf{FST}^{x+1}_{\mathsf{fa}}) \to \mathsf{con}(\mathsf{P}^{x+1}\mathsf{V}_{\mathsf{fa}})).$$

Hence,

$$\begin{aligned}\mathsf{Q} + \mathsf{con}(\mathsf{FST}^{x+1}_{\mathsf{fa}}) &\rhd I\Delta_0 + \Omega_1 + \mathsf{con}(\mathsf{FST}^{x+1}_{\mathsf{fa}})\\ &\supseteq \mathsf{Q} + \mathsf{con}(\mathsf{P}^{x+1}\mathsf{V}_{\mathsf{fa}}).\end{aligned}$$

steps (34) and (35) These steps follow by combining Eqs. (22) and (23) with reasoning as in step (33).
step (36) This is by Theorem 6.3.
step (37) This step uses Theorem 4.1. □

Corollary 8.4. *The theory* EA *verifies that, for all n,* $\mathsf{P}^{n+1}\mathsf{V} \equiv_{\mathsf{loc}} (\mathsf{Q} + \mathsf{con}^n(\mathsf{Q}))$. *Hence,* EA *also verifies:* $\mathsf{P}^{\omega}\mathsf{V} \equiv \mathsf{Q}_{\omega}$.

9. Consequences

A first consequence is an alternative characterization of $\mathsf{P}^{\omega}\mathsf{V}$ in terms of EA.

Lemma 9.1. *Verifiably in* $I\Delta_0 + \Omega_1$, *for any* Π^0_1*-sentence P, we have:* $(\mathsf{Q} + \mathsf{con}(\mathsf{Q} + P)) \equiv (\mathsf{EA} + P)$.

Proof. This is an immediate consequence of Lemma 4.1 of [19]. Here is a quick sketch of the proof.

From right to left, we may prove, in $\mathsf{EA} + P$, the cut free consistency of $\mathsf{Q} + P$. See [22], for how this may be done. There is a definable EA-cut J, such that, in EA, for all x in J, $\mathsf{supexp}(x)$ is defined. On J, we will have, by cut-elimination, the ordinary consistency of $\mathsf{Q} + P$. Thus, J gives us the desired interpretation.

From left to right, we use that:

$$\mathsf{Q} + \mathsf{con}(\mathsf{Q} + P) \equiv I\Delta_0 + \Omega_1 + \mathsf{con}(I\Delta_0 + \Omega_1 + P).$$

So it is sufficient to prove our result with $I\Delta_0 + \Omega_1$ substituted for Q. We construct an interpretation for $\mathsf{EA} + P$ in $I\Delta_0 + \Omega_1 + \mathsf{con}(I\Delta_0 + \Omega_1 + P)$. We first construct the Henkin–Feferman interpretation:

$$H : (I\Delta_0 + \Omega_1 + \mathsf{con}(I\Delta_0 + \Omega_1 + P)) \rhd (I\Delta_0 + \Omega_1 + P).$$

By a result of Pudlák, there is a cut I such that $\forall x : I\ \Box_{I\Delta_0+\Omega_1+P}\mathsf{supexp}(\dot{x})\downarrow$. See [14]. We may take I shorter than the cut $\Im_H$ on which H is coded. Hence, by Theorem 5.2(2), $\forall x : I\ \mathscr{H}(\mathsf{supexp}(\dot{x})\downarrow)$. Now we restrict the domain of H to those y that such that, for some $x \in I$, $\mathscr{H}(c_y < \mathsf{supexp}(\dot{x}))$. It is easily seen that H restricted to this domain is in an interpretation of $\mathsf{EA} + P$. □

Lemma 9.2. *We have, verifiably in* $I\Delta_0 + \Omega_1$, *that, for all n:* $(\mathsf{Q} + \mathsf{con}^{2n+1}(\mathsf{Q})) \equiv (\mathsf{EA} + \mathsf{con}^n(\mathsf{EA}))$.

Proof. Let $A^\star x := ((\mathsf{Q} + \mathsf{con}^{2x+1}(\mathsf{Q})) \equiv (\mathsf{EA} + \mathsf{con}^x(\mathsf{EA})))$. By Löb's Rule, it is sufficient to prove: $I\Delta_0 + \Omega_1 \vdash \Box_{I\Delta_0+\Omega_1}\forall x\, A^\star x \to \forall x\, A^\star x$.

