Interpreting weak König's lemma in theories of nonstandard arithmetic

Bruno Dinis Fernando Ferreira

Faculdade de Ciências
Universidade de Lisboa

Abstract

We show how to interpret weak König's lemma in some recently defined theories of nonstandard arithmetic in all finite types. Two types of interpretations are described, with very different verifications. The celebrated conservation result of Harvey Friedman about weak König's lemma can be proved using these interpretations. We also address some issues concerning the collecting of witnesses in herbrandized functional interpretations.

1 Introduction

Benno van den Berg, Eyvind Briseid and Pavol Safarik introduced and studied in [4] a cluster of functional interpretations of theories of nonstandard arithmetic. They were able to prove pertinent soundness theorems and, as a consequence, to extract computational information from proofs in these theories. In their work, van den Berg *et al.* started by considering intuitionistic nonstandard theories but their analysis was also adapted to the classical setting. Their interpretations are dubbed *herbrandized* because the witnesses of the existential quantifiers of the interpreting formulas are collected into finite sets. More recently, the second author of this paper and Jaime Gaspar studied in [9] a variation in which the existential witnesses are given below certain given bounds (using the notion of strong majorizability of Marc Bezem [6]). The goal for studying this variation was to show that the bounded functional interpretation of [10] can be seen (at least for the classical seeting of [7]) as the trace left by an interpretation of nonstandard arithmetic into a purely standard setting.

The theories considered in [4] and [9] are suitable for interpreting weak König's lemma. Both can interpret this lemma by internally considering Scott sets given by so-called standard systems (see [11] for these concepts). The theory in [9] can nevertheless interpret weak König's lemma in a more direct fashion, as we will see. This is in agreement with what Jeremy Avigad says in his report of [9] in [2]: "[it] is a natural setting for carrying out and eliminating certain kinds of compactness arguments." Even though the elimination of weak König's lemma makes perfect sense over a theory of the strength of Peano Arithmetic (without arithmetical comprehension), it is more traditional to consider weak König's lemma over theories enjoying at most Σ^0_1-induction. This is what we do in this paper and, accordingly, we recover the celebrated result of Harvey

Friedman that the theory of Reverse Mathematics WKL_0 is conservative over the theory $\mathsf{I}\Sigma^0_1$ for Π^0_2-sentences (Stephen Simpson's *opus* [16] is the place where the reader can find a thorough discussion of these matters). Friedman's result can also be proved using functional interpretations for theories of *standard* arithmetic. The first such proof is due to Ulrich Kohlenbach in [12]. The proof in [7] is related to the argument that we give in the next section.

This paper has another theme, *pace* the title of the paper. We have found that the manner of collecting finitely many witnesses in the herbrandized functional interpretations of [4] is somewhat artificial. This is not the place to elaborate on this opinion, but we advance now two reasons for this: one conceptual, the other connected with the work at hand. The conceptual reason lies in the fact that the herbrandized functional interpretations do not work directly with the arrow type $\sigma \to \tau^*$ between two types σ and τ^* (the superscript * means that one is considering the type of finite sequences of elements of type τ), but rather with the type $(\sigma \to \tau^*)^*$. One is nevertheless forced to see the elements of the type $(\sigma \to \tau^*)^*$ as, somehow, functions from the elements of type σ to the elements of type τ^*. This is possible, but at the cost of some artificiality. The other reason is that the properties of finiteness needed to pull through the herbrandized interpretation are very minimal. The herbrandized functional interpretation can actually be given a sense for pure classical logic, as shown in [8]. Of course, in the context of arithmetic, the notions of finiteness and (natural) number must be closely related on pain of straining one (or both) of them. In this paper we strove to separate the minimal properties of finiteness needed to take care of the herbrandized functional interpretation from the properties of finiteness that come hand in hand with the properties of number in our system.

The paper is organized as follows. In the next section we consider a (primitive recursive) version of the theory of [9] and show how WKL_0 can be interpreted in it. The following section does the same thing for the (classical) theory of [4]. The verifications of the interpretations are different in each case. The other theme of the paper, concerning the herbrandized functional interpretation, is developed in Section 4. In the closing section, we address the question of the treatment of finiteness in our weak theories and show how the conservation result of Friedman can be obtained from the interpretations discussed.

2 The interpretation of weak König's lemma in $\mathsf{E\text{-}PRA}^\omega_{\mathrm{st}}$

The theory $\mathsf{E\text{-}PA}^\omega_{\mathrm{st}}$ of nonstandard arithmetic was introduced in [9]. It is built upon the well-known theory of (standard) arithmetic $\mathsf{E\text{-}PA}^\omega$: Peano Arithmetic in all finite types with full extensionality (see [13] for the precise definition that we use, with only primitive equality at type 0). As discussed, we need to restrict this theory to a theory of the strength of primitive recursive arithmetic. The restriction amounts to allow only the recursor R_0 of type 0 and restrict induction to the quantifier-free induction principle:

$$\varphi(0) \wedge \forall x\,(\varphi(x) \to \varphi(x+1)) \to \forall x\,\varphi(x),$$

where φ is a quantifier-free formula. This is the theory $\mathsf{E\text{-}PRA}^\omega$. Note that the recursor R_0 is exactly what is needed to introduce the primitive recursive functions. A version of this theory is defined in [1], but we have full extensionality. Also, we do not have, nor need to have, the conditional functionals (although it would be harmless to adjoin them) of page 22 of [1]. For further clarification on this issue the reader can consult note 11 of [3].

