A functional interpretation for nonstandard arithmetic

Benno van den Berg,[*] Eyvind Briseid,[†] and Pavol Safarik[‡]

July 12, 2012

Abstract

We introduce constructive and classical systems for nonstandard arithmetic and show how variants of the functional interpretations due to Gödel and Shoenfield can be used to rewrite proofs performed in these systems into standard ones. These functional interpretations show in particular that our nonstandard systems are conservative extensions of $\mathsf{E\text{-}HA}^\omega$ and $\mathsf{E\text{-}PA}^\omega$, strengthening earlier results by Moerdijk and Palmgren, and Avigad and Helzner. We will also indicate how our rewriting algorithm can be used for term extraction purposes. To conclude the paper, we will point out some open problems and directions for future research, including some initial results on saturation principles.

Contents

[*]Mathematisch Instituut, Universiteit Utrecht, PO. Box 80010, 3508 TA Utrecht. Email address: B.vandenBerg1@uu.nl. Supported by the Netherlands Organisation for Scientific Research (NWO).

[†]Department of Mathematics, The University of Oslo, Postboks 1053, Blindern, 0316 Oslo. Email address: eyvindbriseid@gmail.com. Supported by the Research Council of Norway (Project 204762/V30).

[‡]Fachbereich Mathematik, Technische Universität Darmstadt, Schloßgartenstraße 7, 64289 Darmstadt. Email address: pavol.safarik@googlemail.com. Supported by the German Science Foundation (DFG project KO 1737/5-1).

1 Introduction

In this paper we present functional interpretations for both constructive and classical systems of nonstandard arithmetic. The interpretations have two aspects: they show that the nonstandard systems are conservative over ordinary (standard) ones and they show how terms can be extracted from nonstandard proofs.

After inventing nonstandard analysis and showing that it was a versatile idea leading to nonstandard proofs in various areas of mathematics, Robinson suggested one could look at nonstandard arguments "syntactically" (proof-theoretically). From this point of view, one would see nonstandard analysis as "introduc[ing] *new deductive procedures* rather than new mathematical entities" [43]. Apparently he did not think that adding these new deductive procedures to standard systems would make them more powerful or stronger in a proof-theoretic sense. In fact, in [42] he formulates as a general problem *"to devise a purely syntactic transformation which correlates standard and nonstandard proofs of the same theorems in a large area, e.g., complex function theory"*. In [30] he is reported as asking a more specific question, whether a certain system for nonstandard arithmetic is conservative over PA. These ideas of Robinson have borne fruit in the work of Kreisel [30, 31], Friedman (unpublished), Nelson [38, 39], Moerdijk, Palmgren [36, 37, 41], Avigad [5, 2] and others, who proved for various systems for nonstandard analysis that they are conservative extensions of standard systems. Often their arguments are effective in that one could extract algorithms from their proofs which convert nonstandard arguments into standard ones.

As an example of this, let us have a short look at the work of Nelson, also because it is a major source of inspiration for this article. The idea of Nelson was to add a new unary predicate symbol st to ZFC for "being standard". In addition, he added three new axioms to ZFC governing the use of this new unary predicate, called Idealization, Standardization and Transfer. The resulting system he called IST, which stands for Internal Set Theory. The main logical result about IST is that it is a conservative extension of ZFC, so any theorem provable in IST which does not involve the st-predicate is provable in ZFC as well. Hence such theorems are genuine mathematical results, acceptable from the generally shared foundational standpoint of ZFC.

The conservativity of IST over ZFC was proved twice. In the original paper where he introduces Internal Set Theory [38] (recently reprinted with a foreword by G. F. Lawler in Volume 48, Number 4 of the *Bulletin of the American Mathematical Society* in recognition of its status as a classic), Nelson gives a model-theoretic argument which he attributes to Powell. In a later publication [39], he proves the same result syntactically by providing a "reduction algorithm" (a rewriting algorithm) for converting proofs performed in IST to ordinary ZFC-proofs. There is a remarkable similarity between his reduction algorithm and the Shoenfield interpretation [45]; this observation was the starting point for this paper.

We will work with systems in higher types, such as HA$^\omega$ and PA$^\omega$, rather than set theory, because, as we mentioned before, we will not just be interested in establishing conservation results, but also in extracting terms from nonstandard proofs and "proof mining". Proof mining is an area of applied logic in which one uses proof-theoretic techniques to extract quantitive information (such as bounds on the growth rate of certain functions) from proofs in ordinary mathematics. In addition, such techniques

can reveal certain uniformities leading to new qualitative results as well. Functional interpretations are one of the main tools in proof mining (for an introduction to this part of applied proof theory, see [25]). To extract interesting bounds, however, it is important that the mathematical arguments one analyses can be performed in sufficiently weak systems: therefore one considers systems such as HA^ω or PA^ω, or fragments thereof, rather than ZFC. The reason for considering systems in higher types (rather than PA, for instance) is not just because they are more expressive, but also because higher types are precisely what makes functional interpretations work.

Although establishing conservation and term extraction results for systems of non-standard arithmetic is what this paper is about, there is another way of looking at the results of this paper, which has more to do with the ideas of Lifschitz on calculable numbers [35], of Berger on uniform Heyting arithmetic [7] and of Hernest on the light Dialectica interpretation [19], than with nonstandard arithmetic. Their idea was to have two kinds of quantifiers, one with computational content and one without. On the realizability interpretation of Lifschitz, the computationally meaningful quantifiers are interpreted in the usual way (which, in the case of the existential quantifier, means that a realizer needs to exhibit a witness, while a realizer for a universal statement $\forall n\, \varphi(n)$ is a program which computes a realizer of $\varphi(n)$ from the value n). The computationally empty quantifiers, on the other hand, are to be interpreted uniformly (which means that it need not exhibit a witness in the existential case: a witness simply has to exist; while in the case of the universal quantifier it has to be a realizer which, uniformly, realizes $\varphi(n)$ for all n).

A new unary predicate st introduces two types of quantifiers as well: the *internal* quantifiers $\forall x$ and $\exists x$, as well as the *external* quantifiers $\forall^{\mathrm{st}} x$ and $\exists^{\mathrm{st}} x$ (which can be seen as abbreviations of $\forall x\, (\,\mathrm{st}(x) \to \ldots)$ and $\exists x\, (\,\mathrm{st}(x) \wedge \ldots)$ respectively). Our initial idea was to interpret the former uniformly, while interpreting the latter in the usual way, in complete analogy with the ideas of Lifschitz. But this led to several, to us, undesirable effects (in particular, one could not realize the statement that the standard natural numbers are closed downwards). Our solution was to weaken the computational meaning of the external quantifiers: in particular, to realize $\exists^{\mathrm{st}} n\, \varphi(n)$ it suffices to exhibit a finite lists of candidates $n_1, \ldots, n_k$ such that at least one of the statements $\varphi(n_i)$ is realized. Because of the analogy with Herbrand disjunctions from proof theory, we have dubbed this type of realizability "Herbrand realizability".

To make Herbrand realizability work in a higher type setting, it is convenient to work in an extension of HA^ω with types for finite sequences. More precisely, we will assume that there is a type σ^* for sequences of objects of type σ. The type σ^* carries the structure of a preorder (with $x \preceq y$ if every element in the list coded by x also occurs in the list coded by y) and one would naturally expect realizers to be closed upwards with respect to this preorder. To make this work nicely, it will be useful to introduce a new kind of application (of functions to arguments) which is monotone in the first component. This can be done and with this additional ingredient Herbrand realizability can be defined. We will do this in Section 4.

It has to be admitted that the connection of Herbrand realizability to nonstandard arithmetic is rather tangential and, as a matter of fact, we will not be very interested in Herbrand realizability *per se*. This will change radically if we turn to the functional interpretation introduced in Section 5, which bears the same relation to Herbrand realizability as the usual Dialectica interpretation does to modified realizability. It turns out that if one defines a Dialectica-type functional interpretation using the new application, with implication interpreted *à la* Diller–Nahm [9], basing it on some of the characteristic principles of Herbrand realizability, and the idea of having Herbrand disjunctions realize existential statements, this interpretation will, almost by magic, interpret and eliminate principles recognizable from nonstandard analysis. Our main reason for including Herbrand realizability is to have an easy point of access to and to provide some intuition for this functional interpretation.

So far the techniques we mentioned work only for constructive systems. But by

combining the functional interpretation we mentioned above with negative translation, we are able to define a Shoenfield-type functional interpretation for classical nonstandard systems as well. In this way we also obtain conservation and term extraction results for classical systems. We will work out the technical details in Sections 6 and 7 below.

The resulting functional interpretations have some striking similarities with the bounded functional interpretations introduced by Ferreira and Oliva in [12] and [10] (see also [13]). In the same way, Herbrand realizability seems related to the bounded modified realizability interpretation due to Ferreira and Nunes (for which, see [11]). We will briefly comment on this in Section 5 below.

The contents of this paper are therefore as follows. In Section 2 we will introduce an intuitionistic base system for our investigations into constructive nonstandard arithmetic. In Section 3, we will discuss some principles from nonstandard analysis and their relations. This will give one some ideas of how interpretations of nonstandard systems have to look like and will provide us with some "benchmarks" with which one can measure the success of an interpretation. In Section 4 we will introduce Herbrand realizability and discuss its merits as an interpretation of nonstandard arithmetic. Our nonstandard functional interpretation will be introduced in Section 5 and we will use it to prove several conservation and term extraction results in an intuitionistic context. In Section 6 we introduce a classical base system for nonstandard arithmetic and discuss two variants of the negative translation. These we will use in Section 7 to obtain a Shoenfield-type functional interpretation and derive conservation and term extraction results in a classical context. Finally, Section 8 discusses work in progress on saturation principles and other plans for future work.

2 Formalities

In this section, we introduce our base systems for investigating nonstandard arithmetic.

2.1 The system $\mathsf{E\text{-}HA}^{\omega*}$

In this paper, $\mathsf{E\text{-}HA}^{\omega*}$ will be the extension of the system called $\mathsf{E\text{-}HA}^\omega_0$ in [48] and $\mathsf{E\text{-}HA}^\omega_\to$ in [49] with types for finite sequences. More precisely, the collection of types $\boldsymbol{T}^*$ will be smallest set closed under the following rules:

(i) $0 \in \boldsymbol{T}^*$;

(ii) $\sigma, \tau \in \boldsymbol{T}^* \Rightarrow (\sigma \to \tau) \in \boldsymbol{T}^*$;

(iii) $\sigma \in \boldsymbol{T}^* \Rightarrow \sigma^* \in \boldsymbol{T}^*$.

Because we have not included product types in $\boldsymbol{T}^*$, we will often be handling tuples of types or terms. We will always refer to such lists of types and terms as tuples and never as sequences, so as not to confuse them with terms of sequence type (i.e., of type σ^* for some $\sigma \in \boldsymbol{T}^*$). In dealing with tuples, we will follow the notation and conventions of [48] and [25]. In particular, if $\underline{x} = x_0, \ldots, x_{m-1}$ and $\underline{y} = y_0, \ldots, y_{n-1}$, then

1. $[]$ stands for the empty tuple, while $\underline{x}, \underline{y}$ stands for $x_0, \ldots, x_{m-1}, y_0, \ldots, y_{n-1}$;

2. $x_i \underline{y}$ stands for $(\ldots((x_i y_0)y_1)\ldots)y_{n-1}$, while $\underline{x}y$ stands for $x_0 y, \ldots, x_{m-1} y$ (and never for $x_0, \ldots, x_{m-1}\underline{y}$);

3. $\lambda \underline{x}.\underline{y}$ stands for $\lambda \underline{x}.y_0, \ldots, \lambda \underline{x}.y_{n-1}$;

4. and, finally, if $\underline{x} = x_0^{\sigma_0}, \ldots, x_{m-1}^{\sigma_{m-1}}$ and $\underline{y} = y_0^{\sigma_0}, \ldots, y_{m-1}^{\sigma_{m-1}}$ are tuples having the same length and types, we will write $\underline{x} =_{\underline{\sigma}} \underline{y}$ for

$$\bigwedge_{j=0}^{m-1} x_j =_{\sigma_j} y_j.$$

Because we have included sequence types, we will have to enrich the term language (Gödel's $\mathcal{T}$); it now also includes a constant $\langle\rangle_\sigma$ of type σ^* and an operation c of type $\sigma \to (\sigma^* \to \sigma^*)$ (for the empty sequence and the operation of prepending an element to a sequence, respectively), as well as a list recursor $\underline{L}_{\sigma,\rho}$ satisfying the following axioms:

$$\begin{aligned} \underline{L}_{\sigma,\rho} \, \langle\rangle_\sigma \underline{y}\underline{z} &=_{\underline{\rho}} & \underline{y}, \\ \underline{L}_{\sigma,\rho} \, c(a,x)\underline{y}\underline{z} &=_{\underline{\rho}} & \underline{z}(\underline{L}_{\sigma,\rho} \, x\underline{y}\underline{z})a, \end{aligned}$$

where $\underline{\rho} = \rho_1, \ldots, \rho_k$ is a k-tuple of types, $\underline{y} = y_1, \ldots, y_k$ is a k-tuple of terms with y_i of type ρ_i and $\underline{z} = z_1, \ldots, z_k$ is a k-tuple of terms with z_i of type $\rho_1 \to \ldots \to \rho_k \to \sigma \to \rho_i$ (compare [49, p. 456] or [25, p. 48]). In addition, we have the recursors and combinators for all the new types in Gödel's $\mathcal{T}$, satisfying the usual equations. The resulting extension we will denote by $\mathcal{T}^*$.

We will have a primitive notion of equality at every type and equality axioms expressing that equality is a congruence (as in [49, p. 448-9]). Since decidability of quantifier-free formulas is not essential for this paper, this choice will not create any difficulties. In addition, we assume the axiom of extensionality for functions:

$$f =_{\sigma\to\tau} g \leftrightarrow \forall x^\sigma \, fx =_\tau gx.$$

Of course, the underlying logic of $\mathsf{E\text{-}HA}^{\omega*}$ is constructive; we will assume it is axiomatized as in [25, p. 42]. We also have all the usual axioms of $\mathsf{E\text{-}HA}^\omega$, as in [25, p. 48-9], for example, where it is to be understood that the induction axiom applies to all formulas in the language (i.e., also those containing variables of sequence type and the new terms that belong to $\mathcal{T}^*$). Finally, we add the following sequence axiom:

$$\mathsf{SA}: \quad \forall y^{\sigma^*} \, (\, y = \langle\rangle_\sigma \vee \exists a^\sigma, x^{\sigma^*} \, y = c(a,x) \,).$$

In normal $\mathsf{E\text{-}HA}^\omega$, as in [25] or [49], for example, one can also talk about sequences, but these have to be coded up (see [25, p. 59]). As a result, $\mathsf{E\text{-}HA}^{\omega*}$ is a definitional extension of, and hence conservative over, $\mathsf{E\text{-}HA}^\omega$ as defined in [25] or [49].