Reason in $I\Delta_0 + \Omega_1$. Suppose $\Box_{I\Delta_0+\Omega_1}\forall x\, A^\star x$. Note that our assumption implies that:

$$\Box_{I\Delta_0+\Omega_1}\forall x\,(\mathsf{con}^{2x+2}(\mathsf{Q}) \leftrightarrow \mathsf{con}^{x+1}(\mathsf{EA})). \tag{38}$$

We want to show $\forall x\, A^\star x$. The case $x = 0$ is immediate using Eq. (38). Suppose $x = y + 1$. By Lemma 9.1 and Eq. (38), we obtain:

$$\begin{aligned}\mathsf{Q} + \mathsf{con}^{2x+1}(\mathsf{Q}) &\equiv \mathsf{Q} + \mathsf{con}^{2y+3}(\mathsf{Q}) && (39)\\ &\equiv \mathsf{EA} + \mathsf{con}^{2y+2}(\mathsf{Q}) && (40)\\ &\equiv \mathsf{EA} + \mathsf{con}^{y+1}(\mathsf{EA}) && (41)\\ &\equiv \mathsf{EA} + \mathsf{con}^{x}(\mathsf{EA}). && (42)\end{aligned}$$

So, for all x, we have $A^\star x$. □

Corollary 9.1. *We have, verifiably in* EA, *that, for all n,* $\mathsf{P}^{2n+1}\mathsf{V} \equiv (\mathsf{EA} + \mathsf{con}^n(\mathsf{EA}))$ *and , thus,* $\mathsf{P}^{\omega}\mathsf{V} \equiv_{\mathsf{loc}} (\mathsf{EA})_{\omega}$.

Proof. We combine Lemma 9.2 with Corollary 8.4. □

The fact that $(\mathsf{Q} + \mathsf{con}(\mathsf{Q})) \equiv \mathsf{EA}$, suggests the following question.

Open Question 9.2. Is there a (natural) hierarchy of recursive functions F_n below superexponentiation, such that $I\Delta_0 +$ "F_n is total" is mutually interpretable with $\mathsf{Q} + \mathsf{con}^n(\mathsf{Q})$?

In his paper [1], Lev Beklemishev has shown that, verifiably in $\mathsf{EA}^+ := I\Delta_0 + \mathsf{SUPEXP}$, the following three theories have the same Π^0_1-consequences: EA_ω, $I\Delta_0 + \mathsf{SUPEXP}$ and $I\Delta_0 + I\Pi_1^-$ (see Propositions 4.5 and 11.5 of [1]).[14] Combining these results with Corollary 9.1, we obtain:

Corollary 9.3. *The theories* $\mathsf{P}^\omega\mathsf{V}$, EA^+ *and* $I\Delta_0 + I\Pi_1^-$ *are equiconsistent over* EA^+.

We close this paper with a treatment of strong reflexivity. It will follow that $\mathsf{P}^\omega\mathsf{V}$ is not finitely axiomatizable in a strong sense. From this we get that $\mathsf{P}^\omega\mathsf{V}$ does not interpret EA^+.

A theory T is *strongly reflexive* if it interprets:

$$\mho_T^+ := \mathsf{S}^1_2 + \{\mathsf{con}(T \restriction n) \mid n \in \omega\}.$$ [15, 16, 17]

Here $T \restriction n$ is the result of restricting T to its axioms with Gödel numbers below n.[18] As is easily seen $(\mathsf{Q})_\omega$ is strongly reflexive. Other examples of strongly reflexive theories are PA and ZFC. A theory T is strongly ℓ-reflexive if it locally interprets $\mho_T^+$.[19] We have:

Theorem 9.4. *The following are equivalent:*

i. *T is strongly ℓ-reflexive,*
ii. *$T \equiv_{\mathsf{loc}} U$, for some strongly reflexive U,*
iii. *$T \equiv_{\mathsf{loc}} U$, for some strongly ℓ-reflexive U.*

Proof. (i) ⇒ (ii) Suppose T is strongly ℓ-reflexive, i.e. $T \rhd_{\mathsf{loc}} \mho_T^+$. We have $\mho_T^+ \rhd T$, by an argument due to Feferman. See [4]. So, $T \equiv_{\mathsf{loc}} \mho_T^+$. We show that $\mho_T^+$ is strongly reflexive. Consider any n. We have, for some m, that:

$$(T \restriction m) \rhd (\mathsf{S}^1_2 + \mathsf{con}(T \restriction n)).$$

It follows that $\mathsf{S}^1_2 \vdash \mathsf{con}(T \restriction m) \to \mathsf{con}(\mathsf{S}^1_2 + \mathsf{con}(T \restriction n))$. Hence $\mho_T^+$ proves the consistency of each of its finitely axiomatized subtheories.