The theory $\mathsf{E\text{-}PRA}^{\omega}_{\mathrm{st}}$ of nonstandard arithmetic extends this theory in the following way. With regard to the language one adds unary predicate symbols st^{σ} for each finite type σ (the predicates for *standardness*). The axioms are extended with the following standardness axioms:

1. $x =_{\sigma} y \to (\mathrm{st}^{\sigma}(x) \to \mathrm{st}^{\sigma}(y))$
2. $\mathrm{st}^{\sigma}(y) \to (x \leq^{*}_{\sigma} y \to \mathrm{st}^{\sigma}(x))$
3. $\mathrm{st}^{\sigma}(t)$, for each closed term t of type σ
4. $\mathrm{st}^{\sigma\to\tau}(z) \to (\mathrm{st}^{\sigma}(x) \to \mathrm{st}^{\tau}(zx))$

where the types σ and τ are arbitrary. Except for the second axiom, there is hardly any need for comment. Just notice that equality in type σ is defined extensionally. In the statement of the second standardness axiom, $x \leq^{*}_{\sigma} y$ means that x is strongly majorizable by y. The definition of strong majorizability is by induction on the finite type:

$x \leq^{*}_{0} y$ is $x \leq y$

$x \leq^{*}_{\rho\to\sigma} y$ is $\forall v \forall u \leq^{*}_{\rho} v\, (xu \leq^{*}_{\sigma} yv \wedge yu \leq^{*}_{\sigma} yv)$

For more information the reader can consult [13] where, instead of our notation $x \leq^{*}_{\sigma} y$, it is used y *s-maj*$_{\sigma}$ x. Notice that in type 0 the strong majorizability relation is the usual less than or equal relation. The second standardness axiom for type 0 is an obvious requirement and it permits do deduce the following form of induction:

$$\varphi(0) \wedge \forall^{\mathrm{st}} x\, (\varphi(x) \to \varphi(x+1)) \to \forall^{\mathrm{st}} x\, \varphi(x),$$

where φ is an internal quantifier-free formula. (An *internal* formula is a formula in which the standardness predicates do not occur. Note that the formulas in schematic position in both induction principles are the same.)

However, in higher types, the second standardness axioms are very peculiar. They are intrinsic to the functional interpretation of [9] and, somewhat surprisingly, they refute the so-called transfer principle of nonstandard arithmetic (even for universal formulas: cf. the appendix of [9]). The transfer principle is one of a triad of Edward Nelson's theory IST for nonstandard set theory (cf. [14]). Both *idealization* (the I of IST) and *standardization* (the S of IST) have, as we will comment, counterparts in the theories studied in [9] and [4]. The third piece of the triad, *transfer* (the T of IST) is missing because, as we have observed, it is incompatible with $\mathsf{E\text{-}PRA}^{\omega}_{\mathrm{st}}$. In the context of [4], transfer is related with comprehension-like principles (cf. [5]).

There are three principles that play a fundamental role in the functional interpretation of [9] (the formulas φ below are internal):

I. *Monotone choice* $\mathsf{mAC}^{\omega}_{\mathrm{st}}$: $\tilde{\forall}^{\mathrm{st}} x^{\tau} \tilde{\exists}^{\mathrm{st}} y^{\sigma}\, \varphi(x,y) \to \tilde{\exists}^{\mathrm{st}} h\, \tilde{\forall}^{\mathrm{st}} x \tilde{\exists} y \leq^{*}_{\sigma} hx\, \varphi(x,y)$.

II. *Idealization* I^{ω}: $\tilde{\forall}^{\mathrm{st}} z^{\sigma} \exists x^{\tau} \forall y \leq^{*}_{\sigma} z\, \varphi\,(x,y) \to \exists x \forall^{\mathrm{st}} y\, \varphi\,(x,y)$.

III. *Majorizability axioms* MAJ^ω: $\forall^{\mathrm{st}} x^\sigma \exists^{\mathrm{st}} y^\sigma (x \leq^*_\sigma y)$.

Idealization is fundamental from a nonstandard point of view because it guarantees the existence of nonstandard elements (of every type). As the name indicates, it is a form of idealization in the sense of Nelson, where one replaces being in a finite set by being majorized. Of course, the quantifications $\forall^{\mathrm{st}} x\,(\ldots)$ and $\exists^{\mathrm{st}} x\,(\ldots)$ abbreviate $\forall x\,(\mathrm{st}(x) \to \ldots)$ and $\exists x\,(\mathrm{st}(x) \wedge \ldots)$, respectively. The tilde above the quantifiers indicates that the quantification is restricted to monotone functionals, i.e., functionals x such that $x \leq^* x$. Monotone choice can be seen as a form of standardization. The rationale for the majorizability axioms is ultimately connected with the bounded functional interpretation of the second author and Paulo Oliva [10].

Theorem. *The theory* WKL_0 *is interpretable in the theory* $\mathsf{E\text{-}PRA}^\omega_{\mathrm{st}} + \mathsf{mAC}^\omega_{\mathrm{st}} + \mathsf{I}^\omega$.

Proof. The interpretation is as follows: the number sort is interpreted by the standard numbers, and the set sort is interpreted by the type 1 functionals x such that $x \leq 1$. (Here, $x \leq 1$ abbreviates $\forall w^0 (xw \leq 1)$. Note that this is the same as saying that $x \leq^* 1$, where here 1 stands for the constant function of type 1 which always takes the number 1 as value.) The membership operation $w^0 \in x^1$ is interpreted by $xw = 0$. This is the interpretation.

Except for the Σ^0_1-induction scheme, the first-order axioms of WKL_0 are obviously interpretable in $\mathsf{E\text{-}PRA}^\omega_{\mathrm{st}}$ alone. The interpretations of Σ^0_1-induction and Δ^0_1-comprehension scheme follow – by an adaptation of well-known arguments – from the following form of choice:

$$\text{(C)} \qquad \forall^{\mathrm{st}} x^0 \exists^{\mathrm{st}} y^0\, \varphi(x, y) \to \exists^{\mathrm{st}} f^1 \forall^{\mathrm{st}} x^0\, \varphi(x, fx),$$