2.2 The system $\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*}$

Definition 2.1. The language of the system $\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*}$ is obtained by extending that of $\mathsf{E\text{-}HA}^{\omega*}$ with unary predicates st^σ as well as two new quantifiers $\forall^{\mathrm{st}}x^\sigma$ and $\exists^{\mathrm{st}}x^\sigma$ for every type $\sigma \in \boldsymbol{T}^*$. Formulas in the language of $\mathsf{E\text{-}HA}^{\omega*}$ (i.e., those that do not contain the new predicate st_σ or the two new quantifiers $\forall^{\mathrm{st}}x^\sigma$ and $\exists^{\mathrm{st}}x^\sigma$) will be called *internal*. Formulas which are not internal will be called *external*.

We will adopt the following

> IMPORTANT CONVENTION: We follow Nelson [39] in using small Greek letters to denote internal formulas and capital Greek letters to denote formulas which can be external.

Definition 2.2 ($\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*}$)**.** The system $\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*}$ is obtained by adding to $\mathsf{E\text{-}HA}^{\omega*}$ the axioms $\mathsf{EQ}, \mathcal{T}_{\mathrm{st}}^*$ and $\mathsf{IA}^{\mathrm{st}}$, where

- EQ stands for the defining axioms of the external quantifiers:

$$\forall^{\mathrm{st}}x\,\Phi(x) \quad \leftrightarrow \quad \forall x\,(\,\mathrm{st}(x) \to \Phi(x)\,),$$
$$\exists^{\mathrm{st}}x\,\Phi(x) \quad \leftrightarrow \quad \exists x\,(\,\mathrm{st}(x) \wedge \Phi(x)\,),$$

  with $\Phi(x)$ an arbitrary formula, possibly with additional free variables.

- $\mathcal{T}^*_{\mathrm{st}}$ consists of:

    1. the axioms $\mathrm{st}(x) \wedge x = y \to \mathrm{st}(y)$,
    2. the axiom $\mathrm{st}(t)$ for each closed term t in $\mathcal{T}^*$,
    3. the axioms $\mathrm{st}(f) \wedge \mathrm{st}(x) \to \mathrm{st}(fx)$.

- $\mathsf{IA}^{\mathrm{st}}$ is the external induction axiom:

$$\mathsf{IA}^{\mathrm{st}} \quad : \quad \big(\Phi(0) \wedge \forall^{\mathrm{st}}n^0(\Phi(n) \to \Phi(n+1))\big) \to \forall^{\mathrm{st}}n^0\Phi(n),$$

  where $\Phi(n)$ is an arbitrary formula, possibly with additional free variables.

Here it is to be understood that in $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ the laws of intuitionistic logic apply to all formulas, while the induction axiom from $\mathsf{E\text{-}HA}^{\omega*}$

$$\big(\varphi(0) \wedge \forall n^0(\varphi(n) \to \varphi(n+1))\big) \to \forall n^0\varphi(n)$$

applies to internal formulas φ only.

Lemma 2.3. $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} \vdash \Phi(x) \wedge x = y \to \Phi(y)$ *for every formula* Φ.

Proof. By induction on the logical structure of Φ. Note that the st-predicate is extensional by $\mathcal{T}^*_{\mathrm{st}}$ and the case of the external quantifiers $\forall^{\mathrm{st}}$ and $\exists^{\mathrm{st}}$ can be reduced to that of the internal quantifiers $\forall$ and $\exists$ by using the EQ-axiom. $\qquad\square$

Lemma 2.4. $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} \vdash \mathrm{st}^0(x) \wedge y \leq x \to \mathrm{st}^0(y)$.

Proof. Apply external induction to the formula $\Phi(x) :\equiv \forall y\,(\,y \leq x \to \mathrm{st}(y))$. $\qquad\square$

Remark 2.5. The previous lemma implies that, *whenever n is a standard natural number*, a bounded internal quantifier of the form $\exists^{\mathrm{st}}i \leq n$ can always be replaced by $\exists i \leq n$ and vice versa (the same applies to the universal quantifiers, of course). So we can regard such bounded quantifiers as internal or external, depending on what suits us best. Most often, however, it will be convenient to regard them as internal quantifiers.

Definition 2.6. For any formula Φ in the language of $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$, we define its *internalization* Φ^{int} to be the formula one obtains from Φ by replacing $\mathrm{st}(x)$ by $x = x$, and $\forall^{\mathrm{st}}x$ and $\exists^{\mathrm{st}}x$ by $\forall x$ and $\exists x$, respectively.

One of the reasons $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ is such a convenient system for our proof-theoretic investigations is because we have the following easy result:

Proposition 2.7. *If a formula Φ is provable in* $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$*, then its internalization Φ^{int} is provable in* $\mathsf{E\text{-}HA}^{\omega*}$*. Hence* $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ *is a conservative extension of* $\mathsf{E\text{-}HA}^{\omega*}$ *and* $\mathsf{E\text{-}HA}^{\omega}$*.*

Proof. Clear, because the internalizations of the axioms of $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ are provable in $\mathsf{E\text{-}HA}^{\omega*}$. $\qquad\square$

2.3 Operations on finite sequences

Using the list recursor $L_{\sigma,\rho}$ one can define a length function $|\cdot| : \sigma^* \to 0$ satisfying

$$
\begin{aligned}
|\langle\rangle_\sigma| &= 0, \\
|c(a,x)| &= S|x|.
\end{aligned}
$$

Moreover, we can fix for every type σ a term $\mathcal{O}_\sigma$ in $\mathcal{T}^*$ of that type. One can then also define a projection function $\sigma^* \to (0 \to \sigma)$; we will write $(x)_i$ for the ith projection of x. It satisfies:

$$
\begin{aligned}
(\langle\rangle_\sigma)_n &= \mathcal{O}_\sigma, \\
(c(a,x))_0 &= a, \\
(c(a,x))_{Sn} &= (x)_n.
\end{aligned}
$$

In addition, we will have an operation which given $x_0, \ldots, x_{n-1}$ of type σ builds an object $x = \langle x_0, \ldots, x_{n-1} \rangle$ of type σ^* for which we have $|x| = n$ and

$$
\begin{aligned}
(x)_i &= x_i \quad \text{if } i < |x|, \\
(x)_i &= \mathcal{O}_\sigma \quad \text{otherwise.}
\end{aligned}
$$

We will also need a concatenation operation $*_{\sigma^*} : \sigma^* \to (\sigma^* \to \sigma^*)$ defined by:

$$
\begin{aligned}
\langle\rangle_\sigma *_{\sigma^*} y &= y, \\
c(a,x) *_{\sigma^*} y &= c(a, x * y).
\end{aligned}
$$

This we can use to define an n-fold concatenation: If $F : 0 \to \sigma^*$ and n is of type 0, then we can set:

$$
(F(0) *_{\sigma^*} \ldots *_{\sigma^*} F(n-1)) = \begin{cases} \langle\rangle_\sigma & \text{if } n = 0, \\ (F(0) *_{\sigma^*} \ldots *_{\sigma^*} F(n-2)) *_{\sigma^*} F(n-1) & \text{if } n > 0. \end{cases}
$$

Note that $F(0) *_{\sigma^*} \ldots *_{\sigma^*} F(n-1) = F(0)$ if $n = 1$.

Lemma 2.8.   *1.* $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} \vdash \mathrm{st}(x^{\sigma^*}) \to \mathrm{st}(|x|)$,

    2. $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} \vdash \mathrm{st}(x^{\sigma^*}) \to \mathrm{st}((x)_i)$,

    3. $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} \vdash \mathrm{st}(x_0^\sigma) \wedge \ldots \wedge \mathrm{st}(x_n^\sigma) \to \mathrm{st}(\langle x_0^\sigma, \ldots, x_n^\sigma \rangle)$,

    4. $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} \vdash \mathrm{st}(x^{\sigma^*}) \wedge \mathrm{st}(y^{\sigma^*}) \to \mathrm{st}(x *_\sigma y)$.

    5. $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} \vdash \mathrm{st}(F^{0 \to \sigma^*}) \wedge \mathrm{st}(n^0) \to \mathrm{st}(F(0) * \ldots * F(n-1))$.

Proof.   Follows from the $\mathcal{T}^*_{\mathrm{st}}$-axioms together with the fact that the list recursor L belongs to $\mathcal{T}^*$. $\qquad\square$

Notation 2.9.   1. Given $\underline{x} = x_0^{\sigma_0}, \ldots, x_{m-1}^{\sigma_{m-1}}$ and $\underline{i} = i_0^0, \ldots, i_{m-1}^0$ we will write $|\underline{x}|$ for $|x_0^{\sigma_0}|, \ldots, |x_{m-1}^{\sigma_{m-1}}|$ and $(\underline{x})_{\underline{i}}$ for $(x_0^{\sigma_0})_{i_0}, \ldots, (x_{m-1}^{\sigma_{m-1}})_{i_{m-1}}$.

    2. Given $\underline{x} = x_0^{\sigma_0}, \ldots, x_{m-1}^{\sigma_{m-1}}$ and $\underline{y} = y_0^{\sigma_0}, \ldots, y_{m-1}^{\sigma_{m-1}}$, we will write $\underline{\langle x \rangle}$ for $\langle x_0^{\sigma_0} \rangle, \ldots, \langle x_{m-1}^{\sigma_{m-1}} \rangle$, and $\underline{\langle x, y \rangle}$ for $\langle x_0, y_0 \rangle, \ldots, \langle x_{m-1}, y_{m-1} \rangle$.

    3. Given $\underline{x} = x_0^{\sigma_0^*}, \ldots, x_{m-1}^{\sigma_{m-1}^*}$ and $\underline{y} = y_0^{\sigma_0^*}, \ldots, y_{m-1}^{\sigma_{m-1}^*}$, we will write $\underline{x} *_{\underline{\sigma}^*} \underline{y}$ for

$$
x_0 *_{\sigma_0^*} y_0, \ldots, x_{m-1} *_{\sigma_{m-1}^*} y_{m-1}.
$$

2.4 Finite sets

Most of the time, we will regard finite sequences as stand-ins for finite sets. In fact, we will need the notion of an element and that of one sequence being contained in another, as given in the definitions below.

Definition 2.10. For s^σ, t^{σ^*} we write $s \in_\sigma t$ and say that s *is an element of* t if

$$\exists i < |t|(\, s =_\sigma (t)_i \,).$$

For $\underline{s}^{\underline{\sigma}} = s_0^{\sigma_0}, \ldots, s_{n-1}^{\sigma_{n-1}}$ and $\underline{t}^{\underline{\sigma}^*} = t_0^{\sigma_0^*}, \ldots, t_{n-1}^{\sigma_{n-1}^*}$ we write $\underline{s} \in_{\underline{\sigma}} \underline{t}$ and say that $\underline{s}$ *is an element of* $\underline{t}$ if

$$\bigwedge_{k=0}^{n-1} s_k \in_{\sigma_k} t_k.$$

In case no confusion can arise, we will drop the subscript and write simply $\in$ instead of $\in_\sigma$ or $\in_{\sigma^*}$.

Lemma 2.11. $\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*} \vdash \mathrm{st}(x^{\sigma^*}) \wedge y \in_\sigma x \to \mathrm{st}(y^\sigma)$.

Proof. Follows from Lemma 2.8.2 and the extensionality of the st-predicate (first part of the $\mathcal{T}_{\mathrm{st}}^*$-axiom). $\qquad\square$

Definition 2.12. For $s^{\sigma^*}, t^{\sigma^*}$ we write $s \preceq_\sigma t$ and say that s *is contained in* t if

$$\forall x^\sigma \,(\, x \in s \to x \in t\,),$$

or, equivalently,

$$\forall i < |s| \,\exists j < |t| \,(s)_i =_\sigma (t)_j.$$

For $\underline{s}^{\underline{\sigma}^*} = s_0^{\sigma_0^*}, \ldots, s_{n-1}^{\sigma_{n-1}^*}$ and $\underline{t}^{\underline{\sigma}^*} = t_0^{\sigma_0^*}, \ldots, t_{n-1}^{\sigma_{n-1}^*}$ we write $\underline{s} \preceq_{\underline{\sigma}} \underline{t}$ and say that $\underline{s}$ *is contained in* $\underline{t}$ if

$$\bigwedge_{k=0}^{n-1} s_k \preceq_{\sigma_k} t_k.$$

Lemma 2.13. $\mathsf{E\text{-}HA}^{\omega*}$ *proves that* $\preceq_\sigma$ *determines a preorder on the set of objects of type* σ^*. *More precisely, for all* x^{σ^*} *we have* $x \preceq_\sigma x$, *and for all* $x^{\sigma^*}, y^{\sigma^*}, z^{\sigma^*}$ *with* $x \preceq_\sigma y$ *and* $y \preceq_\sigma z$, *we have* $x \preceq_\sigma z$.

Proof. Obvious. $\qquad\square$

In relation to this ordering the notion of a property being *upwards closed* in a variable x will be of importance.

Definition 2.14. A property $\Phi(\underline{x}^{\underline{\sigma}^*})$ is called *upwards closed in* $\underline{x}$ if $\Phi(\underline{x}) \wedge \underline{x} \preceq \underline{y} \to \Phi(\underline{y})$ and *downwards closed in* $\underline{x}$ if $\Phi(\underline{x}) \wedge \underline{y} \preceq \underline{x} \to \Phi(\underline{y})$.

2.5 Induction and extensionality for sequences

The aim of this subsection is to prove induction and extensionality principles for sequences. It all relies on the following lemma:

Lemma 2.15. 1. $\mathsf{E\text{-}HA}^{\omega*} \vdash \forall x^{\sigma^*} (\, |x| = 0 \leftrightarrow x = \langle\rangle_\sigma \,)$.

2. $\mathsf{E\text{-}HA}^{\omega*} \vdash \forall n^0, x^{\sigma^*} (\, |x| = Sn \leftrightarrow \exists a, y^{\sigma^*} (\, x = c(a, y) \wedge |y| = n \,) \,)$.

Proof. The right-to-left directions hold by definition of the length function $|\cdot|$. So suppose we have an element x of type σ^*. Then, by the sequence axiom SA, either $x = \langle\rangle_\sigma$ or there are a^σ, y^{σ^*} such that $x = c(a, y)$. In the latter case, $|x| = S|y| > 0$, so if $|x| = 0$, then $x = \langle\rangle_\sigma$. But if $|x| = Sn$, then $x \neq \langle\rangle_\sigma$ and there are a, y with $x = c(a, y)$ and $|y| = n$. $\qquad\square$

Proposition 2.16. E-HA$^{\omega*}$ *proves the induction schema for sequences:*

$$\varphi(\langle\rangle_\sigma) \wedge \forall a^\sigma, y^{\sigma^*} \, (\, \varphi(y) \to \varphi(c(a,y)) \,) \to \forall x^{\sigma^*} \, \varphi(x).$$

Proof. Suppose $\varphi(\langle\rangle_\sigma)$ and $\forall a^\sigma, y^{\sigma^*} \, (\, \varphi(y) \to \varphi(c(a,y)) \,)$. It now follows from the previous lemma that we can prove

$$\forall n^0 \, \forall x^{\sigma^*} \, (\, |x| = n \to \varphi(x) \,)$$

by ordinary induction. $\qquad\square$

One consequence is the following useful fact:

Lemma 2.17. E-HA$^{\omega*}$ *proves that for any two elements $x^{\sigma^*}, y^{\sigma^*}$ we have $|x*y| = |x|+|y|$ and*
$$
\begin{array}{rcll}
(x * y)_i & = & (x)_i & \text{if } i < |x|, \\
(x * y)_i & = & (y)_{i-|x|} & \text{otherwise.}
\end{array}
$$

Therefore it also proves that $x \preceq_\sigma x * y$ and $y \preceq_\sigma x * y$.