(ii) ⇒ (i) Suppose $T \equiv_{\mathsf{loc}} U$, where U is strongly reflexive. Consider any n. For some m, we have $(U \restriction m) \rhd (T \restriction n)$. So,

$$\mathsf{S}^1_2 \vdash \mathsf{con}(U \restriction m) \to \mathsf{con}(T \restriction n).$$

It follows that: $T \rhd U \rhd (\mathsf{S}^1_2 + \mathsf{con}(U \restriction m)) \supseteq (\mathsf{S}^1_2 + \mathsf{con}(T \restriction n))$.

The step (ii) ⇒ (iii) is trivial. We treat (iii) ⇒ (ii). Suppose $T \equiv_{\mathsf{loc}} U$, for some strongly ℓ-reflexive U. By the equivalence of (i) and (ii), $U \equiv_{\mathsf{loc}} V$, for some strongly reflexive V. It follows that $T \equiv_{\mathsf{loc}} V$, for some strongly reflexive V. □

We may conclude that $\mathsf{P}^\omega\mathsf{V}$ is strongly ℓ-reflexive. Since, as is easy to verify, every strongly reflexive theory locally interprets $(\mathsf{Q})_\omega$, we find that $\mathsf{P}^\omega\mathsf{V}$ is minimal among strongly ℓ-reflexive theories in the local interpretability preorder.

The following theorem shows that the functors PC and PCF yield coding free definitions of strong ℓ-reflexivity.

Theorem 9.5. *Let U be a sequential theory. Then, the following are equivalent:* (i) *U is strongly ℓ-reflexive;* (ii) $U \equiv_{\mathsf{loc}} U^{\mathsf{pc}}$; (iii) $U \equiv_{\mathsf{loc}} U^{\mathsf{pcf}}$.

Proof. (i) ⇒ (iii) Suppose U is strongly ℓ-reflexive. Then, $U \rhd (\mathsf{Q} + \mathsf{con}(U \restriction n))$. We assume that n is so large that sequentiality can be verified in $U \restriction n$. By Theorem 6.3, $(\mathsf{Q} + \mathsf{con}(U \restriction n)) \rhd (U \restriction n)^{\mathsf{pcf}}$. Clearly, $(U \restriction n)^{\mathsf{pcf}} \supseteq (U^{\mathsf{pcf}} \restriction n)$. We may conclude: $U \rhd (U^{\mathsf{pcf}} \restriction n)$. Thus, $U \rhd_{\mathsf{loc}} U^{\mathsf{pcf}}$. Trivially, we have $U^{\mathsf{pcf}} \rhd U$.

(iii) ⇒ (ii) This is immediate.

(ii) ⇒ (i) Let n be so large that $U \restriction n$ is sequential. We have, by Theorem 7.1:

$$\begin{aligned} U^{\mathsf{pc}} &\supseteq (U \restriction n)^{\mathsf{pc}} \\ &\rhd \mathsf{Q} + \mathsf{con}(U \restriction n) \\ &\rhd \mathsf{S}^1_2 + \mathsf{con}(U \restriction n). \end{aligned}$$

We may conclude that $U^{\mathsf{pc}} \rhd_{\mathsf{loc}} \mho_U^+$. □

14 One uses that EA^+ is EA plus uniform Σ^0_1-reflection for EA.

15 We use S^1_2 here in stead of $I\Delta_0 + \Omega_1$, because it is finitely axiomatized. Any sufficiently strong finitely axiomatized subtheory of $I\Delta_0 + \Omega_1$ would suffice for present purposes.

16 We pronounce $\mho$ to rhyme with 'Joe'.

17 An important difference between the functor $\mho^+$ and the mappings Θ and Ω is that $\mho^+$ is extensional.

18 We use '*strongly*' here to distinguish this notion from a notion of reflexiveness involving not just restriction of axioms but also restricted provability. We call this notion: *(weak) reflexiveness*.

19 Sequential theories are (weakly) ℓ-reflexive. Thus, there is an analogy between sequential theories and sequential strongly ℓ-reflexive theories.

We show that strong ℓ-reflexivity is a sufficient condition for non-finite axiomatizability.