where φ is an internal quantifier-free formula. This form of choice is a consequence of monotone choice. Monotone choice applied to x and y of type 0 gives the consequent $\exists^{\mathrm{st}} h^1 \forall^{\mathrm{st}} x^0 \exists y \leq hx\, \varphi(x, y)$. A functional witness f^1 as in (C) can be obtained from h by a bounded search (this is possible because φ is an internal quantifier-free formula). Let us briefly rehearse the above mentioned well-known arguments within our setting. Suppose that $\exists^{\mathrm{st}} y\, \varphi(0, y)$ and that $\forall^{\mathrm{st}} x\,(\exists^{\mathrm{st}} y\, \varphi(x, y) \to \exists^{\mathrm{st}} y\, \varphi(x + 1, y))$, where x and y are of type 0 and φ is an internal quantifier-free formula (possibly with parameters). Take a standard number y_0 such that $\varphi(0, y_0)$ and, by the choice principle (C), take a standard f^1 such that $\forall^{\mathrm{st}} x, y\,(\varphi(x, y) \to \varphi(x, fxy))$. Using the recursor R_0 it is possible to define g^1 such that $g0 = y_0$ and $\forall x\,(g(x + 1) = f(x, gx))$. By the standardness axioms 3 and 4, this g is standard. By the second form of induction, it is clear that $\forall^{\mathrm{st}} x\, \varphi(x, gx)$. Therefore, by the standardness axiom 4, $\forall^{\mathrm{st}} x \exists^{\mathrm{st}} y\, \varphi(x, y)$. We have proved that the interpretation of Σ^0_1-induction holds. In order to show that the interpretation of Δ^0_1-comprehension holds, suppose that $\forall^{\mathrm{st}} x\,(\exists y^{\mathrm{st}}\, \varphi(x, y) \leftrightarrow \forall^{\mathrm{st}} z\, \psi(x, z))$, where x, y and z are of type 0 and both φ and ψ are internal quantifier-free formulas (possibly with parameters). By the right-to-left direction of the above equivalence, we have $\forall^{\mathrm{st}} x \exists^{\mathrm{st}} y, z\,(\psi(x, z) \to \varphi(x, y))$. By the choice principle (C), take f^1 and g^1 standard such that $\forall^{\mathrm{st}} x\,(\psi(x, gx) \to \varphi(x, fx))$. Since φ is an internal quantifier-free formula, there is a type 1 functional h such that $\forall x\,(hx = 0 \leftrightarrow \varphi(x, fx))$. It is clear that we can take this functional so that $h \leq 1$. It is easy to see that $\forall^{\mathrm{st}} x\,(hx = 0 \leftrightarrow \exists^{\mathrm{st}} y\, \varphi(x, y))$ is provable in the theory $\mathsf{E\text{-}PRA}^\omega_{\mathrm{st}} + \mathsf{mAC}^\omega_{\mathrm{st}}$. This is what we need to interpret Δ^0_1-comprehension.

To finish the proof, it remains to verify weak König's lemma. This is the place where the principle of idealization enters. Weak König's lemma says that every infinite binary tree has an infinite path. Let us analyze the interpretation of this statement (sentence). Let T be a functional of type 1 such that $T \leq 1$. To say that T is a binary tree is interpreted by the conjunction of the sentence $\forall^{\mathrm{st}} x^0 \, (Tx = 0 \to Seq_2(x))$, where $Seq_2(x)$ is a primitive recursive predicate saying that x^0 is the code of a binary sequence, with the sentence

$$\forall^{\mathrm{st}} \sigma, \tau \, (T\sigma = 0 \wedge \tau \subseteq \sigma \to T\tau = 0),$$

where $\tau \subseteq \sigma$ says that the binary sequence coded by τ is an initial segment of the binary sequence coded by σ. The interpretation of the infinitude of T is

$$\forall^{\mathrm{st}} w^0 \exists^{\mathrm{st}} \sigma \, (T\sigma = 0 \wedge |\sigma| = w),$$

where $|\sigma|$ denotes the length of the sequence coded by σ. We claim that

$$\forall^{\mathrm{st}} w \exists \alpha \leq 1 \forall k \leq w \, (T(\bar{\alpha}k) = 0).$$

To see this, pick a standard number w. By the infinitude of T, take a binary sequence σ with $|\sigma| = w + 1$ and $T\sigma = 0$. Given that T is a tree, note that $T\tau = 0$, for all $\tau \subseteq \sigma$. Now, let α^1 be such that $\alpha k = \sigma_k$ for all $k \leq w$, and $\alpha k = 0$ otherwise. It is clear that this α does the job. Here we are using the notation σ_k for the kth entry of σ and $\bar{\alpha}k$ for the (code of the) sequence $\langle \alpha 0, \ldots, \alpha(k-1) \rangle$. By I^ω, the claim readily entails

$$\exists \alpha \leq 1 \forall^{\mathrm{st}} k \, (T(\bar{\alpha}k) = 0).$$

This is the interpretation of the sentence saying that T has an infinite path.

□

3 The interpretation of weak König's lemma in $\mathsf{E\text{-}PRA}^{\omega^*}_{\mathrm{st}}$

The theory of nonstandard arithmetic $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$ was introduced in [4]. It is based upon the theory of (standard) arithmetic $\mathsf{E\text{-}PA}^{\omega*}$. They are both theories of arithmetic in all finite types and the notational star is meant to indicate that there is a type constructor that, to each finite type σ, associates the type σ^* of all finite sequences of elements of type σ (this is known as Kleene's star operation in semigroup theory). In fact, what is really needed for the following is that, given a type σ, we can speak of (non-empty) finite sets of elements of type σ. This is accomplished by speaking instead of finite sequences of elements of type σ and by saying that an element belongs to such a finite sequence if it is an entry of that sequence. One needs to introduce some special functionals and axioms in order to deal with the new types. Among these new functionals figure the so-called list recursors which, *prima facie*, pose a problem for adapting the theories $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$ and $\mathsf{E\text{-}PA}^{\omega*}$ to the weaker setting of primitive recursion. The present paper has to deal with this technical issue somehow. What we chose to do is the following. In this section we define the theories $\mathsf{E\text{-}PRA}^{\omega*}_{\mathrm{st}}$ and $\mathsf{E\text{-}PRA}^{\omega*}$ but we omit the treatment of the star type. This will be done in the last section of this paper. Meanwhile, in order to follow this section, the reader only needs to know that one can define in the language of $\mathsf{E\text{-}PRA}^{\omega *}_{\mathrm{st}}$ a closed term $\mathrm{M} : 0^* \to 0$ such that the theory $\mathsf{E\text{-}PRA}^{\omega *}_{\mathrm{st}}$ proves

$$\text{(P)} \qquad \forall c^{0^*} \forall x^0 (x \in c \to x \leq \mathrm{M}c).$$

So, we will assume this property in the present section.