Proof. Easy argument using the recursive definitions of $|\cdot|$ and $*$ and the induction schema for sequences. $\qquad\square$

Another consequence is the principle of extensionality for sequences. We will call two elements $x^{\sigma^*}, y^{\sigma^*}$ extensionally equal, and write $x =_{e,\sigma*} y$, if

$$|x| =_0 |y| \wedge \forall i < |x| \, (\, (x)_i =_\sigma (y)_i \,).$$

Proposition 2.18. E-HA$^{\omega*}$ *proves*

$$\forall x^{\sigma^*}, y^{\sigma^*} \, (\, x =_{e,\sigma*} y \to x =_{\sigma*} y \,).$$

Proof. Proof by sequence induction on x.

If $x =_e y$ and $x = \langle\rangle_\sigma$, then $|y| = |x| = 0$. So $y = \langle\rangle_\sigma$.

If $x =_e y$ and $x = c(a, x')$, then $|x| = Sn$ where $n = |x'|$. So also $|y| = Sn$ and hence $y = c(b, y')$ for some b^σ, y'^{σ^*} with $|y'| = n$. Since $x =_e y$, we have $a = b$ and $x' =_e y'$. From the latter we get $x' = y'$ by induction hypothesis, so $x = c(a, x') = c(b, y') = y$. $\qquad\square$

Corollary 2.19. E-HA$^{\omega*}_{\text{st}}$ *proves*

$$\forall x^{\sigma^*} \, \text{st}(|x|) \wedge \forall i < |x| \, \text{st}((x)_i) \to \text{st}(x).$$

Proof. Suppose x^{σ^*} is a sequence of standard length and all components $(x)_i$ are standard. Then $x' := \langle x_0, \ldots, x_{|x|-1}\rangle$ is also standard (by Lemma 2.8). But $x =_{e,\sigma*} x'$, so $x = x'$ by extensionality for sequences and $\text{st}(x)$ by extensionality of the standardness predicate. $\qquad\square$

Corollary 2.20. E-HA$^{\omega*}_{\text{st}}$ *proves the external induction axiom for sequences:*

$$\Phi(\langle\rangle_\sigma) \wedge \forall^{\text{st}} a^\sigma, y^{\sigma^*} \, (\, \Phi(y) \to \Phi(c(a,y)) \,) \to \forall^{\text{st}} x^{\sigma^*} \, \Phi(x).$$

Proof. Suppose $\Phi(\langle\rangle_\sigma)$ and $\forall^{\text{st}} a^\sigma, y^{\sigma^*} \, (\, \Phi(y) \to \Phi(c(a,y)) \,)$. The idea now is to prove

$$\forall^{\text{st}} n^0 \, \forall^{\text{st}} x^{\sigma^*} \, (\, |x| = n \to \Phi(x) \,)$$

by external induction IA^{st}, using the previous corollary to argue that if $x = c(a, y)$ and x^{σ^*} is standard, then both a and y are standard as well. $\qquad\square$

2.6 Finite sequence application

The following operations will be crucial for what follows.

Definition 2.21 (Finite sequence application and abstraction). If s is of type $(\sigma \to \tau^*)^*$ and t is of type σ, then

$$s[t] := (s)_0(t) * \ldots * (s)_{|s|-1}(t) : \tau^*.$$

For every term s of type $\sigma \to \tau^*$ we set

$$\Lambda x^\sigma . s(x) := \langle \lambda x^\sigma . s(x) \rangle : (\sigma \to \tau^*)^*.$$

The point is that these two operations act like application and abstraction, for we have

$$(\Lambda x.s(x))[t] =_{\tau^*} (\lambda x.s(x))(t) =_{\tau^*} s(t).$$

We will often write $\underline{s}[\underline{t}]$ and $\Lambda \underline{x}.\underline{t}$; in handling these expressions, the same conventions as for ordinary application and abstraction apply (see Section 2.1).

Note that the defining equations for the sequence application and Λ-abstraction imply that finite sequence application and ordinary application are (provably) interdefinable, in the following sense: E-HA$^{\omega *}$ proves that for every $s : (\sigma \to \tau^*)^*$ there is a $t : \sigma \to \tau^*$ (viz., $t = \lambda x.s[x]$) such that $s[x] = t(x)$ for all x, as well as that for every $t : \sigma \to \tau^*$ there is an s of type $(\sigma \to \tau^*)^*$ (viz., $s = \Lambda x.t(x)$) such that $s[x] = t(x)$ for all x.

In what follows we will need that one can define recursors $\underline{\mathcal{R}}_\rho$ for each tuple of types $\underline{\rho}^* = \rho_0^*, \ldots, \rho_k^*$, such that

$$
\begin{aligned}
\underline{\mathcal{R}}_\rho(0, \underline{y}, \underline{z}) &=_{\underline{\rho}^*} \quad \underline{y}, \\
\underline{\mathcal{R}}_\rho(n+1, \underline{y}, \underline{z}) &=_{\underline{\rho}^*} \quad \underline{z}[n, \underline{\mathcal{R}}_\rho(n, \underline{y}, \underline{z})],
\end{aligned}
$$

(where y_i is of type ρ_i^* and z_i is of type $(0 \to \rho_0^* \to \ldots \to \rho_k^* \to \rho_i^*)^*$). Indeed, by letting

$$\underline{\mathcal{R}}_\rho := \lambda n^0, \underline{y}, \underline{z}.\underline{R}_{\underline{\rho}^*}(n, \underline{y}, (\lambda \underline{s}^{\underline{\rho}^*}, t^0.\underline{z}[t, \underline{s}])),$$

where $\underline{R}_\rho$ are constants for simultaneous primitive recursion as in [25], we get

$$\underline{\mathcal{R}}_\rho(0, \underline{y}, \underline{z}) =_{\underline{\rho}^*} \underline{R}_{\underline{\rho}^*}(0, \underline{y}, (\lambda \underline{s}^{\underline{\rho}^*}, t^0.\underline{z}[t, \underline{s}])) =_{\underline{\rho}^*} \underline{y}$$

and

$$
\begin{aligned}
\underline{\mathcal{R}}_\rho(n+1, \underline{y}, \underline{z}) &=_{\underline{\rho}^*} \quad \underline{R}_{\underline{\rho}^*}(n+1, \underline{y}, (\lambda \underline{s}^{\underline{\rho}^*}, t^0.\underline{z}[t, \underline{s}])) \\
&=_{\underline{\rho}^*} \quad (\lambda \underline{s}^{\underline{\rho}^*}, t^0.\underline{z}[t, \underline{s}])(\underline{R}_{\underline{\rho}^*}(n, \underline{y}, (\lambda \underline{s}^{\underline{\rho}^*}, t^0.\underline{z}[t, \underline{s}])), n) \\
&=_{\underline{\rho}^*} \quad \underline{z}[n, \underline{R}_{\underline{\rho}^*}(n, \underline{y}, (\lambda \underline{s}^{\underline{\rho}^*}, t^0.\underline{z}[t, \underline{s}]))] \\
&=_{\underline{\rho}^*} \quad \underline{z}[n, \underline{\mathcal{R}}_\rho(n, \underline{y}, \underline{z})].
\end{aligned}
$$

Notice that when compared to the case of the ordinary primitive recursors $\underline{R}_\rho$ we have switched the order of the arguments of z. This is simply to make the realizer for the interpretation of the induction schema nicer.

With respect to the preorder $\preceq$ from Definition 2.12 the new application is monotone in the first component, in the following sense:

Lemma 2.22. E-HA$^{\omega *}$ *proves*

1. *If $s^{(\sigma \to \tau^*)^*} \preceq \tilde{s}^{(\sigma \to \tau^*)^*}$, then $s[t] \preceq \tilde{s}[t]$, for all t^σ.*

2. *If $\underline{s} \preceq \tilde{s}$, then $\underline{s}[\underline{t}] \preceq \tilde{s}[\underline{t}]$ for all $\underline{t}$ of suitable types.*

3. If $\underline{s} \preceq \tilde{\underline{s}}$, then $\underline{s}[\underline{t}] \preceq \tilde{\underline{s}}[\underline{t}]$ for all $\underline{t}$ of suitable types.

Proof. We will only prove the first point, as the other two are similar. Let $i < |s[t]|$, and consider $(s[t])_i$. Since $s[t] =_{\tau^*} (s)_0(t) * \ldots * (s)_{|s|-1}(t)$ there is $k < |s|$ and $m < |(s)_k(t)|$ such that

$$(s[t])_i =_{\tau^*} ((s)_k(t))_m.$$

Since $s \preceq \tilde{s}$ there is $j < |\tilde{s}|$ such that $(s)_k =_{\sigma \to \tau^*} (\tilde{s})_j$. Thus $m < |(\tilde{s})_j(t)|$ and

$$(s[t])_i =_{\tau^*} ((\tilde{s})_j(t))_m,$$

and so since $\tilde{s}[t] =_{\tau^*} (\tilde{s})_0(t) * \ldots * (\tilde{s})_{|\tilde{s}|-1}(t)$ there is some $n < |\tilde{s}[t]|$ such that

$$(\tilde{s}[t])_n =_{\tau^*} ((\tilde{s})_j(t))_m =_{\tau^*} (s[t])_i.$$

$\square$

Lemma 2.23. $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ *proves*

$$\mathrm{st}^{(\sigma \to \tau^*)^*}(x) \wedge \mathrm{st}^\sigma(y) \to \mathrm{st}^{\tau^*}(x[y])$$

and

$$\mathrm{st}^{\sigma \to \tau^*}(s) \to \mathrm{st}^{(\sigma \to \tau^*)^*}(\Lambda x^\sigma.s(x)).$$

Proof. Follows from Lemma 2.8. $\square$

3 Nonstandard principles

Semantic approaches to nonstandard analysis exploit the existence of nonstandard models of the first-order theory of the natural numbers or the reals. In fact, one may use the compactness theorem for first-order logic or the existence of suitable nonprincipal ultrafilters to show that there are extensions of the natural numbers, the reals or any other first-order structure one might be interested in, that are *elementary*: that is, satisfy the same first-order sentences, even when allowing for parameters from the original structure. For the natural numbers, for instance, this means that there are structures $^*\mathbb{N}$ and embeddings $i : \mathbb{N} \to {^*\mathbb{N}}$ that satisfy

$$^*\mathbb{N} \models \varphi(i(n_0), \ldots, i(n_k)) \iff \mathbb{N} \models \varphi(n_0, \ldots, n_k)$$

for all first-order formulas $\varphi(x_0, \ldots, x_k)$ and natural numbers $n_0, \ldots, n_k$. Usually, one identifies the elements in the image of i with the natural numbers and calls these the *standard* natural numbers, while those that do not lie in the image of i are the *non-standard* natural numbers. Sometimes, one adds a new predicate st to the structure $^*\mathbb{N}$, which is true only of the standard natural numbers. One can then use the elementarity of the embedding to show that $^*\mathbb{N}$ is still a linear order in which the nonstandard natural numbers must be infinite (i.e., bigger than any standard natural number). The charm and power of nonstandard proofs is that one can use these infinite natural numbers to prove theorems in the nonstandard structure $^*\mathbb{N}$, which must then be true in $\mathbb{N}$ as well, as the embedding i is elementary. The same applies to nonstandard extensions $^*\mathbb{R}$ of the reals, in which there are besides infinite reals, also infinitesimals (nonstandard reals having an absolute value smaller than any positive standard real): these infinitesimals can then be used to prove theorems in analysis in $^*\mathbb{R}$; again, one can then go on to use the elementarity of the embedding to show that they must hold in $\mathbb{R}$ as well. The catch is that only *first-order, internal* statements can be lifted in this way: so using nonstandard models requires some understanding about what can and what can not be expressed in first-order logic as well as some careful verifications as to whether formulas are internal.

Besides creating an interesting world in which there are infinite natural numbers and infinitesimals, nonstandard analysis also comes with some new proof principles, among which the following are the most important:

1. Overspill: if $\varphi(x)$ is internal and holds for all standard x, then $\varphi(x)$ also holds for some nonstandard x.

2. Underspill: if $\varphi(x)$ is internal and holds for all nonstandard x, then $\varphi(x)$ also holds for some standard x.

3. Transfer: an internal formula φ (possibly with standard parameters) holds in $^*\mathbb{N}$ iff it holds in $\mathbb{N}$.

Of course, transfer expresses the elementarity of the embedding. The other two principles are consequences of the fact that it is impossible to define standardness in $^*\mathbb{N}$ using an internal formula: for if one could, then $\varphi(n)$ would hold in $^*\mathbb{N}$, and hence in $\mathbb{N}$, for every natural number n. This would imply that $\forall x\, \varphi(x)$ holds in $\mathbb{N}$ and hence in $^*\mathbb{N}$ as well; but that contradicts the existence of nonstandard elements in $^*\mathbb{N}$.

In the remainder of this section, we will discuss these principles in more detail, for two related reasons. First of all, they will provide us with three benchmarks with which we will be able to measure the success of the different interpretations. Also, because they have some nontrivial consequences (especially in the intuitionistic context), discussing these will give us some important clues as to how any interpretation of nonstandard analysis will have to look like.

Remark 3.1. Unless we state otherwise, the principles we will subsequently introduce in this paper may have additional parameters besides those explicitly shown. Also recall that we follow Nelson's convention in using small Greek letters to denote internal formulas and capital Greek letters to denote formulas which can be external.

3.1 Overspill

When formalised in $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$, overspill (in type 0) is the following statement:

$$\mathsf{OS}_0 : \forall^{\mathrm{st}} x^0\, \varphi(x) \to \exists x^0\, (\,\neg\,\mathrm{st}(x) \wedge \varphi(x)\,).$$

Proposition 3.2. [40] *In* $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$, *the principle* OS_0 *implies the existence of nonstandard natural numbers,*

$$\mathsf{ENS}_0 : \exists x^0\, \neg\,\mathrm{st}(x),$$

as well as:

$$\mathsf{LLPO}^{\mathrm{st}}_0 : \forall^{\mathrm{st}} x^0, y^0\, (\,\varphi(x) \vee \psi(y)\,) \to \forall^{\mathrm{st}} x^0\, \varphi(x) \vee \forall^{\mathrm{st}} y^0\, \psi(y).$$

Proof. ENS_0 follows trivially from OS_0 by taking for $\varphi(x)$ some trivially true formula (for instance, $x = x$).