Theorem 9.6. *No strongly ℓ-reflexive theory is finitely axiomatizable.*

Proof. Suppose T is strongly ℓ-reflexive and finitely axiomatized. We have, for some n, that $T = (T \upharpoonright n)$. Hence, $T \rhd (\mathsf{S}^1_2 + \mathsf{con}(T))$, contradicting the Second Incompleteness Theorem.[20] □

Note that it follows that theories like PA and ZF are not finitely axiomatizable. The most relevant corollary in the context of this paper is, of course:

Corollary 9.7. *The theory* $\mathsf{P}^{\omega}\mathsf{V}$ *is not finitely axiomatizable, nor is any theory that is mutually locally interpretable with* $\mathsf{P}^{\omega}\mathsf{V}$.

Corollary 9.8. *The theory* $\mathsf{P}^{\omega}\mathsf{V}$ *does not interpret* EA^+.

Proof. Since $\mathsf{EA}^+ \supseteq (\mathsf{Q})_\omega$, we have $\mathsf{EA}^+ \rhd_{\mathsf{loc}} \mathsf{P}^{\omega}\mathsf{V}$. Hence, it is impossible that we have $\mathsf{P}^{\omega}\mathsf{V} \rhd \mathsf{EA}^+$, since EA^+ is finitely axiomatizable. □

Open Question 9.9. It seems to me that every extension of EA studied in the literature is either finitely axiomatizable or (strongly) reflexive. On the other hand, it seems very probable to me that the theory $\mathsf{P}^{\omega}\mathsf{V}$ is not strongly reflexive, but just strongly locally reflexive. This makes the following question interesting. Is $\mathsf{P}^{\omega}\mathsf{V}$ strongly reflexive?

Acknowledgments

I am grateful to Lev Beklemishev, John Burgess, Mihai Ganea, Richard Heck, Rosalie Iemhoff, Joost Joosten and Charles Parsons for enlightening discussions.

Appendix A. Finite axiomatizability III

In this appendix we prove Theorem 3.3. We show the following. Suppose U is a one-sorted theory with identity and pairing. Then, U^{pc} can be finitely axiomatized *over* U.

Proof. We code sequences in U by:

- $\langle x \rangle := x, \sigma \circledast y := \langle \sigma, y \rangle$.

Here $\langle x_0, \ldots, x_{n-1}\rangle \circledast y = \langle x_0, \ldots, x_{n-1}, y\rangle$. Our definition by recursion is external. Internally, we only define, for any external n, sequences of length n. Note that a sequence of length n may coincide with a sequence of length k, for $n < k$. Clearly, we can define the notion n-seq of *sequence of length n* in U. It is convenient to also have the right associating notion of sequence $[x_0, \ldots, x_{n-1}]$.

- $[x] := x, y \boxast \sigma := \langle y, \sigma \rangle$.

We now define a number of operations on concepts.

- Suppose P is a k-ary predicate symbol, then $F_{0,P} := \wr \langle x_0, \ldots, x_{k-1}\rangle \mid P(x_0, \ldots, x_{k-1}) \int$.
- $F_1(X) := \wr \langle x, y\rangle \mid X\langle x, y, y\rangle \int$, $F_1^-(X) := \wr y \mid X\langle y, y\rangle \int$.
  Note that by our conventions this also identifies the last two elements of longer sequences.
- $F_2(x) := \wr x \int$.
- $F_3(X) := \wr \langle x, z, y\rangle \mid X\langle x, y, z\rangle \int$, $F_3^-(X) := \wr \langle z, y\rangle \mid X\langle y, z\rangle \int$.
- $F_4(X) := \wr \langle y, z, w\rangle \mid X[y, z, w] \int$.
- $F_5(X) := \wr [x, y, z, w] \mid X[x, \langle y, z, w\rangle] \int$.
- $F_6(X) := \wr \langle x, y\rangle \mid Xx \int$.
- $F_7(X, Y) := X \setminus Y$.
- $F_8(X) := \wr y \mid \exists x\, X\langle y, x\rangle \int$.

Let $\mathcal{F}$ be the class of functions generated by the functions of our list.

---

20 For the proof of the Second Incompleteness Theorem it is irrelevant on which interpretation of S^1_2 in T we have $\mathsf{con}(T)$. The Theorem is, in this sense, coordinate free.