As indicated, the theory $\mathsf{E\text{-}PRA}^{\omega *}$ extends $\mathsf{E\text{-}PRA}^{\omega}$ by enriching the term language with combinators (and associated axioms) for the new types, and some special constants and axioms in order to deal with the star types. Note that, as in $\mathsf{E\text{-}PRA}^{\omega *}$, there is only the recursor R_0 and induction is still restricted to quantifier-free formulas (there are now more quantifier-free formulas because there are new types and new constants, but this is not a very important feature of the new system). Similarly to the previous section, the theory $\mathsf{E\text{-}PRA}^{\omega *}_{\mathrm{st}}$ of nonstandard arithmetic extends $\mathsf{E\text{-}PRA}^{\omega *}$ by having unary standardness predicates for each type (including, of course, the new star types) and the following standardness axioms:

1. $x =_\sigma y \to (\mathrm{st}^\sigma(x) \to \mathrm{st}^\sigma(y))$
2. $\mathrm{st}^0(y) \to (x \leq y \to \mathrm{st}^0(x))$
3. $\mathrm{st}^\sigma(t)$, for each closed term t of type σ
4. $\mathrm{st}^{\sigma \to \tau}(z) \to (\mathrm{st}^\sigma(x) \to \mathrm{st}^\tau(zx))$

Observe that the second standardness axiom is restricted to type 0 (Berg *et al.* do not need this second axiom because they can prove it using external induction). There are two principles that play a fundamental role in the interpretation of [4] (the formulas φ below are internal):

I. *Herbrandized choice* $\mathsf{HAC}_{\mathrm{int}}$: $\forall^{\mathrm{st}} x^\tau \exists^{\mathrm{st}} y^\sigma\, \varphi(x, y) \to \exists^{\mathrm{st}} h \forall^{\mathrm{st}} x \exists y \in hx\, \varphi(x, y)$.

II *Idealization* I: $\forall^{\mathrm{st}} z^{\sigma^*} \exists x^\tau \forall y \in z\, \varphi(x, y) \to \exists x^\tau \forall^{\mathrm{st}} y^\sigma \varphi(x, y)$.

The first thing to notice is that the principle of choice (C) of the previous section is a consequence of $\mathsf{HAC}_{\mathrm{int}}$ and of the existence of the functional M. In fact, for x and y standard of type 0, hx is standard of type 0^* and, by property (P), all elements of hx fall below $\mathrm{M}(hx)$. As usual, a precise (standard) witness functional f can be obtained by bounded search. On the other hand, idealization entails the existence of nonstandard elements. To see this, notice that we have $\forall^{\mathrm{st}} c^{0^*} \exists u^0 \forall x \in c\, (x \leq u)$. This is immediate by property (P). Hence, by idealization, $\exists u^0 \forall^{\mathrm{st}} x^0 (x \leq u)$. This u is obviously nonstandard.

Theorem. *The theory* WKL_0 *is interpretable in the theory* $\mathsf{E\text{-}PRA}^{\omega *}_{\mathrm{st}} + \mathsf{HAC}_{\mathrm{int}} + \mathsf{I}$.

Proof. As it is well known, in primitive recursive arithmetic we can code finite sets by numbers. Therefore, in $\mathsf{E\text{-}PRA}^{\omega *}_{\mathrm{st}}$ there are two ways of speaking about finite sets of numbers: via elements of type 0^* and by elements of type 0 which are primitive recursive codes of finite sets of numbers. In our setting, these two manners amount essentially to the same thing (see Section 5). However, in this proof, we make a distinction between finite sets of numbers (given by elements of type 0^*) and codes of finite sets of numbers (elements of type 0).

The interpretation is as follows: the number sort is interpreted by the standard numbers, and the set sort is interpreted by number codes (standard and nonstandard) of finite sets of numbers (standard and nonstandard). The membership operation $w^0 \in x^0$ is interpreted by "w is an element of the finite set coded by x." As opposed to the interpretation of the last section, the second-order sort is interpreted by elements, not of type 1, but of type 0. It is, in short, interpreted by the standard system considered internally within the theory $\mathsf{E\text{-}PRA}^{\omega *}_{\mathrm{st}} + \mathsf{HAC}_{\mathrm{int}} + \mathsf{I}$.

As before, with the exception of Σ^0_1-induction, the first-order axioms of WKL_0 are clearly interpretable in $\mathsf{E\text{-}PRA}^{\omega *}_{\mathrm{st}}$ alone, and Σ^0_1-induction itself follows, as in the last section, from the choice principle (C). For recursive comprehension and weak König's lemma, we repeat the argument of Theorem 4.4 of [1]. The argument is as follows. Using the proof of Lemma IV.4.4 of [16], it is clear that both of these principles are a consequence of the so-called scheme of Σ^0_1-separation. An instance of this scheme is an implication with antecedent

$$\forall x \neg (\exists y\, \varphi(x,y) \wedge \exists z\, \psi(x,z))$$

and consequent

$$\exists V \forall x\, ((\exists y\, \varphi(x,y) \to x \in V) \wedge (\exists z\, \psi(x,z) \to x \notin V)).$$

The formulas in schematic position φ and ψ are bounded internal formulas, possibly with parameters. The sentence $\forall^{\mathrm{st}} x \neg (\exists^{\mathrm{st}} y\, \varphi(x,y) \wedge \exists^{\mathrm{st}} z\, \psi(x,z))$ is the interpretation of the antecedent. Fix a nonstandard number ω. Consider the code v^0 of the finite set

$$\{x < \omega : \exists y < \omega\, (\varphi(x,y) \wedge \forall z < y \neg \psi(x,z))\}.$$

This code exists because the theory $\mathsf{E\text{-}PRA}^{\omega *}_{\mathrm{st}}$ is able to code bounded sets defined by comprehension using bounded (internal) formulas. It is easy to check that v^0 codes a set that separates the standard elements x such that $\exists^{\mathrm{st}} y\, \varphi(x,y)$ from those that satisfy $\exists^{\mathrm{st}} z\, \psi(x,z)$. $\square$

Plainly, WKL_0 can also be interpreted in the theory $\mathsf{E\text{-}PRA}^{\omega}_{\mathrm{st}} + \mathsf{mAC}^{\omega}_{\mathrm{st}} + \mathsf{I}^{\omega}$ of the previous section using the interpretation of the above proof.

The interpretations of this and the previous section are, in some sense, the same because it is possible to go from (type 1) Boolean functions to finite sets below certain nonstandard elements, and vice-versa. However, the *checking* that the interpretations work are *very different* in each case.