If $\forall^{\mathrm{st}} x^0, y^0\, (\,\varphi(x) \vee \psi(y)\,)$, then one can use external induction to prove:

$$\forall^{\mathrm{st}} n^0\, (\,\forall m \le n\, \varphi(m) \vee \forall m \le n\, \psi(m)\,).$$

By applying overspill to this statement, we see that it holds for some nonstandard n. Using external induction again, we can prove that a nonstandard natural number must be bigger than any standard natural number. Hence $\forall^{\mathrm{st}} x^0\, \varphi(x) \vee \forall^{\mathrm{st}} y^0\, \psi(y)$. $\qquad\square$

Of course, overspill can be formulated for all types:

$$\mathsf{OS} : \forall^{\mathrm{st}} x^\sigma\, \varphi(x) \to \exists x^\sigma\, (\,\neg\,\mathrm{st}(x) \wedge \varphi(x)\,).$$

But, actually, the interpretations that we will discuss will not only verify this principle, but also a far-reaching generalization of it, *viz.* a higher-type version of Nelson's idealization principle [38]:

$$\mathsf{I} : \forall^{\mathrm{st}} x^{\sigma^*} \exists y^\tau \forall x' \in_\sigma x\, \varphi(x', y) \to \exists y^\tau \forall^{\mathrm{st}} x^\sigma\, \varphi(x, y).$$

Proposition 3.3. [40] *In* $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$, *the idealization principle* I *implies overspill, as well as the statement that for every type σ there is a nonstandard sequence containing all the standard elements of that type:*

$$\mathsf{USEQ} : \exists y^{\sigma^*} \, \forall^{\mathrm{st}} x^{\sigma} \; x \in_{\sigma} y.$$

Proof. OS for an internal formula $\psi(y)$ follows from I by taking $\varphi(x,y) :\equiv y \neq x \wedge \psi(y)$, while USEQ follows by taking $\varphi(x,y) :\equiv x \in_{\sigma} y$. $\qquad \square$

Proposition 3.4. *In* $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$, *the idealization principle* I *implies the existence of nonstandard elements of any type,*

$$\mathsf{ENS} : \exists x^{\sigma} \, \neg \, \mathrm{st}(x),$$

as well as $\mathsf{LLPO}^{\mathrm{st}}$ *for any type:*

$$\mathsf{LLPO}^{\mathrm{st}} : \forall^{\mathrm{st}} x^{\sigma}, y^{\sigma} \, (\, \varphi(x) \vee \psi(y) \,) \to \forall^{\mathrm{st}} x^{\sigma} \, \varphi(x) \vee \forall^{\mathrm{st}} y^{\sigma} \, \psi(y).$$

Proof. The first statement is obvious, so we concentrate on the second.

Suppose $\forall^{\mathrm{st}} x^{\sigma}, y^{\sigma} \, (\, \varphi(x) \vee \psi(y) \,)$. Then one easily proves by external sequence induction and by taking $v := u$ that

$$\forall^{\mathrm{st}} u^{\sigma^*} \, \exists v^{\sigma^*} \, \forall u' \in u \, \big(\, u' \in v \wedge (\, \forall v' \in v \, \varphi(v') \vee \forall v' \in v \, \psi(v') \,) \,\big).$$

By applying idealization to this statement we obtain

$$\exists v^{\sigma^*} \, \forall^{\mathrm{st}} u^{\sigma} \, \big(\, u \in v \wedge (\, \forall v' \in v \, \varphi(v') \vee \forall v' \in v \, \psi(v') \,) \,\big),$$

from which $\mathsf{LLPO}^{\mathrm{st}}$ follows. $\qquad \square$

Classically, idealization is equivalent to its dual, which we have dubbed the *realization principle* (intuitionistically, things are not so clear):

$$\mathsf{R} : \forall y^{\tau} \exists^{\mathrm{st}} x^{\sigma} \, \varphi(x,y) \to \exists^{\mathrm{st}} x^{\sigma^*} \forall y^{\tau} \exists x' \in x \, \varphi(x',y).$$

As it turns out, both our interpretations will eliminate this principle as well. Actually, both interpretations for constructive nonstandard analysis eliminate the stronger *nonclassical realization principle*:

$$\mathsf{NCR} : \forall y^{\tau} \exists^{\mathrm{st}} x^{\sigma} \, \Phi(x,y) \to \exists^{\mathrm{st}} x^{\sigma^*} \forall y^{\tau} \exists x' \in x \, \Phi(x',y),$$

where $\Phi(x,y)$ can be any formula. This is quite remarkable, as NCR is incompatible with classical logic (hence the name) in that one can prove:

Proposition 3.5. *In* $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$, *the nonclassical realization principle* NCR *implies the undecidability of the standardness predicate:*

$$\neg \forall x^{\sigma} \, (\, \mathrm{st}(x) \vee \neg \, \mathrm{st}(x) \,).$$

Proof. Assume that standardness would be decidable. Then we would have

$$\forall y^{\sigma} \exists^{\mathrm{st}} x^{\sigma} \, (\, \mathrm{st}(y) \to x = y \,).$$

Applying NCR to this statement yields:

$$\exists^{\mathrm{st}} x^{\sigma^*} \forall y^{\sigma} \exists x' \in x \, (\, \mathrm{st}(y) \to x' = y \,),$$

which is the statement that there are only finitely many standard elements of type σ. This is clearly absurd. $\qquad \square$

3.2 Underspill

Underspill (in type 0) is the following statement:

$$\mathsf{US}_0 : \forall x^0 \, (\, \neg \operatorname{st}(x) \to \varphi(x) \,) \to \exists^{\mathrm{st}} x^0 \, \varphi(x).$$

In a constructive context it has the following nontrivial consequence (compare [5]):

Proposition 3.6. *In* $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$, *the underspill principle* US_0 *implies*

$$\mathsf{MP}^{\mathrm{st}}_0 : \big(\, \forall^{\mathrm{st}} x^0 \, (\, \varphi(x) \vee \neg\varphi(x) \,) \wedge \neg\neg \exists^{\mathrm{st}} x^0 \varphi(x) \,\big) \to \exists^{\mathrm{st}} x^0 \varphi(x).$$

In particular, $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} + \mathsf{US}_0 \vdash \neg\neg \operatorname{st}^0(x) \to \operatorname{st}^0(x).$

Proof. We reason in $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} + \mathsf{US}_0$. Suppose $\forall^{\mathrm{st}} x^0 \, (\, \varphi(x) \vee \neg\varphi(x) \,)$ and $\neg\neg\exists^{\mathrm{st}} x^0 \varphi(x)$.

Since the latter is intuitionistically equivalent to $\neg\forall^{\mathrm{st}} x^0 \neg\varphi(x)$, this means that for every infinite natural number ω, we have $\neg\forall x \leq \omega \neg\varphi(x)$. In other words, we have

$$\forall \omega^0 \, (\, \neg \operatorname{st}(\omega) \to \neg\forall x \leq \omega \neg\varphi(x) \,).$$

So by US_0 we have

$$\exists^{\mathrm{st}} y \, \neg\forall x \leq y \neg\varphi(x),$$

which implies $\exists^{\mathrm{st}} x \, \varphi(x)$ by decidability of $\varphi(x)$ for standard values of x.

As a special case we have $\neg\neg \operatorname{st}^0(x) \to \operatorname{st}^0(x)$, because $\operatorname{st}^0(x)$ is equivalent to $\exists^{\mathrm{st}} y^0 \, (\, x =_0 y \,)$ and equality of objects of type 0 is decidable. $\qquad \square$

Also underspill has a direct generalization to higher types:

$$\mathsf{US} : \forall x^\sigma \, (\, \neg \operatorname{st}(x) \to \varphi(x) \,) \to \exists^{\mathrm{st}} x^\sigma \, \varphi(x).$$

A natural question is whether this implies a version of Markov's Principle for all types. Our suspicion is that this is not the case, but we were unable to prove this.

3.3 Transfer

Following Nelson [38], the transfer principle is usually formulated as follows:

$$\mathsf{TP}_\forall : \forall^{\mathrm{st}} \underline{t} \, (\, \forall^{\mathrm{st}} x \, \varphi(x, \underline{t}) \to \forall x \, \varphi(x, \underline{t}) \,).$$

(Here, for once, we do not allow parameters: so it is important that x and $\underline{t}$ include all free variables of the formula φ.) This is classically, but not intuitionistically, equivalent to the following:

$$\mathsf{TP}_\exists : \forall^{\mathrm{st}} \underline{t} \, (\, \exists x \, \varphi(x, \underline{t}) \to \exists^{\mathrm{st}} x \, \varphi(x, \underline{t}) \,),$$

where, once again, we do not allow parameters.

It turns out that interpreting transfer is very difficult, especially in a constructive context (in fact, Avigad and Helzner have devoted an entire paper [5] to this issue). There are, at least, the following three problems:

1. Transfer principles together with overspill imply instances of the law of excluded middle, as was first shown by Moerdijk and Palmgren in [37]. In our setting we have:

   Proposition 3.7. *(a) In* $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$, *the combination of* ENS_0 *and* $\mathsf{TP}_\forall$ *implies the law of excluded middle for all internal arithmetical formulas.*

   (b) In $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$, *the combination of* USEQ *and* $\mathsf{TP}_\forall$ *implies the law of excluded middle for all internal formulas.*

Proof. Ad (a): assume ENS_0 and $\mathsf{TP}_\forall$. We show that every internal arithmetic formula φ is decidable by induction on the number of internal quantifiers in φ. Atomic formulas are decidable anyway, so the base case is easy.

If $\varphi(x^0, \underline{t})$ is an internal formula which is decidable and arithmetical (and with all free variables shown), then one can use internal induction to show

$$\forall y^0 \, (\, \exists x \leq y \, \neg\varphi(x, \underline{t}) \vee \forall x \leq y \, \varphi(x, \underline{t}) \,).$$

Let $\underline{t}$ be some arbitrary standard value and let y be some infinite natural number ω, using ENS_0. Then we either have $\exists x \leq \omega \, \neg\varphi(x, \underline{t})$ and in particular $\exists x \, \neg\varphi(x, \underline{t})$ or we have $\forall x \leq \omega \, \varphi(x, \underline{t})$ and in particular $\forall^{\mathrm{st}} x \, \varphi(x, \underline{t})$. So:

$$\forall^{\mathrm{st}}\underline{t} \, (\, \exists x \, \neg\varphi(x, \underline{t}) \vee \forall^{\mathrm{st}} x \, \varphi(x, \underline{t}) \,).$$

Applying $\mathsf{TP}_\forall$ once we get

$$\forall^{\mathrm{st}}\underline{t} \, (\, \exists x \, \neg\varphi(x, \underline{t}) \vee \forall x \, \varphi(x, \underline{t}) \,)$$

and applying it another time we get

$$\forall\underline{t} \, (\, \exists x \, \neg\varphi(x, \underline{t}) \vee \forall x \, \varphi(x, \underline{t}) \,).$$

This completes the induction step.

In (b) we argue similarly. First, we use the extensionality principles for functions and sequences to eliminate all equality predicates at higher types in favour of equalities at type 0. This makes all atomic formulas decidable. Since decidable formulas are closed under all propositional connectives, this leaves the case of the quantifiers. So suppose $\varphi(x^\sigma, \underline{t})$ is a internal formula which is decidable and let u be a sequence containing all standard elements of type σ (using USEQ). Then we have:

$$\exists u' \in u \, \neg\varphi(u', \underline{t}) \vee \forall u' \in u \, \varphi(u', \underline{t}).$$

In the former case it holds that $\exists x^\sigma \, \neg\varphi(x, \underline{t})$ and in the latter that $\forall^{\mathrm{st}} x^\sigma \varphi(x, \underline{t})$. And from here the argument proceeds as before. $\qquad\square$

2. As Avigad and Helzner observe in [5], also the combination of transfer principles with underspill results in a system which is no longer conservative over Heyting arithmetic. More precisely, adding US_0 and $\mathsf{TP}_\forall$, or US_0 and $\mathsf{TP}_\exists$, to $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ results in a system which is no longer conservative over Heyting arithmetic HA. The reason is that there are quantifier-free formulas $A(x)$ such that

$$\mathsf{HA} \nvdash \neg\neg\exists x \, A(x) \rightarrow \exists x \, A(x).$$

Since one can prove a version of Markov's Principle in $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} + \mathsf{US}_0$, adding either $\mathsf{TP}_\forall$ or $\mathsf{TP}_\exists$ to it would result in a nonconservative extension of HA (and hence of $\mathsf{E\text{-}HA}^{\omega*}$). We refer to [5] for more details.

3. The last point applies to functional interpretations only. As is well-known, in the context of functional interpretations the axiom of extensionality always presents a serious problem and when developing a functional interpretation of nonstandard arithmetic, the situation is no different. Now, $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ includes an internal axiom of extensionality (as it is part of $\mathsf{E\text{-}HA}^{\omega*}$), but for the functional interpretation that we will introduce in Section 5 that will be harmless. What will be very problematic for us, however, is the following version of the axiom of extensionality: if for two elements f, g of type $\sigma_1 \rightarrow (\sigma_2 \rightarrow \ldots \rightarrow 0))$, we define

$$f =^{\mathrm{st}} g \; :\equiv \; \forall^{\mathrm{st}} x_1^{\sigma_1}, x_2^{\sigma_2}, \ldots \, (\, f\underline{x} =_0 g\underline{x} \,),$$

then extensionality formulated as

$$\forall^{\mathrm{st}} f \, \forall^{\mathrm{st}} x, y \, (\, x =^{\mathrm{st}} y \to fx =^{\mathrm{st}} fy \,)$$

will have no witness definable in ZFC. But that means that also $\mathsf{TP}_\forall$ can have no witness definable in ZFC: for in the presence of $\mathsf{TP}_\forall$ both versions of extensionality are equivalent.

One way out of this quandary, which is strongly suggested by the last point and is the route taken in most sources (beginning with [36]), is to have transfer not as a principle, but as a *rule*. As we will see, this turns out to be feasible. In fact, we will have two transfer rules (which are not equivalent, not even classically):

$$\frac{\forall^{\mathrm{st}} x \, \varphi(x)}{\forall x \, \varphi(x)} \;\; \mathsf{TR}_\forall \qquad \frac{\exists x \, \varphi(x)}{\exists^{\mathrm{st}} x \, \varphi(x)} \;\; \mathsf{TR}_\exists$$

(This time round there are no special requirements on the parameters of φ.)

Remark 3.8. In this section we have also explored several connections between nonstandard principles. More principles will be introduced below and we will prove one more implication (see Proposition 5.11). We have not tried to determine the precise relationships between these principles – in particular, we do not know precisely which principles follow and do not follow over $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ from combinations of other principles. In fact, we believe that mapping these connections would be an interesting research project.

4 Herbrand realizability

In this section we will introduce a new realizability interpretation, which will allow us to prove our first consistency and conservation results in the context of $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$. Our treatment here will be entirely proof-theoretic; for a semantic approach towards Herbrand realizability, see [6].

4.1 The interpretation

The interpretation works by associating to every formula $\Phi(\underline{x})$ a formula $\Psi(\underline{t}, \underline{x})$, also denoted by $\underline{t} \, \mathsf{hr} \, \Phi(\underline{x})$, where $\underline{t}$ is a tuple of new variables *all of which are of sequence type*, determined solely by the logical form of $\Phi(\underline{x})$. The soundness proof will then involve showing that for every formula $\Phi(\underline{x})$ of $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ with $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} \vdash \Phi(\underline{x})$ there is an appropriate tuple $\underline{t}$ of terms from $\mathcal{T}^*$ such that $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} \vdash \underline{t} \, \mathsf{hr} \, \Phi(\underline{x})$.