Let $\mathsf{renorm}_n([x_0, \ldots x_{n-1}]) := \langle x_0, \ldots, x_{n-1}\rangle$. Suppose X is a set of sequences $[x_0, \ldots, x_{n-1}]$. We claim that $G_{0,n}$, with

$$G_{0,n}(X) := \mathsf{renorm}_n[X] := \wr \mathsf{renorm}_n(\xi) \mid \xi \in X \int,$$

is in $\mathcal{F}$. For $n = 1$ this is trivial. For $n > 1$, note that:

$$\langle\langle x_0, \ldots, x_k\rangle, [x_{k+1}, \ldots, x_{n-1}]\rangle \in F_4(X) \Leftrightarrow \langle\langle x_0, \ldots, x_{k-1}\rangle, [x_k, \ldots, x_{n-1}]\rangle \in X.$$

Hence, we can take $G_{0,n} := F_4^{n-2}$.

Let $\mathsf{renorm}^\star_n(\langle x_0, \langle x_1, \ldots, x_{n-1}\rangle\rangle) := [x_0, \ldots, x_{n-1}]$. Suppose X is a set of pairs $\langle x_0, \langle x_1, \ldots, x_{n-1}\rangle\rangle$. We claim that $G_{1,n}$, with $G_{0,n}(X) := \mathsf{renorm}^\star_n[X]$, is in $\mathcal{F}$. For $n = 1, 2$ this is trivial. For $n > 2$, we take $G_{1,n} := F_5^{n-3}$.

Suppose $n > 1$. Let $\mathsf{shift}_n(\langle x_0, \ldots, x_{n-1}\rangle) := \langle x_{n-1}, x_0, \ldots, x_{n-2}\rangle$. We claim that there is a $G_{2,n}$ in $\mathcal{F}$, such that if X is a concept of sequences of length n, then $G_{2,n}(X) := \mathsf{shift}_n[X]$. In case $n = 2$, we take $G_{2,2} := F_3^-$. Suppose $n > 2$. We take $G_{2,n} := G_{0,n} \circ G_{1,n} \circ F_3^-$.

Suppose $n > 2$ and $k < n - 2$. Define:

$$\mathsf{place}_{n,k}(\langle x_0, \ldots, x_{n-1}\rangle) := \langle x_0, \ldots, x_k, x_{n-1}, x_{k+1}, \ldots x_{n-2}\rangle.$$

Let $G_{4,n,k}(X) := \mathsf{place}_{n,k}[X]$. Then $G_{4,n,k}(X)$ is in $\mathcal{F}$. We can take $G_{4,n,k} := G_{2,n}^{k+2} \circ (G_{2,n} \circ F_{3,n})^{n-k-2}$.

Let π be any permutation of n. Let $\pi^\star(\langle x_0, \ldots, x_{n-1}\rangle) := \langle x_{\pi 0}, \ldots, x_{\pi(n-1)}\rangle$. We define: $G_{5,n,\pi}(X) := \pi^\star[X]$. Then $G_{5,n,\pi}$ is in $\mathcal{F}$. Using the previously defined operations, we can easily define $G_{5,n,\pi}$ for the case that π is a transposition. Then we use the fact that any permutation is a product of transpositions.

To any formula of predicate logic A (possibly with free concept variables), we assign the set of sequences ξ of length ℓ, were ℓ is the supremum-plus-one of the indices of the variables occurring in A, such that ξ, considered as an assignment satisfies A. We now prove by induction on A that this set of sequences is definable using the F_i. We treat the example of the atomic formula $P(v_2, v_0)$. The rest is more or less obvious. We have:

a. $X_0 := F_{0,P} = \wr \langle x_0, x_1\rangle \mid P(x_0, x_1) \int$.
b. $X_1 := F_6(X_0) = \wr \langle x_0, x_1, x_2\rangle \mid P(x_0, x_1) \int$.
c. $X_2 := G_{5,3,(02)(12)} = \wr \langle x_1, x_2, x_0\rangle \mid P(x_0, x_1) \int = \wr \langle x_0, x_1, x_2\rangle \mid P(x_2, x_0) \int$.

It follows that every definable set is generated by the F_i. Clearly, the F_i are definable using predicative comprehension. □

Is the proof of Theorem 3.3 verifiable in $I\Delta_0 + \Omega_1$? We did not verify this in detail, but it seems that if we proceed carefully and choose the variables in the syntactic representation of the formulas involved in comprehension wisely, then is should be feasible.