4 The herbrandized functional interpretation

In this section we present a more natural way of dealing with the finite collections of witnesses in the herbrandized functional interpretations. As it happens, the solution is – in the setting of classical logic – rather straightforward. As an aside, there is also a solution, not as straightforward (but also rather simple), for semi-intuitionistic systems. This will be the subject of a future paper of the second author.

We present a herbrandized functional interpretation of the theory $\mathsf{E\text{-}PRA}^{\omega *}_{\mathrm{st}} + \mathsf{HAC}_{\mathrm{int}} + \mathsf{I}$ and prove a corresponding soundness theorem (the verification of the soundness is done in the theory $\mathsf{E\text{-}PRA}^{\omega*}$). Even though we have not yet fully described $\mathsf{E\text{-}PRA}^{\omega*}$ (nor its nonstandard version), this is not important for most of the arguments in this section. In fact, the missing axioms are internal and the interpretation keeps the internal formulas unchanged.

What is important for this section is to isolate the properties of finiteness needed to prove the soundness theorem. We need three kinds of functionals. Functionals of type $\sigma \to \sigma^*$ whose intended meaning is to map each element x^σ to the singleton set $\{x\}$. Functionals of type $\sigma^* \to \sigma^* \to \sigma^*$ whose intended meaning is to map elements c and d of type σ^* to their union $c \cup d$. Finally, we also need functionals of type $(\sigma \to \tau^*) \to (\sigma^* \to \tau^*)$ whose intended meaning is to map $f : \sigma \to \tau^*$ and $c : \sigma^*$ to the union $\bigcup_{w \in c} fw$. From a purely logical point of view, we only need the following three properties (in our case, provably in $\mathsf{E\text{-}PRA}^{\omega*}$):

(a) $x \in \{x\}$

(b) $x \in c \vee x \in d \to x \in c \cup d$

(c) $z \in c \wedge x \in fz \to x \in \bigcup_{w \in c} fw$

We will also apply the usual logical properties of equality (indiscernibilty of identicals) but it is curious to observe that we do not need to have that co-extensional elements of star type are equal. It is also curious that the empty set is nowhere needed.

The herbrandized functional interpretation below is done directly for classical logic. Our direct treatment is such that the existential variables of the interpreting formulas are always of star type and, moreover, a monotonicity property is sustained. This is quite different from the treatment in [4] of the classical case. In fact, while the intuitionistic interpretation of Berg *et al.* in [4] is herbrandized with a typical monotonicity property associated, their treatment of the classical case does not seem to have an associated monotonicity property, even though it is obtained from the intuitionistic case (via a negative translation).

We work with the logical calculus of Joseph Shoenfield in [15], with primitives $\neg$, $\vee$ and $\forall$.

Definition. *To each formula Φ of the language of $\mathsf{E\text{-}PRA}^{\omega*}_{\mathrm{st}}$, we assign formulas $\Phi^{\mathrm{S_{st}}}$ and $\Phi_{\mathrm{S_{st}}}$ so that $\Phi^{\mathrm{S_{st}}}$ is of the form $\forall^{\mathrm{st}} a \exists^{\mathrm{st}} b \, \Phi_{\mathrm{S_{st}}}(a, b)$, with $\Phi_{\mathrm{S_{st}}}(a, b)$ an* internal *formula, according to the following clauses:*

1. $\Phi^{\mathrm{S_{st}}}$ *and* $\Phi_{\mathrm{S_{st}}}$ *are simply* Φ*, for internal formulas* Φ*,*

2. $(\mathrm{st}(t))^{\mathrm{S_{st}}}$ *is* $\exists^{\mathrm{st}} c \, (t \in c)$*.*

For the remaining *cases, if we have already interpretations for Φ and Ψ given (respectively) by $\forall^{\mathrm{st}} a \exists^{\mathrm{st}} b \, \Phi_{\mathrm{S_{st}}}(a, b)$ and $\forall^{\mathrm{st}} d \exists^{\mathrm{st}} e \, \Psi_{\mathrm{S_{st}}}(d, e)$ then we define:*

3. $(\Phi \vee \Psi)^{\mathrm{S_{st}}}$ *is* $\forall^{\mathrm{st}} a, d \, \exists^{\mathrm{st}} b, e \, [\Phi_{\mathrm{S_{st}}}(a, b) \vee \Psi_{\mathrm{S_{st}}}(d, e)]$*,*

4. $(\neg \Phi)^{\mathrm{S_{st}}}$ *is* $\forall^{\mathrm{st}} f \exists^{\mathrm{st}} a' \, [\exists a \in a' \, \neg \Phi_{\mathrm{S_{st}}}(a, fa)]$*,*

5. $(\forall x \, \Phi(x))^{\mathrm{S_{st}}}$ *is* $\forall^{\mathrm{st}} a \exists^{\mathrm{st}} b \, [\forall x \, \Phi_{\mathrm{S_{st}}}(a, b, x)]$*,* *where the internal formulas between square brackets are the corresponding* lower $\mathrm{S_{st}}$*-formulas.*

The letters a and b (and similar) above stand for (possibly empty) tuples of variables, but we speak of them as if they were only one variable. The abbreviations are as usual, but we need to be clear about the quantification $\exists^{\mathrm{st}} x\,(\ldots)$: it abbreviates $\neg\forall x\,(\neg\mathrm{st}(x) \vee \neg(\ldots))$. Do notice, and this is important, that if $\Phi^{\mathrm{S_{st}}}$ is $\forall^{\mathrm{st}} a \exists^{\mathrm{st}} b\, \Phi_{\mathrm{S_{st}}}(a, b)$, then b is of star type (not so for a though, but this does not cause trouble). Hence, the following monotonicity property makes sense and, moreover, it is easily seen to be provable in $\mathsf{E\text{-}PRA}^{\omega *}$ (cf. Section 5 for the complete definition of this theory):

$$\Phi_{\mathrm{S_{st}}}(a, b) \wedge b \subseteq c \to \Phi_{\mathrm{S_{st}}}(a, c).$$

Of course, $b \subseteq c$ abbreviates $\forall x\,(\neg x \in b \vee x \in c)$.