The idea of the interpretation is that we interpret internal quantifers uniformly and that we do not attempt to give them any computational content. In a sense, the only predicate to which we will assign any computational content is the standardness predicate st: to realize $\mathrm{st}^\sigma(x)$, however, it suffices to provide a nonempty finite list of terms of type σ, one of which will have to be equal to x. Therefore to realize a statement of the form $\exists^{\mathrm{st}} x^\sigma \, \Phi(x)$ one only needs to provide a finite list $\langle y_0, \ldots, y_n \rangle$ and to make sure that $\Phi(y_i)$ is realized for some $i \leq n$ (which is like giving a Herbrand disjunction; hence the name "Herbrand realizability").

The precise definition is as follows:

Definition 4.1 (Herbrand realizability for $\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*}$)**.**

$$
\begin{array}{rcl}
[\,]\,\mathsf{hr}\,\varphi &:\equiv& \varphi \quad \text{for an internal atomic formula } \varphi, \\[4pt]
s\,\mathsf{hr}\,\mathsf{st}(x) &:\equiv& x \in s, \\[4pt]
\underline{s},\underline{t}\,\mathsf{hr}(\Phi \vee \Psi) &:\equiv& \underline{s}\,\mathsf{hr}\,\Phi \vee \underline{t}\,\mathsf{hr}\,\Psi, \\[4pt]
\underline{s},\underline{t}\,\mathsf{hr}(\Phi \wedge \Psi) &:\equiv& \underline{s}\,\mathsf{hr}\,\Phi \wedge \underline{t}\,\mathsf{hr}\,\Psi, \\[4pt]
\underline{s}\,\mathsf{hr}(\Phi \to \Psi) &:\equiv& \forall^{\mathrm{st}}\underline{t}\,(\,\underline{t}\,\mathsf{hr}\,\Phi \to \underline{s}\,[\underline{t}\,]\,\mathsf{hr}\,\Psi), \\[4pt]
\underline{s}\,\mathsf{hr}\,\exists x\,\Phi(x) &:\equiv& \exists x\,(\underline{s}\,\mathsf{hr}\,\Phi(x)), \\[4pt]
\underline{s}\,\mathsf{hr}\,\forall x\,\Phi(x) &:\equiv& \forall x\,(\underline{s}\,\mathsf{hr}\,\Phi(x)), \\[4pt]
s,\underline{t}\,\mathsf{hr}\,\exists^{\mathrm{st}}x\,\Phi(x) &:\equiv& \exists s' \in s\,\big(\underline{t}\,\mathsf{hr}\,\Phi(s')\,\big), \\[4pt]
\underline{s}\,\mathsf{hr}\,\forall^{\mathrm{st}}x\,\Phi(x) &:\equiv& \forall^{\mathrm{st}}x\,\big(\underline{s}\,[x]\,\mathsf{hr}\,\Phi(x)\big).
\end{array}
$$

Before we show the soundness of the interpretation, we first prove some easy lemmas:

Definition 4.2 (The $\exists^{\mathrm{st}}$-free formulas)**.** We call a formula (in the language of $\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*}$) $\exists^{\mathrm{st}}$-free, if it is built up from atomic formulas (including $\bot$) using the connectives $\wedge$, $\vee$, $\to$ and the quantifiers $\exists x$, $\forall x$ and $\forall^{\mathrm{st}}x$. Alternatively, one could say that these are the formulas in which st and $\exists^{\mathrm{st}}$ do not occur. We denote such formulas by $\Phi_{\not\exists^{\mathrm{st}}}$.

Lemma 4.3 (Interpretation of $\exists^{\mathrm{st}}$-free formulas)**.** *All the interpretations $\underline{t}\,\mathsf{hr}\,\Phi(\underline{x})$ of formulas $\Phi(\underline{x})$ of $\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*}$ are $\exists^{\mathrm{st}}$-free. In addition, every $\exists^{\mathrm{st}}$-free formula is interpreted by itself. Hence the interpretation is idempotent.*

The following lemma will be crucial for what follows:

Lemma 4.4 (Realizers are provably upwards closed)**.** *The formula $\underline{t}\,\mathsf{hr}\,\Phi(\underline{x})$ is provably upwards closed in $\underline{t}$, that is:*

$$
\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*} \vdash \underline{s}\,\mathsf{hr}\,\Phi \wedge \underline{s} \preceq \underline{t} \to \underline{t}\,\mathsf{hr}\,\Phi.
$$

Proof. By induction on the structure of Φ, using the monotonicity of the new application in the first component in the clauses for $\to$ and $\forall^{\mathrm{st}}$. $\qquad\qquad\square$

Theorem 4.5 (Soundness of Herbrand realizability)**.** *Let Φ be an arbitrary formula of $\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*}$ and $\Delta_{\not\exists^{\mathrm{st}}}$ be an arbitrary set of $\exists^{\mathrm{st}}$-free sentences. Whenever*

$$
\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*} + \Delta_{\not\exists^{\mathrm{st}}} \quad \vdash \quad \Phi(\underline{x}),
$$

then one can extract from the formal proof closed terms $\underline{t}$ in $\mathcal{T}^*$, such that

$$
\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*} + \Delta_{\not\exists^{\mathrm{st}}} \quad \vdash \quad \underline{t}\,\mathsf{hr}\,\Phi(\underline{x}).
$$

Proof. As for the logical axioms and rules, the differences with the usual soundness proof of modified realizability for $\mathsf{E\text{-}HA}^{\omega}$ (as in [48] or Theorem 5.8 in [25]) are

(a) that we require the realizing terms to be closed,

(b) that we have a nonconstructive interpretation of disjunction, and

(c) that we interpret the quantifiers in a uniform fashion.

Therefore one has to make the following modifications:

1. The contraction axiom $A \vee A \to A$ is realized by $\Lambda\underline{x},\underline{y}.\underline{x}*\underline{y}$, using that the collection of realizers is provably upwards closed.

2. The weakening axiom $A \to A \vee B$ is realized by $\Lambda\underline{x}.\underline{x},\mathcal{O}$.

17

3. The permutation axiom $A \vee B \to B \vee A$ is realized by $\Lambda \underline{x}, \underline{y}.\underline{y}, \underline{x}$.

4. The axioms of $\forall$-elimination $\forall x \Phi(x) \to \Phi(t)$ and $\exists$-introduction $\Phi(t) \to \exists x \Phi(x)$ are realized by the identity tuple $\Lambda \underline{x} . \underline{x}$.

5. The expansion rule $\frac{A \to B}{A \vee C \to A \vee C}$: if $\underline{t} \, \mathsf{hr} \, A \to B$, then $\Lambda \underline{x}, \underline{y}.\underline{t}[\underline{x}], \underline{y}$ is a Herbrand realizer of $A \vee C \to A \vee C$.

6. The $\forall$-introduction rule $\frac{\Phi \to \Psi(x)}{\Phi \to \forall x \Psi(x)}$ is interpreted, because $\underline{s} \, \mathsf{hr}(\Phi \to \Psi(x))$ implies $\underline{s} \, \mathsf{hr}(\Phi \to \forall x \Psi(x))$: for if $\underline{t} \, \mathsf{hr} \, \Phi$ and $\mathrm{st}(\underline{t})$, then $\underline{s}[\underline{t}] \, \mathsf{hr} \, \Psi(x)$ and therefore $\underline{s}[\underline{t}] \, \mathsf{hr} \, \forall x \Psi(x)$.

7. The $\exists$-introduction rule $\frac{\Phi(x) \to \Psi}{\exists x \Phi(x) \to \Psi}$ is interpreted, because $\underline{s} \, \mathsf{hr}(\Phi(x) \to \Psi)$ implies $\underline{s} \, \mathsf{hr}(\exists x \Phi(x) \to \Psi)$: for if $\underline{t} \, \mathsf{hr} \, \exists x \Phi(x)$ and $\mathrm{st}(\underline{t})$, then there is an x such that $\underline{t} \, \mathsf{hr} \, \Phi(x)$, from which it follows that $\underline{s}[\underline{t}] \, \mathsf{hr} \, \Psi$.

The sentences from $\Delta_{\not\exists^{\mathrm{st}}}$ and the axioms of $\mathsf{E\text{-}HA}^{\omega*}$, including SA and the defining axioms for equality, successor, combinators and recursion, are $\exists^{\mathrm{st}}$-free and therefore realized by themselves. Therefore it remains to show the soundness of the following rules and axioms:

1. The external quantifier axioms EQ: both directions in $\exists^{\mathrm{st}} x \Phi(x) \leftrightarrow \exists x(\mathrm{st}(x) \wedge \Phi(x))$ are interpreted by the the identity. In $\forall^{\mathrm{st}} x \Phi(x) \leftrightarrow \forall x(\mathrm{st}(x) \to \Phi(x))$ the right-to-left direction is realized by $\Lambda \underline{s}, x.\underline{s}\,[\langle x \rangle]$, while the left-to-right direction is realized by $\Lambda \underline{s}, x.\underline{s}[x_0] * \ldots * \underline{s}[x_{|x|-1}]$.

2. The axiom schemes $\mathcal{T}_{\mathrm{st}}^*$: the principle $\mathrm{st}(x) \wedge x = y \to \mathrm{st}(y)$ is realized by the identity, while $\mathrm{st}(t)$ is realized by $\langle t \rangle$. In addition, $\mathrm{st}(f) \wedge \mathrm{st}(x) \to \mathrm{st}(fx)$ is realized by $\Lambda f, x.\langle f_i(x_j) \rangle_{i<|f|,j<|x|}$.

3. The induction schema $\mathsf{IA}^{\mathrm{st}}$: suppose $\underline{s} \, \mathsf{hr} \, \Phi(0)$ and $\underline{t} \, \mathsf{hr} \, \forall^{\mathrm{st}} n(\Phi(n) \to \Phi(n+1))$, with $\mathrm{st}(\underline{s})$ and $\mathrm{st}(\underline{t})$. Then $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ proves by external induction that for standard natural numbers n, the term $\underline{\mathcal{R}}(n, \underline{s}, \underline{t})$ is standard and $\underline{\mathcal{R}}(n, \underline{s}, \underline{t}) \, \mathsf{hr} \, \Phi(n)$. Therefore $\Lambda \underline{x}, \underline{y}, n . \underline{\mathcal{R}}(n, \underline{x}, \underline{y}) \, \mathsf{hr} \, \mathsf{IA}^{\mathrm{st}}$.

$\square$

4.2 The characteristic principles of Herbrand realizability

In this section we will prove that HAC (the herbrandized axiom of choice), $\mathsf{HIP}_{\not\exists^{\mathrm{st}}}$ (the herbrandized independence of premise principle for $\exists^{\mathrm{st}}$-free formulas) and NCR axiomatize Herbrand realizability:

1. HAC :
$$\forall^{\mathrm{st}} x \exists^{\mathrm{st}} y \, \Phi(x, y) \to \exists^{\mathrm{st}} F \forall^{\mathrm{st}} x \exists y \in F(x) \, \Phi(x, y),$$

   where $\Phi(x, y)$ can be any formula. If $\Phi(x, y)$ is upwards closed in y, then this is equivalent to
$$\forall^{\mathrm{st}} x \exists^{\mathrm{st}} y \, \Phi(x, y) \to \exists^{\mathrm{st}} F \forall^{\mathrm{st}} x \, \Phi(x, F(x)).$$

2. $\mathsf{HIP}_{\not\exists^{\mathrm{st}}}$:
$$\big(\Phi \to \exists^{\mathrm{st}} y \, \Psi(y) \big) \to \exists^{\mathrm{st}} y \, \big(\Phi \to \exists y' \in y \, \Psi(y') \big),$$

   where Φ has to be an $\exists^{\mathrm{st}}$-free formula and $\Psi(y)$ can be any formula. If $\Psi(y)$ is upwards closed in y, then this is equivalent to
$$\big(\Phi \to \exists^{\mathrm{st}} y \Psi(y) \big) \to \exists^{\mathrm{st}} y \, \big(\Phi \to \Psi(y) \big).$$

3. NCR:
$$\forall x \exists^{\text{st}} y \, \Phi(x,y) \to \exists^{\text{st}} y \, \forall x \, \exists y' \in y \, \Phi(x,y'),$$

where $\Phi(x,y)$ can be any formula. If $\Phi(x,y)$ is upwards closed in y, then this is equivalent to
$$\forall x \exists^{\text{st}} y \, \Phi(x,y) \to \exists^{\text{st}} y \forall x \, \Phi(x,y).$$

Theorem 4.6 (Characterization theorem for Herbrand realizability)**.**

1. *For any instance* Φ *of* HAC, $\text{HIP}_{\overline{\exists}^{\text{st}}}$ *or* NCR*, there are closed terms* $\underline{t}$ *in* $\mathcal{T}^*$ *such that*
$$\text{E-HA}^{\omega*}_{\text{st}} \vdash \underline{t} \, \text{hr} \, \Phi.$$

2. *For any formula* Φ *of* $\text{E-HA}^{\omega*}_{\text{st}}$, *we have*
$$\text{E-HA}^{\omega*}_{\text{st}} + \text{HAC} + \text{HIP}_{\overline{\exists}^{\text{st}}} + \text{NCR} \vdash \Phi \leftrightarrow \exists^{\text{st}} \underline{x} \, (\underline{x} \, \text{hr} \, \Phi).$$

Proof.   Soundness of HAC: If $\underline{r} = s, \underline{t}$ and $\underline{r} \, \text{hr} \, \forall^{\text{st}} x \exists^{\text{st}} y \Phi(x,y)$, then for every standard x there is an $s' \in s[x]$ such that $\underline{t}[x] \, \text{hr} \, \Phi(x,s')$.   Hence $\langle \lambda x.s[x] \rangle, \underline{t} \, \text{hr} \, \exists^{\text{st}} F \forall^{\text{st}} x \exists y \in F(x) \, \Phi(x,y)$. So HAC is realized by $\Lambda x, \underline{y}.\langle \lambda z.x[z] \rangle, \underline{y}$.

Soundness of $\text{HIP}_{\overline{\exists}^{\text{st}}}$: Suppose Φ is $\exists^{\text{st}}$-free, $\underline{r} = s, \underline{t}$ and $\underline{r} \, \text{hr} \, \Phi \to \exists^{\text{st}} y \Phi(\underline{y})$. This means that if Φ would hold, then there would be an $s' \in s$ such that $\underline{t} \, \text{hr} \, \Psi(s')$. Hence $\langle s \rangle, \underline{t} \, \text{hr} \, \exists^{\text{st}} y (\Phi \to \exists y' \in y \Psi(y'))$. So $\text{HIP}_{\overline{\exists}^{\text{st}}}$ is realized by $\Lambda x, \underline{y}.\langle x \rangle, \underline{y}$.

In a similar manner one checks that also NCR is realized by $\Lambda x, \underline{y}.\langle x \rangle, \underline{y}$. This completes the proof of item 1.