Appendix B. Questions

Q1. Is $\mathsf{P}^\omega\mathsf{V}$ bi-interpretable with a theory of finite signature? This is Question 4.4.
Q2. We do not know whether the argument for the finite axiomatizability of $\mathsf{P}^1\mathsf{V}$ can be formalized in $I\Delta_0 + \Omega_1$.
  (a) Can $I\Delta_0 + \Omega_1$ verify the finite axiomatizability of $\mathsf{P}^1\mathsf{V}$?
  (b) If not, is the finite axiomatizability of $\mathsf{P}^{n+1}\mathsf{V}$, for $n \geq 1$, verifiable in $I\Delta_0 + \Omega_1$?
  This is Question 4.8.
Q3. Is there a (natural) hierarchy of recursive functions F_n below superexponentiation, such that $I\Delta_0 +$ "F_n is total" is mutually interpretable with $\mathsf{Q} + \mathsf{con}^n(\mathsf{Q})$? This is Question 9.2.
Q4. Is $\mathsf{P}^\omega\mathsf{V}$ strongly reflexive? This is Question 9.9.

References

[1] L. Beklemishev, Proof-theoretic analysis by iterated reflection, Archive for Mathematical Logic 42 (2003) 515–552, doi:10.1007/s00153-002-0158-7.
[2] J. Burgess, Fixing Frege, in: Princeton Monographs in Philosophy, Princeton University Press, Princeton, 2005.
[3] G. Collins, J. Halpern, On the interpretability of arithmetic in set theory, The Notre Dame Journal of Formal Logic 11 (1970) 477–483.
[4] S. Feferman, Arithmetization of metamathematics in a general setting, Fundamenta Mathematicae 49 (1960) 35–92.
[5] S. Feferman, My route to arithmetization, Theoria 63 (1997) 168–181.
[6] M. Ganea, Burgess' PV is Robinson's Q, The Journal of Symbolic Logic 72 (2) (2007) 619–624.
[7] P. Hájek, P. Pudlák, Metamathematics of first-order arithmetic, in: Perspectives in Mathematical Logic, Springer, Berlin, 1991.
[8] D. Hilbert, P. Bernays, Grundlagen der mathematik II, Springer, Berlin, 1939, second edition: 1970.
[9] S. Iwan, On the untenability of Nelson's predicativism, Erkenntnis 53 (1–2) (2000) 147–154.
[10] S. MacLane, Categories for the Working Mathematician, in: Graduate Texts in Mathematics, vol. 5, Springer, New York, 1971.
[11] F. Montagna, A. Mancini, A minimal predicative set theory, The Notre Dame Journal of Formal Logic 35 (1994) 186–203.
[12] J. Mycielski, P. Pudlák, A. Stern, A lattice of chapters of mathematics (interpretations between theorems), in: Memoirs of the American Mathematical Society, vol. 426, AMS, Providence, Rhode Island, 1990.
[13] E. Nelson, Predicative Arithmetic, Princeton University Press, Princeton, 1986.
[14] P. Pudlák, Cuts, consistency statements and interpretations, The Journal of Symbolic Logic 50 (1985) 423–441.

[15] P. Pudlák, Review of Edward Nelson: Predicative arithmetic, The Journal of Symbolic Logic 53 (1988) 987–989.
[16] R.M. Solovay, Interpretability in set theories, Unpublished letter to P. Hájek. http://www.cs.cas.cz/~hajek/RSolovayZFGB.pdf, August 17 1976.
[17] A. Visser, Interpretability logic, in: P. Petkov (Ed.), Mathematical Logic (Proceedings of the Heyting 1988 Summer School in Varna, Bulgaria), Plenum Press, Boston, 1990, pp. 175–209.
[18] A. Visser, The formalization of interpretability, Studia Logica 51 (1991) 81–105.
[19] A. Visser, An inside view of exp, The Journal of Symbolic Logic 57 (1992) 131–165.
[20] A. Visser, Categories of theories and interpretations, in: A. Enayat, I. Kalantari, M. Moniri (Eds.), Logic in Tehran. Proceedings of the Workshop and Conference on Logic, Algebra and Arithmetic, Held October 18–22, 2003, in: Lecture Notes in Logic, vol. 26, ASL, A.K. Peters, Ltd., Wellesley, MA, 2006, pp. 284–341.
[21] H. Wang, Arithmetical models for formal systems, Methodos 3 (1951) 217–232.
[22] A. Wilkie, J. Paris, On the scheme of induction for bounded arithmetic formulas, Annals of Pure and Applied Logic 35 (1987) 261–302.