Lemma. *Let φ and ψ be internal formulas. Then:*

(i) $(\forall^{\mathrm{st}} x\, \varphi(x))^{\mathrm{S_{st}}}$ *is* $\forall^{\mathrm{st}} c \forall x \in c\, \varphi(x)$.

(ii) $(\exists^{\mathrm{st}} x\, \varphi(x))^{\mathrm{S_{st}}}$ *is* $\exists^{\mathrm{st}} c' \exists c \in c' \exists x \in c\, \varphi(x)$.

(iii) $(\forall^{\mathrm{st}} x \exists^{\mathrm{st}} y\, \psi(x, y))^{\mathrm{S_{st}}}$ *is* $\forall^{\mathrm{st}} d \exists^{\mathrm{st}} c' \forall x \in d \exists c \in c' \exists y \in c\, \psi(x, y)$.

It is easy to check the above claims. *Morally*, the interpretation of $\exists^{\mathrm{st}} x\, \varphi(x)$ should be $\exists^{\mathrm{st}} c \exists x \in c\, \varphi(x)$ but formally it is as in (ii) above.

Theorem (Soundness)**.** *Suppose that*

$$\mathsf{E\text{-}PRA}^{\omega *}_{\mathrm{st}} + \mathsf{HAC}_{\mathrm{int}} + \mathsf{I} \vdash \Phi,$$

where Φ is an arbitrary formula (it may have free variables). Then there are closed *terms t of appropriate types such that*

$$\mathsf{E\text{-}PRA}^{\omega *} \vdash \forall a\, \Phi_{\mathrm{S_{st}}}(a, ta).$$

Proof. The proof is by induction on the number of steps of the derivation of Φ. We rely on the complete axiomatization of classical logic described in sections 2.6 and 8.3 of [15]. We omit the study of the law of excluded middle and of the propositional rules since the checking is essentially the one done in [8] (the property of monotonicity is used in this part). The situation with the axiom of substitution and the rule of $\forall$-introduction is different because the clause for the interpretation of the universal quantifier is unusual in nonstandard arithmetic.

Substitution: $\forall x\, \Phi(x) \to \Phi(t)$. Suppose that $\Phi(x)^{\mathrm{S_{st}}}$ is $\forall^{\mathrm{st}} a \exists^{\mathrm{st}} b\, \Phi_{\mathrm{S_{st}}}(a, b, x)$. Then $(\forall x\, \Phi(x) \to \Phi(t))^{\mathrm{S_{st}}}$ is

$$\forall^{\mathrm{st}} f, d \exists^{\mathrm{st}} a', e\, (\forall a \in a' \forall x\, \Phi_{\mathrm{S_{st}}}(a, fa, x) \;\to\; \Phi_{\mathrm{S_{st}}}(d, e, t)).$$

We need to exhibit close terms t and q such that

$$\forall f, d\, (\forall a \in tfd \forall x\, \Phi_{\mathrm{S_{st}}}(a, fa, x) \;\to\; \Phi_{\mathrm{S_{st}}}(d, qfd, t)).$$

It is clear that we can take terms t and q such that $tfd = \{d\}$ and $qfd = fd$.

$\forall$-*introduction*: $\Phi(x) \vee \Psi \Rightarrow \forall x\, \Phi(x) \vee \Psi$, where x does not occur free in Ψ. Let $\Phi(x)^{S_{st}}$ be as above and $\Psi^{S_{st}}$ be $\forall^{st} d \exists^{st} e\, \Psi_{S_{st}}(d, e)$. By induction hypothesis there are closed terms t and q such that, for all x,

$$\forall a, d\, (\Phi_{S_{st}}(a, tad, x) \vee \Psi_{S_{st}}(d, qad)).$$

Hence, $\forall a, d\, (\forall x\, \Phi_{S_{st}}(a, tad, x) \vee \Psi_{S_{st}}(d, qad))$, as wished. The fact that the terms t and q are closed (hence, they do not depend on x) is crucial in this step.

All the arithmetical axioms are internal and, hence, trivially interpreted. Let us now turn to the four standardness axioms of Section 3. The interpretation of the first of these axioms is

$$\forall^{st} c \exists^{st} d\, (x = y \to (x \in c \to y \in d)).$$

Obviously, we can put d to be c. The interpretation of the third standardness axiom is $\exists^{st} c\, (t \in c)$. Here the witnessing term $\{t\}$ works (this term is closed because t is). The interpretation of the fourth standardness axiom is $\forall^{st} c, d\, \exists^{st} e\, (z \in c \to (x \in d \to zx \in e))$. We need to find a closed term t such that

$$\forall c, d\, (z \in c \to (x \in d \to zx \in tcd)).$$

We can just take t such that $tcd = \bigcup_{z \in c} \bigcup_{x \in d} \{zx\}$. It remains to check the second standardness axiom of Section 3. Its interpretation is

$$\forall^{st} c^{0^*} \exists^{st} d^{0^*}\, (y \in c \to (x \leq y \to y \in d)).$$

Here we need a functional $\mathrm{E} : 0 \to 0^*$ such that the theory $\mathsf{E\text{-}PRA}^{\omega *}$ proves $\forall x^0, y^0\, (x \leq y \to x \in \mathrm{E}y)$. This functional exists: see Section 5. Using the functional E, we can define a closed term t with $tc = \bigcup_{w \in c} \mathrm{E}w$. It is clear that

$$\forall c^{0^*} \forall x^0, y^0 (y \in c \to (x \leq y \to x \in tc)).$$

Let us now check the principles of idealization and herbrandized choice. We look at idealization first. A simple computation shows that the interpretation of idealization is

$$\forall^{st} e' \exists^{st} c'\, (\forall c \in c' \forall z \in c \exists x \forall y \in z\, \varphi(x, y) \to \exists x \forall e \in e' \forall y \in e\, \varphi(x, y)).$$

Therefore, we need to exhibit a closed term t such that

$$\forall e'\, (\forall c \in te' \forall z \in c \exists x \forall y \in z\, \varphi(x, y) \to \exists x \forall e \in e' \forall y \in e\, \varphi(x, y)).$$

We can take t such that $te' = \{\{\bigcup_{e \in e'} e\}\}$.