Item 2 one proves by induction on the logical structure of Φ. We discuss implication as an illustrative case, as it is by far the hardest, and leave the other cases to the reader. We reason in $\text{E-HA}^{\omega*}_{\text{st}} + \text{HAC} + \text{HIP}_{\overline{\exists}^{\text{st}}} + \text{NCR}$. By induction hypothesis, we have that $\Phi \leftrightarrow \exists^{\text{st}} \underline{t} (\underline{t} \, \text{hr} \, \Phi)$ and $\Psi \leftrightarrow \exists^{\text{st}} \underline{s} (\underline{s} \, \text{hr} \, \Psi)$ and therefore

$$\Phi \to \Psi$$

is equivalent to
$$\exists^{\text{st}} \underline{t} (\underline{t} \, \text{hr} \, \Phi) \to \exists^{\text{st}} \underline{s} (\underline{s} \, \text{hr} \, \Psi),$$

which in turn is equivalent to:
$$\forall^{\text{st}} \underline{t} \, \big(\, \underline{t} \, \text{hr} \, \Phi \to \exists^{\text{st}} \underline{s} \, (\, \underline{s} \, \text{hr} \, \Psi \,) \, \big).$$

Because $\underline{t} \, \text{hr} \, \Phi$ is $\exists^{\text{st}}$-free and $\underline{s} \, \text{hr} \, \Psi$ is upwards closed in $\underline{s}$, we can use $\text{HIP}_{\overline{\exists}^{\text{st}}}$ to rewrite this as:
$$\forall^{\text{st}} \underline{t} \, \exists^{\text{st}} \underline{s} \, \big(\, \underline{t} \, \text{hr} \, \Phi \to \underline{s} \, \text{hr} \, \Psi \, \big).$$

As $\underline{t} \, \text{hr} \, \Phi \to \underline{s} \, \text{hr} \, \Psi$ is upwards closed in $\underline{s}$ and finite sequence application and ordinary application are interdefinable, we can use HAC to see that this is equivalent to:
$$\exists^{\text{st}} \underline{s} \, \forall^{\text{st}} \underline{t} \, \big(\, \underline{t} \, \text{hr} \, \Phi \to \underline{s}[\underline{t}] \, \text{hr} \, \Psi \, \big),$$

which is precisely the meaning of $\exists^{\text{st}} \underline{s} \, (\, \underline{s} \, \text{hr}(\Phi \to \Psi) \,)$. $\square$

Theorem 4.7 (Main theorem on program extraction by hr)**.** *Let* $\forall^{\text{st}} x \exists^{\text{st}} y \Phi(x,y)$ *be a sentence of* $\text{E-HA}^{\omega*}_{\text{st}}$ *and* $\Delta_{\overline{\exists}^{\text{st}}}$ *be an arbitrary set* $\exists^{\text{st}}$*-free sentences. Then the following rule holds*

$$\text{E-HA}^{\omega*}_{\text{st}} + \text{HAC} + \text{HIP}_{\overline{\exists}^{\text{st}}} + \text{NCR} + \Delta_{\overline{\exists}^{\text{st}}} \vdash \forall^{\text{st}} x \, \exists^{\text{st}} y \, \Phi(x,y) \Rightarrow$$
$$\text{E-HA}^{\omega*}_{\text{st}} + \text{HAC} + \text{HIP}_{\overline{\exists}^{\text{st}}} + \text{NCR} + \Delta_{\overline{\exists}^{\text{st}}} \vdash \forall^{\text{st}} x \, \exists y \in t(x) \, \Phi(x,y),$$

where t *is a closed term from* $\mathcal{T}^*$ *which is extracted from the original proof using Herbrand realizability.*

In the particular case where both $\Phi(x,y)$ and $\Delta_{\nexists^{st}}$ are internal, the conclusion yields

$$\mathsf{E\text{-}HA}^{\omega *} + \Delta_{\nexists^{st}} \vdash \forall x\, \exists y \in t(x)\, \Phi(x,y).$$

If we assume that the sentences from $\Delta_{\nexists^{st}}$ are not just internal, but also true (in the set-theoretic model), the conclusion implies that $\forall x\, \exists y \in t(x)\, \Phi(x,y)$ must be true as well.

Proof. If

$$\mathsf{E\text{-}HA}^{\omega *}_{st} + \mathsf{HAC} + \mathsf{HIP}_{\nexists^{st}} + \mathsf{NCR} + \Delta_{\nexists^{st}} \vdash \forall^{st}x\, \exists^{st}y\, \Phi(x,y),$$

then the soundness proof yields terms $r, \underline{s}$ such that

$$\mathsf{E\text{-}HA}^{\omega *}_{st} + \Delta_{\nexists^{st}} \vdash r, \underline{s}\, \mathsf{hr}\, \forall^{st}x\, \exists^{st}y\, \Phi(x,y).$$

Since $r, \underline{s}\, \mathsf{hr}\, \forall^{st}x\, \exists^{st}y\, \Phi(x,y)$ is by definition $\forall^{st}x\, \exists y \in r[x]\, (\underline{s}\, \mathsf{hr}\, \Phi(x,y))$, the first statement follows by taking $t = \lambda x.r[x]$.

If both $\Phi(x,y)$ and $\Delta_{\nexists^{st}}$ are internal, then $\underline{s}$ is empty and we get

$$\mathsf{E\text{-}HA}^{\omega *}_{st} + \Delta_{\nexists^{st}} \vdash \forall^{st}x\, \exists y \in t(x)\, \Phi(x,y).$$

By internalizing the statement, we obtain

$$\mathsf{E\text{-}HA}^{\omega *} + \Delta_{\nexists^{st}} \vdash \forall x\, \exists y \in t(x)\, \Phi(x,y).$$

Since all the axioms of $\mathsf{E\text{-}HA}^{\omega *}$ are true, this implies that $\forall x\, \exists y \in t(x)\, \Phi(x,y)$ will be true, whenever $\Delta_{\nexists^{st}}$ is. $\qquad\square$

4.3 Discussion

The main virtue of Herbrand realizability may be that it points one's attention to principles like HAC and NCR and that it gives one a simple proof of their consistency. However, as a method for eliminating nonstandard principles from proofs, Herbrand realizability has serious limitations. It does eliminate the realization principle R (it even eliminates the nonclassical principle NCR), but overspill, the idealization principle I and the transfer rules are $\exists^{st}$-free and therefore simply passed to the verifying system. Even worse, the underspill principle US_0 does not have a computable Herbrand realizer:

Proposition 4.8. MP^{st}_0 *and* US_0 *do not have computable Herbrand realizers.*

Proof. It is well-known that there can be no computable function witnessing the modified realizability interpretation of Markov's principle, because its existence would imply the decidability of the halting problem. A similar argument shows that MP^{st}_0 does not have a computable Herbrand realizer: Kleene's T-predicate $T(e,x,n)$ is primitive recursive and hence

$$\mathsf{E\text{-}HA}^{\omega *}_{st} \vdash \forall^{st}e, x, n\, \big(T(e,x,n) \vee \neg T(e,x,n) \big).$$

So if MP^{st}_0 would have a computable realizer, then so would

$$\forall^{st}e\, \big(\neg\neg\exists^{st}n\, T(e,e,n) \to \exists^{st}n\, T(e,e,n) \big).$$

But if t would be such a realizer, we could decide the halting problem by checking $T(e,e,n)$ for all $n \in t[e]$.

Since US_0 implies MP^{st}_0 (see Proposition 3.6), it follows that US_0 does not have a computable realizer either. $\qquad\square$

In the next section, we will show that these problems can be overcome by moving from realizability to, more complicated, functional interpretations.

5 A functional interpretation for $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$

In this section we will introduce and study a functional interpretation for $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$.

5.1 The interpretation

The basic idea of the D_{st}-interpretation (the nonstandard Dialectica interpretation) is to associate to every formula $\Phi(\underline{a})$ a new formula $\Phi(\underline{a})^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},\underline{a})$ such that

1. all variables in $\underline{x}$ are of sequence type and

2. $\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},\underline{a})$ is upwards closed in $\underline{x}$.

We will interpret the standardness predicate st^{σ} similarly to the case for Herbrand realizability: For a realizer for the interpretation of $\mathrm{st}^{\sigma}(x)$ we will require a standard finite list $\langle y_0,\ldots,y_n\rangle$ of candidates, one of which must be equal to x.

Definition 5.1 (The D_{st}-interpretation for $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$). We associate to every formula $\Phi(\underline{a})$ in the language of $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ (with free variables among $\underline{a}$) a formula $\Phi(\underline{a})^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},\underline{a})$ in the same language (with the same free variables) by:

(i) $\varphi(\underline{a})^{D_{\mathrm{st}}} :\equiv \varphi_{D_{\mathrm{st}}}(\underline{a}) :\equiv \varphi(\underline{a})$ for internal atomic formulas $\varphi(\underline{a})$,

(ii) $\mathrm{st}^{\sigma}(u^{\sigma})^{D_{\mathrm{st}}} :\equiv \exists^{\mathrm{st}}x^{\sigma^*} u \in_{\sigma} x$.

Let $\Phi(\underline{a})^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},\underline{a})$ and $\Psi(\underline{b})^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\psi_{D_{\mathrm{st}}}(\underline{u},\underline{v},\underline{b})$. Then

(iii) $(\Phi(\underline{a}) \wedge \Psi(\underline{b}))^{D_{\mathrm{st}}} :\equiv \exists^{\mathrm{st}}\underline{x},\underline{u}\forall^{\mathrm{st}}\underline{y},\underline{v}\,\big(\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},\underline{a}) \wedge \psi_{D_{\mathrm{st}}}(\underline{u},\underline{v},\underline{b})\big)$,

(iv) $(\Phi(\underline{a}) \vee \Psi(\underline{b}))^{D_{\mathrm{st}}} :\equiv \exists^{\mathrm{st}}\underline{x},\underline{u}\forall^{\mathrm{st}}\underline{y},\underline{v}\,\big(\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},\underline{a}) \vee \psi_{D_{\mathrm{st}}}(\underline{u},\underline{v},\underline{b})\big)$,

(v) $(\Phi(\underline{a}) \to \Psi(\underline{b}))^{D_{\mathrm{st}}} :\equiv \exists^{\mathrm{st}}\underline{U},\underline{Y}\forall^{\mathrm{st}}\underline{x},\underline{v}\,\big(\forall \underline{y} \in \underline{Y}[\underline{x},\underline{v}]\,\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},\underline{a}) \to \psi_{D_{\mathrm{st}}}(\underline{U}[\underline{x}],\underline{v},\underline{b})\big)$.

Let $\Phi(z,\underline{a})^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},z,\underline{a})$, with the free variable z not occuring among the $\underline{a}$. Then

(vi) $(\forall z\Phi(z,\underline{a}))^{D_{\mathrm{st}}} :\equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\forall z\,\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},z,\underline{a})$,

(vii) $(\exists z\Phi(z,\underline{a}))^{D_{\mathrm{st}}} :\equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\exists z\forall \underline{y}' \in \underline{y}\,\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y}',z,\underline{a})$,

(viii) $(\forall^{\mathrm{st}}z\Phi(z,\underline{a}))^{D_{\mathrm{st}}} :\equiv \exists^{\mathrm{st}}\underline{X}\forall^{\mathrm{st}}z,\underline{y}\,\varphi_{D_{\mathrm{st}}}(\underline{X}[z],\underline{y},z,\underline{a})$,

(ix) $(\exists^{\mathrm{st}}z\Phi(z,\underline{a}))^{D_{\mathrm{st}}} :\equiv \exists^{\mathrm{st}}\underline{x},z\,\forall^{\mathrm{st}}\underline{y}\,\exists z' \in z\,\forall \underline{y}' \in \underline{y}\,\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y}',z',\underline{a})$.

Definition 5.2. We say that a formula Φ is a $\forall^{\mathrm{st}}$-formula if $\Phi \equiv \forall^{\mathrm{st}}\underline{x}\,\varphi(\underline{x})$, with $\varphi(\underline{x})$ internal.

Lemma 5.3. *Let Φ be a $\forall^{\mathrm{st}}$-formula. Then $\Phi^{D_{\mathrm{st}}} \equiv \Phi$.*

Proof. By induction on the structure of Φ. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad \square$

Notice that because of the clause for $\exists^{\mathrm{st}}z$ the interpretation is not idempotent. Similarly to what is the case for Herbrand realizability it will be crucial that realizers are upwards closed:

Lemma 5.4. *Let $\Phi(\underline{a})$ be a formula in the language of $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ with interpretation $\exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},\underline{a})$. Then the formula $\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},\underline{a})$ is provably upwards closed in $\underline{x}$, i.e.,*

$$\mathsf{E\text{-}HA}^{\omega*} \vdash \varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},\underline{a}) \wedge \underline{x} \preceq \underline{x}' \to \varphi_{D_{\mathrm{st}}}(\underline{x}',\underline{y},\underline{a}).$$

Proof. By induction on the structure of $\Phi(\underline{a})$, using Lemma 2.22 in the clauses for $\rightarrow$ and $\forall^{\mathrm{st}}$. $\qquad\square$

The D_{st}-interpretation will allow us to interpret the nonclassical realization principle NCR, and also both I and HAC. Additionally we will be able to interpret a herbrandized independence of premise principle for formulas of the form $\forall^{\mathrm{st}}x\,\varphi(x)$, and also a herbrandized form of a generalized Markov's principle:

1. $\mathsf{HIP}_{\forall^{\mathrm{st}}}$:

$$\left(\forall^{\mathrm{st}}x\,\varphi(x) \rightarrow \exists^{\mathrm{st}}y\Psi(y)\right) \rightarrow \exists^{\mathrm{st}}y\left(\forall^{\mathrm{st}}x\,\varphi(x) \rightarrow \exists y' \in y\,\Psi(y')\right),$$

   where $\Psi(y)$ is a formula in the language of $\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*}$ and $\varphi(x)$ is an internal formula. If $\Psi(y)$ is upwards closed in y, then this is equivalent to

$$\left(\forall^{\mathrm{st}}x\,\varphi(x) \rightarrow \exists^{\mathrm{st}}y\Psi(y)\right) \rightarrow \exists^{\mathrm{st}}y\left(\forall^{\mathrm{st}}x\,\varphi(x) \rightarrow \Psi(y)\right).$$

2. $\mathsf{HGMP}^{\mathrm{st}}$:

$$\left(\forall^{\mathrm{st}}x\,\varphi(x) \rightarrow \psi\right) \rightarrow \exists^{\mathrm{st}}x\left(\forall x' \in x\,\varphi(x') \rightarrow \psi\right),$$

   where $\varphi(x)$ and ψ are internal formulas in the language of $\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*}$. If $\varphi(x)$ is downwards closed in x, then this is equivalent to

$$\left(\forall^{\mathrm{st}}x\,\varphi(x) \rightarrow \psi\right) \rightarrow \exists^{\mathrm{st}}x(\varphi(x) \rightarrow \psi).$$

The latter gives us a form of Markov's principle by taking $\psi \equiv 0 =_0 1$ and $\varphi(x) \equiv \neg\varphi_0(x)$ (with $\varphi_0(x)$ internal and quantifier-free), whence the name.

Theorem 5.5 (Soundness of the D_{st}-interpretation). *Let $\Phi(\underline{a})$ be a formula of $\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*}$ and let Δ_{int} be a set of internal sentences. If*

$$\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*} + \mathsf{I} + \mathsf{NCR} + \mathsf{HAC} + \mathsf{HGMP}^{\mathrm{st}} + \mathsf{HIP}_{\forall^{\mathrm{st}}} + \Delta_{\mathrm{int}} \vdash \Phi(\underline{a})$$

and $\Phi(\underline{a})^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi_{D_{\mathrm{st}}}(\underline{x},\underline{y},\underline{a})$, then from the proof we can extract closed terms $\underline{t}$ in $\mathcal{T}^*$ such that

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathrm{int}} \vdash \forall\underline{y}\,\varphi_{D_{\mathrm{st}}}(\underline{t},\underline{y},\underline{a}).$$

Proof. As in the proof of the soundness of the Dialectica interpretation we proceed by induction on the length of the derivation.

1. We will first consider the logical axioms and rules:

   (a) $A \rightarrow A \wedge A$:
      With $A^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$ we have

$$(A \rightarrow A \wedge A)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{X}',\underline{X}'',\underline{Y}\forall^{\mathrm{st}}\underline{x},\underline{y}',\underline{y}''$$
$$\left(\forall\underline{z} \in \underline{Y}[\underline{x},\underline{y}',\underline{y}'']\varphi(\underline{x},\underline{z},\underline{a}) \rightarrow \varphi(\underline{X}'[\underline{x}],\underline{y}',\underline{a}) \wedge \varphi(\underline{X}''[\underline{x}],\underline{y}'',\underline{a})\right),$$

      and we can take

$$\begin{aligned}
\underline{X}' &:= \Lambda\underline{x}.\underline{x}, \\
\underline{X}'' &:= \Lambda\underline{x}.\underline{x}, \\
\underline{Y} &:= \Lambda\underline{x},\underline{y}',\underline{y}''.\langle\underline{y}',\underline{y}''\rangle.
\end{aligned}$$

   (b) $A \vee A \rightarrow A$:
      With $A^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$ we have

$$(A \vee A \rightarrow A)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{X}'',\underline{Y},\underline{Y}'\forall^{\mathrm{st}}\underline{x},\underline{x}',\underline{y}''$$
$$\left((\forall\underline{z} \in \underline{Y}[\underline{x},\underline{x}',\underline{y}'']\varphi(\underline{x},\underline{z},\underline{a}) \vee \forall\underline{z}' \in \underline{Y}'[\underline{x},\underline{x}',\underline{y}'']\varphi(\underline{x}',\underline{z}',\underline{a}))\right.$$
$$\left.\rightarrow \varphi(\underline{X}''[\underline{x},\underline{x}'],\underline{y}'',\underline{a})\right),$$

22

and we can take

$$\begin{aligned}
\underline{X}'' &:= \Lambda\,\underline{x},\underline{x}'\,.\,\underline{x}*\underline{x}',\\
\underline{Y} &:= \Lambda\,\underline{x},\underline{y},\underline{y}''\,.\,\langle\underline{y}''\rangle,\\
\underline{Y}' &:= \Lambda\,\underline{x},\underline{y},\underline{y}''\,.\,\langle\underline{y}''\rangle.
\end{aligned}$$

(c) $A \to A \vee B$:

With $A^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$ and $B^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\psi(\underline{u},\underline{v},\underline{b})$ we have

$$(A \to A \vee B)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{X}',\underline{U},\underline{Y}\forall^{\mathrm{st}}\underline{x},\underline{y}',\underline{v}$$
$$\Big(\forall\underline{z} \in \underline{Y}[\underline{x},\underline{y}',\underline{v}]\varphi(\underline{x},\underline{z},\underline{a}) \to \varphi(\underline{X}'[\underline{x}],\underline{y}',\underline{a}) \vee \psi(\underline{U}[\underline{x}],\underline{v},\underline{b})\Big),$$

and we can take

$$\begin{aligned}
\underline{X}' &:= \Lambda\,\underline{x}\,.\,\underline{x},\\
\underline{Y} &:= \Lambda\,\underline{x},\underline{y}',\underline{v}\,.\,\langle\underline{y}'\rangle,\\
\underline{U} &:= \Lambda\,\underline{x}\,.\,\mathcal{O}.
\end{aligned}$$

(d) $A \wedge B \to A$:

With $A^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$ and $B^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\psi(\underline{u},\underline{v},\underline{b})$ we have

$$(A \wedge B \to A)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{X}',\underline{Y},\underline{V}\forall^{\mathrm{st}}\underline{x},\underline{u},\underline{y}'$$
$$\Big(\forall\underline{z} \in \underline{Y}[\underline{x},\underline{u},\underline{y}']\,\forall\underline{t} \in \underline{V}[\underline{x},\underline{u},\underline{y}']\big(\varphi(\underline{x},\underline{z},\underline{a}) \wedge \psi(\underline{u},\underline{t},\underline{b})\big) \to \varphi(\underline{X}'[\underline{x},\underline{u}],\underline{y}',\underline{a})\Big),$$

and we can take

$$\begin{aligned}
\underline{X}' &:= \Lambda\,\underline{x},\underline{u}\,.\,\underline{x},\\
\underline{Y} &:= \Lambda\,\underline{x},\underline{u},\underline{y}'\,.\,\langle\underline{y}'\rangle,\\
\underline{V} &:= \Lambda\,\underline{x},\underline{u},\underline{y}'\,.\,\mathcal{O}.
\end{aligned}$$

(e) $A \vee B \to B \vee A$:

With $A^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$ and $B^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\psi(\underline{u},\underline{v},\underline{b})$ we have

$$(A \vee B \to B \vee A)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{U}',\underline{X}',\underline{Y},\underline{V}\forall^{\mathrm{st}}\underline{x},\underline{u},\underline{v}',\underline{y}'$$
$$\Big(\forall\underline{z} \in \underline{Y}[\underline{x},\underline{u},\underline{v}',\underline{y}']\,\forall\underline{t} \in \underline{V}[\underline{x},\underline{u},\underline{v}',\underline{y}']\big(\varphi(\underline{x},\underline{z},\underline{a}) \vee \psi(\underline{u},\underline{t},\underline{b})\big)$$
$$\to \big(\psi(\underline{U}'[\underline{x},\underline{u}],\underline{v}',\underline{b}) \vee \varphi(\underline{X}'[\underline{x},\underline{u}],\underline{y}',\underline{a})\big)\Big),$$

and we can take

$$\begin{aligned}
\underline{U}' &:= \Lambda\,\underline{x},\underline{u}\,.\,\underline{u},\\
\underline{X}' &:= \Lambda\,\underline{x},\underline{u}\,.\,\underline{x},\\
\underline{Y} &:= \Lambda\,\underline{x},\underline{u},\underline{v}',\underline{y}'\,.\,\langle\underline{y}'\rangle,\\
\underline{V} &:= \Lambda\,\underline{x},\underline{u},\underline{v}',\underline{y}'\,.\,\langle\underline{v}'\rangle.
\end{aligned}$$

(f) $A \wedge B \to B \wedge A$:

We can take the same terms as for $A \vee B \to B \vee A$, since $\vee$ and $\wedge$ are handled in the same way by the interpretation.

(g) $\bot \to A$:

With $A^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$ we have

$$(\bot \to A)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{X}\forall^{\mathrm{st}}\underline{y}\,\big(\bot \to \varphi(\underline{X},\underline{y},\underline{a})\big),$$

and we can take $\underline{X} := \mathcal{O}$.

(h) $\forall z A \to A[t/z]$:

With $A^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi(\underline{x},\underline{y},z,\underline{a})$ we have

$$(\forall z A \to A[t/z])^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{X}',\underline{Y}\forall^{\mathrm{st}}\underline{x},\underline{y}'$$
$$\Big(\forall \underline{w} \in \underline{Y}[\underline{x},\underline{y}']\forall z\,\varphi(\underline{x},\underline{w},z,\underline{a}) \to \varphi(\underline{X}'[\underline{x}],\underline{y}',t,\underline{a})\Big),$$

and we can take

$$\begin{aligned}
\underline{X}' &:= \Lambda\,\underline{x}\,.\,\underline{x},\\
\underline{Y} &:= \Lambda\,\underline{x},\underline{y}'\,.\,\langle\underline{y}'\rangle.
\end{aligned}$$

(i) $A[t/z] \to \exists z A$:

With $A^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi(\underline{x},\underline{y},z,\underline{a})$ we have

$$(A[t/z] \to \exists z A)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{X}',\underline{Y}\forall^{\mathrm{st}}\underline{x},\underline{y}'$$
$$\Big(\forall \underline{w} \in \underline{Y}[\underline{x},\underline{y}']\,\varphi(\underline{x},\underline{w},t,\underline{a}) \to \exists z\forall \underline{w} \in \underline{y}'\,\varphi(\underline{X}'[\underline{x}],\underline{w},z,\underline{a})\Big),$$

and we can take

$$\begin{aligned}
\underline{X}' &:= \Lambda\,\underline{x}\,.\,\underline{x},\\
\underline{Y} &:= \Lambda\,\underline{x},\underline{y}'\,.\,\underline{y}'.
\end{aligned}$$

(j) The modus ponens rule:

We assume that we have terms $\underline{t}_1$ and $\underline{T}_2,\underline{T}_3$ realizing the interpretations of respectively A and $A \to B$, and we wish to obtain terms $\underline{T}_4$ realizing the interpretation of B. So with $A^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$ and $B^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\psi(\underline{u},\underline{v},\underline{b})$ we have

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathrm{int}} \vdash \forall\underline{y}\,\varphi(\underline{t}_1,\underline{y},\underline{a})$$

and

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathrm{int}} \vdash \forall\underline{x},\underline{v}\,\big(\forall\underline{z} \in \underline{T}_3[\underline{x},\underline{v}]\,\varphi(\underline{x},\underline{z},\underline{a}) \to \psi(\underline{T}_2[\underline{x}],\underline{v},\underline{b})\big).$$

With $\underline{T}_4 := \underline{T}_2[\underline{t}_1]$ we get

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathrm{int}} \vdash \forall\underline{v}\,\psi(\underline{T}_4,\underline{v},\underline{b}),$$

as desired.

(k) The syllogism rule:

We assume that we have terms $\underline{T}_1,\underline{T}_2$ and $\underline{T}_3,\underline{T}_4$ realizing the interpretations of respectively $A \to B$ and $B \to C$, and we wish to obtain terms $\underline{T}_5,\underline{T}_6$ realizing the interpretation of $A \to C$. So with $A^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$, $B^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\psi(\underline{u},\underline{v},\underline{b})$, and $C^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{w}\forall^{\mathrm{st}}\underline{z}\,\chi(\underline{w},\underline{z},\underline{c})$ we have

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathrm{int}} \vdash \forall\underline{x},\underline{v}\,\big(\forall\underline{i} \in \underline{T}_2[\underline{x},\underline{v}]\,\varphi(\underline{x},\underline{i},\underline{a}) \to \psi(\underline{T}_1[\underline{x}],\underline{v},\underline{b})\big) \qquad (1)$$

and

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathrm{int}} \vdash \forall\underline{u},\underline{z}\,\big(\forall\underline{j} \in \underline{T}_4[\underline{u},\underline{z}]\,\psi(\underline{u},\underline{j},\underline{b}) \to \chi(\underline{T}_3[\underline{u}],\underline{z},\underline{c})\big), \qquad (2)$$

and we wish to obtain $\underline{T}_5,\underline{T}_6$ such that

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathrm{int}} \vdash \forall\underline{x},\underline{z}\,\big(\forall\underline{k} \in \underline{T}_6[\underline{x},\underline{z}]\,\varphi(\underline{x},\underline{k},\underline{a}) \to \chi(\underline{T}_5[\underline{x}],\underline{z},\underline{c})\big).$$

To do this we let $\underline{T}_5 := \Lambda\,\underline{x}\,.\,\underline{T}_3[\underline{T}_1[\underline{x}]]$, $\underline{T}_7 := \Lambda\,\underline{x},\underline{z}\,.\,\underline{T}_4[\underline{T}_1[\underline{x}],\underline{z}]$ and apply (2) with $\underline{u} = \underline{T}_1[\underline{x}]$, such that

$$\text{E-HA}^{\omega*} + \Delta_{\text{int}} \vdash \forall \underline{x},\underline{z}\,\big(\forall \underline{j} \in \underline{T}_7[\underline{x},\underline{z}]\,\psi(\underline{T}_1[\underline{x}],\underline{j},\underline{b}) \to \chi(\underline{T}_5[\underline{x}],\underline{z},\underline{c})\big).$$

As a special case of (1) we get

$$\text{E-HA}^{\omega*} + \Delta_{\text{int}} \vdash \forall\underline{x},\underline{z}\forall\underline{j} \in \underline{T}_7[\underline{x},\underline{z}]$$
$$\big(\forall\underline{i} \in \underline{T}_2[\underline{x},\underline{j}]\,\varphi(\underline{x},\underline{i},\underline{a}) \to \psi(\underline{T}_1[\underline{x}],\underline{j},\underline{b})\big),$$

and thus

$$\text{E-HA}^{\omega*} + \Delta_{\text{int}} \vdash \forall\underline{x},\underline{z}$$
$$\big(\forall\underline{j} \in \underline{T}_7[\underline{x},\underline{z}]\,\forall\underline{i} \in \underline{T}_2[\underline{x},\underline{j}]\,\varphi(\underline{x},\underline{i},\underline{a}) \to \chi(\underline{T}_5[\underline{x}],\underline{z},\underline{c})\big).$$

Hence it is enough to construct terms $\underline{T}_6$ such that

$$\underline{T}_6[\underline{x},\underline{z}] =_{\underline{\sigma}} \underline{T}_2[\underline{x},\underline{T}_7[\underline{x},\underline{z}]_{\underline{0}}] *_{\underline{\sigma}} \ldots *_{\underline{\sigma}} \underline{T}_2[\underline{x},\underline{T}_7[\underline{x},\underline{z}]_{|\underline{T}_7[\underline{x},\underline{z}]|-\underline{1}}],$$

where the concatenation is to include $\underline{T}_2[\underline{x},\underline{T}_7[\underline{x},\underline{z}]_{\underline{j}}]$ for each $\underline{j} < |\underline{T}_7[\underline{x},\underline{z}]|$, which one can easily do using terms from $\mathcal{T}^*$.

(l) The importation and exportation rules:

With $A^{D_{\text{st}}} \equiv \exists^{\text{st}}\underline{x}\forall^{\text{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$, $B^{D_{\text{st}}} \equiv \exists^{\text{st}}\underline{u}\forall^{\text{st}}\underline{v}\,\psi(\underline{u},\underline{v},\underline{b})$, and $C^{D_{\text{st}}} \equiv \exists^{\text{st}}\underline{w}\forall^{\text{st}}\underline{z}\,\chi(\underline{w},\underline{z},\underline{c})$ we get

$$(A \wedge B \to C)^{D_{\text{st}}} \equiv \exists^{\text{st}}\underline{W},\underline{Y},\underline{V}\forall^{\text{st}}\underline{x},\underline{u},\underline{z}$$
$$\Big(\forall\underline{i} \in \underline{Y}[\underline{x},\underline{u},\underline{z}]\,\forall\underline{j} \in \underline{V}[\underline{x},\underline{u},\underline{z}]\big(\varphi(\underline{x},\underline{i},\underline{a}) \wedge \psi(\underline{u},\underline{j},\underline{b})\big) \to \chi(\underline{W}[\underline{x},\underline{u}],\underline{z},\underline{c})\Big)$$

and

$$(A \to (B \to C))^{D_{\text{st}}} \equiv \exists^{\text{st}}\underline{W},\underline{Y},\underline{V}\forall^{\text{st}}\underline{x},\underline{u},\underline{z}$$
$$\Big(\forall\underline{i} \in \underline{Y}[\underline{x},\underline{u},\underline{z}]\,\varphi(\underline{x},\underline{i},\underline{a}) \to \big(\forall\underline{j} \in \underline{V}[\underline{x},\underline{u},\underline{z}]\,\psi(\underline{u},\underline{j},\underline{b}) \to \chi(\underline{W}[\underline{x},\underline{u}],\underline{z},\underline{c})\big)\Big),$$

so that the same terms realize the interpretations of $A \wedge B \to C$ and $A \to (B \to C)$.