The analysis of the principle of herbrandized choice is more interesting. By (iii) of the fact, the interpretation of $\neg \forall^{st} x \exists^{st} y\, \varphi(x, y)$ is

$$\forall^{st} f \exists^{st} d' \exists d \in d' \neg \forall x \in d \exists c \in fd \exists y \in c\, \varphi(x, y).$$

A simple computation shows that the interpretation of the consequent of $\mathsf{HAC}_{\mathrm{int}}$ is

$$\forall^{st} \delta \exists^{st} h'' \exists h' \in h'' \exists h \in h' \forall d \in \delta h' \forall x \in d \exists y \in hx\, \varphi(x, y).$$

The typing is as follows: $x : \tau$, $y : \sigma$, $c : \sigma^*$, $d : \tau^*$, $d' : \tau^{**}$, $f : \tau^* \to \sigma^{**}$, $h : \tau \to \sigma^*$, $h' : (\tau \to \sigma^*)^*$, $h'' : (\tau \to \sigma^*)^{**}$ and $\delta : (\tau \to \sigma^*)^* \to \tau^{**}$. We need to exhibit closed terms t and q such that, for all f and δ,

$$\text{(H)} \qquad \forall d \in tf\delta\, \forall x \in d\, \exists c \in fd\, \exists y \in c\ \varphi(x,y) \to$$
$$\exists h' \in qf\delta\, \exists h \in h'\, \forall d \in \delta h'\, \forall x \in d\, \exists y \in hx\ \varphi(x,y).$$

We take q such that $qf\delta = \{\{\lambda x. \bigcup_{c \in f\{x\}} c\}\}$ and t such that

$$tf\delta = \bigcup_{d \in \delta\left(\left\{\lambda x. \bigcup_{c \in f\{x\}} c\right\}\right)} \bigcup_{w \in d} \{\{w\}\}.$$

Assume that $\forall d \in tf\delta\, \forall x \in d\, \exists c \in fd\, \exists y \in c\ \varphi(x,y)$. In order to see (H), take $d_0 \in \delta(\{\lambda x. \bigcup_{c \in f\{x\}} c\})$ and $x_0 \in d_0$ in order to show $\exists y \in \bigcup_{c \in f\{x_0\}} c\ \varphi(x_0,y)$. Of course, $\{x_0\} \in tf\delta$. So, $\exists c \in f\{x_0\}\, \exists y \in c\ \varphi(x_0,y)$. This is what we want. $\square$

5 Harvey Friedman's conservation result

The interpretations of Section 2 and Section 3 can be used to prove that the theory WKL_0 is Π^0_2-conservative over the theory $\mathsf{I}\Sigma^0_1$. Suppose that WKL_0 proves a certain Π^0_2-sentence $\forall x \exists y\, \varphi(x,y)$, where φ is a bounded formula of the language of arithmetic. Since WKL_0 is interpretable in $\mathsf{E\text{-}PRA}^\omega_{\mathrm{st}} + \mathsf{mAC}^\omega_{\mathrm{st}} + \mathsf{I}^\omega$ (Section 2), this latter theory proves $\forall^{\mathrm{st}} x \exists^{\mathrm{st}} y\, \varphi(x,y)$. The following proposition is a consequence of the soundness theorem of [9] (suitably adapted for the primitive recursive setting):

Proposition. *If* $\mathsf{E\text{-}PRA}^\omega_{\mathrm{st}} + \mathsf{mAC}^\omega_{\mathrm{st}} + \mathsf{I}^\omega + \mathsf{MAJ}^\omega \vdash \forall^{\mathrm{st}} x \exists^{\mathrm{st}} y\, \theta(x,y)$*, where* θ *is an internal quantifier-free formula, then* $\mathsf{E\text{-}PRA}^\omega \vdash \tilde{\forall} x \exists y\, \theta(x,y)$.

So, $\mathsf{E\text{-}PRA}^\omega \vdash \forall x \exists y\, \varphi(x,y)$. We claim that this implies $\mathsf{I}\Sigma^0_1 \vdash \forall x \exists y\, \varphi(x,y)$. This is the case because the theory $\mathsf{E\text{-}PRA}^\omega$ is interpretable into $\mathsf{I}\Sigma^0_1$. The interpretation is given by the model HEO of the hereditarily effective operations. This internal model is well-known in the literature for interpreting $\mathsf{E\text{-}PA}^\omega$ into Peano Arithmetic PA (see, for instance, [17]), but the argument also works for our case. Let us briefly describe the model. We need to define the ranges of the various finite types. To do this, we define in tandem the ranges of the finite types and certain binary relations $=_\sigma$. The definition is inductive (on the building of types): N_0 is the universal predicate and $=_0$ is the equality relation for natural numbers, $N_{\sigma\to\tau}(e)$ is defined by

$$\forall x\, (N_\sigma(x) \to \{e\}(x)\downarrow) \wedge \forall x, y\ (N_\sigma(x) \wedge N_\sigma(y) \wedge x =_\sigma y \to \{e\}(x) =_\tau \{e\}(y))$$

and $e =_{\sigma\to\tau} d\ :\equiv\ \forall x\, (N_\sigma(x) \to \{e\}(x) =_\tau \{d\}(x))$. Application between e such that $N_{\sigma\to\tau}(e)$ and x such that $N_\sigma(x)$ is defined to be Kleene's application $\{e\}(x)$. It is not difficult to complete the definition of this model. (Note that we have claimed that the theory $\mathsf{E\text{-}PRA}^\omega$ is interpretable into $\mathsf{I}\Sigma^0_1$, not in mere primitive recursive arithmetic. If the reader thinks through, Σ^0_1-induction is needed when we try to make the argument for the interpretation explicit.)

As already mentioned, we could have used the interpretation of Section 3 instead. In that case, we would have gotten $\mathsf{E\text{-}PRA}^{\omega*}_{\mathrm{st}} + \mathsf{HAC}_{\mathrm{int}} + \mathsf{I} \vdash \forall^{\mathrm{st}} x \exists^{\mathrm{st}} y\, \varphi(x,y)$.

The following proposition is a consequence of the soundness theorem in [4] for classical arithmetic (suitably adapted for the primitive recursive setting), or alternatively, it is a consequence of the soundness theorem of Section 4:

Proposition. *If* $\mathsf{E\text{-}PRA}^{\omega*}_{\mathrm{st}} + \mathsf{HAC}_{\mathrm{int}} + \mathsf{I} \vdash \forall^{\mathrm{st}} x \exists^{\mathrm{st}} y\, \theta(x, y)$*, where* θ *is an internal quantifier-free formula, then* $\mathsf{E\text{-}PRA}^{\omega*} \vdash \forall x \exists y\, \theta(x, y)$*.*

In order to obtain Friedman's conservation result, we need to extend the internal model HEO of $\mathsf{I}\Sigma^0_1$ to the theory $\mathsf{E\text{-}PRA}^{\omega*}$. At this juncture, we really need to have the complete definition of the theory $\mathsf{E\text{-}PRA}^{\omega*}$. So far, we have omitted some constants and axioms needed to deal with the star types.

The treatment of the finite sequences in star types given in [4] cannot, *prima facie*, be adapted to our weak setting because of the so-called list recursors. List recursors can simulate primitive recursion in finite types and, as a consequence, yield a theory of the strength of Peano Arithmetic. We opt for a direct and simple-minded treatment. For each type σ, we include in $\mathsf{E\text{-}PRA}^{\omega*}$ three constants: C_σ, $|\cdot|_\sigma$ and F_σ. The constant C_σ has type $(0 \to \sigma) \to 0 \to \sigma^*$ and its intended meaning is to map an infinite sequence $f : 0 \to \sigma$ and a natural number l to the truncation of f at length l. The constant $|\cdot|_\sigma$ has type $\sigma^* \to 0$ and its intended meaning is to map an element s of type σ^* (a finite sequence) to its length $|s|_\sigma$. The constant F_σ has type $\sigma^* \to (0 \to \sigma)$ and its intended meaning is to map $s : 0^*$ to a functional $\mathrm{F}_\sigma s : 0 \to \sigma$ such that $\mathrm{F}_\sigma sn = s_n$, for $n < |s|_\sigma$ (here, s_n is the n-th entry of s).

We accept three kinds of axioms: $|\mathrm{C}fl| = l$, $\forall n < l\, (\mathrm{F}(\mathrm{C}fl)n = fn)$ and $\mathrm{C}(\mathrm{F}s)|s| = s$. Given x of type σ and s of type σ^*, we say that $x \in s$ if $\exists n < |s|_\sigma\, (\mathrm{F}_\sigma sn = x)$. Note that, by the standardness axioms of Section 3, if $\mathrm{st}^{\sigma^*}(s)$ and $x \in s$, then $\mathrm{st}^\sigma(x)$. It is clear that with the help of these functionals and axioms one can define the singleton functionals, the binary union functionals and the indexed union functionals of Section 4 and prove the corresponding three properties. The functional M of property (P) of Section 3 can be easily defined: it is $\lambda s^{0^*}. \max_{0 \le n < |s|} \mathrm{F}_0 sn$. The functional E of Section 4 is $\lambda y^0. \mathrm{C}_0(\lambda x^0.x)(y{+}1)$.

We are now ready to extend the internal model HEO to the new star types. Given a type σ, we say that $N_{\sigma^*}(e)$ holds if e is the code of a finite sequence of natural numbers, each of which satisfies the property N_σ. Also, we say that $e =_{\sigma^*} d$ if e and d code finite sequences of the same length l, and $e_n =_\sigma d_n$ for each $n < l$. It is clear that the functionals C_σ, $|\cdot|_\sigma$ and F_σ can be given in HEO, even *independently* of the type σ. The definition of the extension is completed.

References

[1] J. Avigad. Weak theories of nonstandard arithmetic and analysis. In Stephen Simpson, editor, *Reverse Mathematics 2001*, volume 21 of *Lecture Notes in Logic*, pages 19–46. A K Peters, Massachusetts, 2005.

[2] J. Avigad. Review of [9]. *Mathematical Reviews*, MR3325571, 2016.

[3] J. Avigad and S. Feferman. Gödel's functional ("Dialectica") interpretation. In S. R. Buss, editor, *Handbook of Proof Theory*, volume 137 of *Studies in Logic and the Foundations of Mathematics*, pages 337–405. North Holland, Amsterdam, 1998.

[4] B. van den Berg, E. Briseid, and P. Safarik. A functional interpretation for nonstandard arithmetic. *Annals of Pure and Applied Logic*, 163(12):1962 – 1994, 2012.

[5] B. van den Berg and S. Sanders. Transfer equals comprehension. arXiv: 1409.6881, 2014.

[6] M. Bezem. Strongly majorizable functionals of finite type: a model for bar recursion containing discontinuous functionals. *The Journal of Symbolic Logic*, 50:652–660, 1985.

[7] F. Ferreira. Injecting uniformities into Peano arithmetic. *Annals of Pure and Applied Logic*, 157:122–129, 2009.

[8] F. Ferreira and G. Ferreira. A functional interpretation of classical logic. Manuscript, 2016.

[9] F. Ferreira and J. Gaspar. Nonstandardness and the bounded functional interpretation. *Annals of Pure and Applied Logic*, 166(6):701–712, 2015.

[10] F. Ferreira and P. Oliva. Bounded functional interpretation. *Annals of Pure and Applied Logic*, 135:73–112, 2005.

[11] R. Kaye. *Models of Peano Arithmetic*, volume 15 of *Oxford Logic Guides*. Clarendon Press, Oxford, 1991.

[12] U. Kohlenbach. Effective bounds from ineffective proofs in analysis: an application of functional interpretation and majorization. *The Journal of Symbolic Logic*, 57:1239–1273, 1992.

[13] U. Kohlenbach. *Applied Proof Theory: Proof Interpretations and their Use in Mathematics*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2008.

[14] E. Nelson. Internal set theory: A new approach to nonstandard analysis. *Bulletin of the American Mathematical Society*, 83(6):1165–1198, 11 1977.

[15] J. R. Shoenfield. *Mathematical Logic*. Addison-Wesley Publishing Company, 1967. Republished in 2001 by AK Peters.

[16] S. G. Simpson. *Subsystems of Second Order Arithmetic*. Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1999.

[17] A. S. Troelstra (ed.). *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*, volume 344 of *Lecture Notes in Mathematics*. Springer-Verlag, Berlin, 1973.