(m) The expansion rule:

We assume that we have terms $\underline{T}_1,\underline{T}_2$ realizing the interpretation of $A \to B$, and we wish to obtain terms $\underline{T}_3,\underline{T}_4,\underline{T}_5,\underline{T}_6$ realizing the interpretation of $C \vee A \to C \vee B$. So with $A^{D_{\text{st}}} \equiv \exists^{\text{st}}\underline{x}\forall^{\text{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$, $B^{D_{\text{st}}} \equiv \exists^{\text{st}}\underline{u}\forall^{\text{st}}\underline{v}\,\psi(\underline{u},\underline{v},\underline{b})$, and $C^{D_{\text{st}}} \equiv \exists^{\text{st}}\underline{w}\forall^{\text{st}}\underline{z}\,\chi(\underline{w},\underline{z},\underline{c})$ we have

$$\text{E-HA}^{\omega*} + \Delta_{\text{int}} \vdash \forall\underline{x},\underline{v}\,\big(\forall\underline{i} \in \underline{T}_2[\underline{x},\underline{v}]\,\varphi(\underline{x},\underline{i},\underline{a}) \to \psi(\underline{T}_1[\underline{x}],\underline{v},\underline{b})\big),$$

and we want $\underline{T}_3,\underline{T}_4,\underline{T}_5,\underline{T}_6$ such that

$$\text{E-HA}^{\omega*} + \Delta_{\text{int}} \vdash$$
$$\forall\underline{w},\underline{x},\underline{z}',\underline{v}\Big(\forall\underline{j} \in \underline{T}_5[\underline{w},\underline{x},\underline{z}',\underline{v}]\,\forall\underline{i} \in \underline{T}_6[\underline{w},\underline{x},\underline{z}',\underline{v}]\big(\chi(\underline{w},\underline{j},\underline{c}) \vee \varphi(\underline{x},\underline{i},\underline{a})\big)$$
$$\to \chi(\underline{T}_3[\underline{w},\underline{x}],\underline{z}',\underline{c}) \vee \psi(\underline{T}_4[\underline{w},\underline{x}],\underline{v},\underline{b})\Big).$$

Thus we can take

$$
\begin{aligned}
\underline{T}_3 &:= \Lambda\,\underline{w},\underline{x}\,.\,\underline{w}, \\
\underline{T}_4 &:= \Lambda\,\underline{w},\underline{x}\,.\,\underline{T}_1[\underline{x}], \\
\underline{T}_5 &:= \Lambda\,\underline{w},\underline{x},\underline{z}',\underline{v}\,.\,\langle\underline{z}'\rangle, \\
\underline{T}_6 &:= \Lambda\,\underline{w},\underline{x},\underline{z}',\underline{v}\,.\,\underline{T}_2[\underline{x},\underline{v}].
\end{aligned}
$$

(n) The quantifier rules:

    i. We assume that we have terms $\underline{T}_1, \underline{T}_2$ realizing the interpretation of $B \rightarrow A$, and we want terms $\underline{T}_3, \underline{T}_4$ realizing the interpretation of $B \rightarrow \forall z A$, where z is not among the free variables of B. Thus with $A^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi(\underline{x}, \underline{y}, z, \underline{a})$ and $B^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\psi(\underline{u}, \underline{v}, \underline{b})$ we have

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathsf{int}} \vdash \forall \underline{u}, \underline{y}\,\big(\forall \underline{i} \in \underline{T}_2[\underline{u}, \underline{y}]\,\psi(\underline{u}, \underline{i}, \underline{b}) \rightarrow \varphi(\underline{T}_1[\underline{u}], \underline{y}, z, \underline{a})\big),$$

and we want

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathsf{int}} \vdash \forall \underline{u}, \underline{y}\,\big(\forall \underline{i} \in \underline{T}_4[\underline{u}, \underline{y}]\,\psi(\underline{u}, \underline{i}, \underline{b}) \rightarrow \forall z\,\varphi(\underline{T}_3[\underline{u}], \underline{y}, z, \underline{a})\big).$$

Hence we may take $\underline{T}_3 := \underline{T}_1$ and $\underline{T}_4 := \underline{T}_2$.

    ii. We have terms $\underline{T}_1, \underline{T}_2$ realizing the interpretation of $A \rightarrow B$, and we want terms $\underline{T}_3, \underline{T}_4$ realizing the interpretation of $\exists z A \rightarrow B$, where z is not among the free variables of B. Thus with $A^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi(\underline{x}, \underline{y}, z, \underline{a})$ and $B^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\psi(\underline{u}, \underline{v}, \underline{b})$ we have

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathsf{int}} \vdash \forall \underline{x}, \underline{v}\,\big(\forall \underline{i} \in \underline{T}_2[\underline{x}, \underline{v}]\,\varphi(\underline{x}, \underline{i}, z, \underline{a}) \rightarrow \psi(\underline{T}_1[\underline{x}], \underline{v}, \underline{b})\big),$$

and we want

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathsf{int}} \vdash \forall \underline{x}, \underline{v}\,\big(\forall \underline{k} \in \underline{T}_4[\underline{x}, \underline{v}]\,\exists z \forall \underline{i} \in \underline{k}\varphi(\underline{x}, \underline{i}, z, \underline{a}) \rightarrow \psi(\underline{T}_3[\underline{x}], \underline{v}, \underline{b})\big).$$

So we can take $\underline{T}_3 := \underline{T}_1$ and $\underline{T}_4 := \Lambda\,\underline{x}, \underline{v}\,.\,\langle\,\underline{T}_2[\underline{x}, \underline{v}]\,\rangle$.

2. The nonlogical axioms of $\mathsf{E\text{-}HA}^{\omega*}$: These axioms are all internal, and so their D_{st}-interpretations are all realized by the empty tuple of terms.

3. The defining axioms EQ of the external quantifiers:

(a) $\forall^{\mathrm{st}}x\,\Phi(x) \leftrightarrow \forall x(\mathrm{st}(x) \rightarrow \Phi(x))$:
We treat first $\forall^{\mathrm{st}}x\,\Phi(x) \rightarrow \forall x(\mathrm{st}(x) \rightarrow \Phi(x))$. With $\Phi^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\varphi(\underline{u}, \underline{v}, x, \underline{a})$ we get

$$(\forall^{\mathrm{st}}x\,\Phi(x))^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{U}\forall^{\mathrm{st}}x, \underline{v}\,\varphi(\underline{U}[x], \underline{v}, x, \underline{a})$$

and

$$\big(\forall x\,(\mathrm{st}(x) \rightarrow \Phi(x))\big)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{U}'\forall^{\mathrm{st}}z, \underline{v}'\forall x\,\big(x \in z \rightarrow \varphi(\underline{U}'[z], \underline{v}', x, \underline{a})\big),$$

and thus $\big(\forall^{\mathrm{st}}x\,\Phi(x) \rightarrow \forall x\,(\mathrm{st}(x) \rightarrow \Phi(x))\big)^{D_{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}}\tilde{\underline{U}}', X, \underline{V}\,\forall^{\mathrm{st}}\underline{U}, z, \underline{v}'$$
$$\big(\forall i \in X[\underline{U}, z, \underline{v}']\,\forall \underline{j} \in \underline{V}[\underline{U}, z, \underline{v}']\varphi(\underline{U}[i], \underline{j}, i, \underline{a}) \rightarrow \forall x\,\big(x \in z \rightarrow \varphi(\tilde{\underline{U}}'[\underline{U}, z], \underline{v}', x, \underline{a})\big)\big).$$

Let now

$$X \quad := \quad \Lambda\,\underline{U}, z, \underline{v}'\,.\,z,$$
$$\underline{V} \quad := \quad \Lambda\,\underline{U}, z, \underline{v}'\,.\,\langle\underline{v}'\rangle,$$
$$\tilde{\underline{U}}' \quad := \quad \Lambda\,\underline{U}, z\,.\,\underline{U}[z_0] * \ldots * \underline{U}[z_{|z|-1}]$$

(with

$$\underline{U}[z_0] * \ldots * \underline{U}[z_{|z|-1}] := U_0[z_0] * \ldots * U_0[z_{|z|-1}], \ldots, U_{n-1}[z_0] * \ldots * U_n[z_{|z|-1}]$$

for $\underline{U} = U_0, \ldots, U_{n-1}$).

Next we treat $\forall x(\mathrm{st}(x) \to \Phi(x)) \to \forall^{\mathrm{st}}x\,\Phi(x)$. Similarly to the case above we get that $\big(\forall x(\mathrm{st}(x) \to \Phi(x)) \to \forall^{\mathrm{st}}x\,\Phi(x)\big)^{D_{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}}\tilde{\underline{U}}, Z, \underline{V}'\,\forall^{\mathrm{st}}\underline{U}', x, \underline{v}$$

$$\Big(\forall k \in Z[\underline{U}', x, \underline{v}]\,\forall \underline{j} \in \underline{V}'[\underline{U}', x, \underline{v}]\forall x' \in k\,\varphi(\underline{U}'[k], \underline{j}, x', \underline{a}) \to \varphi(\tilde{\underline{U}}[\underline{U}', x], \underline{v}, x, \underline{a})\Big).$$

Hence we can take

$$
\begin{aligned}
Z &:= \Lambda\,\underline{U}', x, \underline{v}\,.\,\langle\langle x\rangle\rangle,\\
\underline{V}' &:= \Lambda\,\underline{U}', x, \underline{v}\,.\,\langle\underline{v}\rangle,\\
\tilde{\underline{U}} &:= \Lambda\,\underline{U}', x\,.\,\underline{U}'[\langle x\rangle].
\end{aligned}
$$

(b) $\exists^{\mathrm{st}}x\,\Phi(x) \leftrightarrow \exists x(\mathrm{st}(x) \wedge \Phi(x))$:

First we treat $\exists^{\mathrm{st}}x\,\Phi(x) \to \exists x(\mathrm{st}(x)\wedge\Phi(x))$. With $\Phi^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\varphi(\underline{u}, \underline{v}, x, \underline{a})$ we get

$$(\exists^{\mathrm{st}}x\,\Phi(x))^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}, x\forall^{\mathrm{st}}\underline{v}\exists i \in x\forall \underline{j} \in \underline{v}\,\varphi(\underline{u}, \underline{j}, i, \underline{a})$$

and

$$\big(\exists x\,(\mathrm{st}(x) \wedge \Phi(x))\big)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}z, \underline{u}'\forall^{\mathrm{st}}\underline{v}'\exists x'\forall \underline{j} \in \underline{v}'\,\big(x' \in z \wedge \varphi(\underline{u}', \underline{j}, x', \underline{a})\big),$$

and thus $\big(\exists^{\mathrm{st}}x\,\Phi(x) \to \exists x(\mathrm{st}(x) \wedge \Phi(x))\big)^{D_{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}}Z, \underline{U}', \underline{V}\,\forall^{\mathrm{st}}\underline{u}, x, \underline{v}'$$

$$\Big(\forall \underline{l} \in \underline{V}[\underline{u}, x, \underline{v}']\,\exists k \in x\,\forall \underline{n} \in \underline{l}\,\varphi(\underline{u}, \underline{n}, k, \underline{a}) \to$$

$$\exists x'\,\forall \underline{j} \in \underline{v}'\big(x' \in Z[\underline{u}, x] \wedge \varphi(\underline{U}'[\underline{u}, x], \underline{j}, x', \underline{a})\big)\Big).$$

Thus we can take

$$
\begin{aligned}
Z &:= \Lambda\,\underline{u}, x\,.\,x,\\
\underline{U}' &:= \Lambda\,\underline{u}, x\,.\,\underline{u},\\
\underline{V} &:= \Lambda\,\underline{u}, x, \underline{v}'\,.\,\langle\underline{v}'\rangle.
\end{aligned}
$$

Finally we consider $\exists x(\mathrm{st}(x)\wedge\Phi(x)) \to \exists^{\mathrm{st}}x\,\Phi(x)$. Similarly to the case above we get that $\big(\exists x(\mathrm{st}(x) \wedge \Phi(x)) \to \exists^{\mathrm{st}}x\,\Phi(x)\big)^{D_{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}}\underline{U}, X, \underline{V}'\,\forall^{\mathrm{st}}z, \underline{u}', \underline{v}$$

$$\Big(\forall \underline{l} \in \underline{V}'[z, \underline{u}', \underline{v}]\,\exists x'\,\forall \underline{j} \in \underline{l}\big(x' \in z \wedge \varphi(\underline{u}', \underline{j}, x', \underline{a})\big) \to$$

$$\exists k \in X[z, \underline{u}']\,\forall \underline{n} \in \underline{v}\,\varphi(\underline{U}[z, \underline{u}'], \underline{n}, k, \underline{a})\Big).$$

Let now

$$
\begin{aligned}
X &:= \Lambda\,z, \underline{u}'\,.\,z,\\
\underline{U} &:= \Lambda\,z, \underline{u}'\,.\,\underline{u}',\\
\underline{V}' &:= \Lambda\,z, \underline{u}', \underline{v}\,.\,\langle\underline{v}\rangle.
\end{aligned}
$$

4. The schemata $\mathcal{T}^*_{\mathrm{st}}$:

(a) $\mathrm{st}(x) \wedge x = y \to \mathrm{st}(y)$: We have that $\big(\mathrm{st}(x) \wedge x = y \to \mathrm{st}(y)\big)^{D_{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}}V\forall^{\mathrm{st}}u\,\big(x \in u \wedge x = y \to y \in V[u]\big),$$

and so we can take $V := \Lambda\,u\,.\,u$.

(b) $\mathrm{st}(t)$ for closed terms t in $\mathcal{T}^*$: Since

$$\big(\mathrm{st}(t)\big)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}} x\, t \in x,$$

we can take $x := \langle t \rangle$.

(c) $\mathrm{st}(f^{\sigma \to \tau}) \wedge \mathrm{st}(x^{\sigma}) \to \mathrm{st}(f(x))$: We have that $\big(\mathrm{st}(f) \wedge \mathrm{st}(x) \to \mathrm{st}(f(x))\big)^{D_{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}} W \forall^{\mathrm{st}} u, v\big(f \in u \wedge x \in v \to f(x) \in W[u, v]\big),$$

hence it is enough to construct a closed term W such that

$$W[u, v] =_{\tau^*} \langle u_i(v_j) : i < |u|, j < |v| \rangle,$$

and this we can do easily using closed terms from $\mathcal{T}^*$.

5. The external induction axiom $\mathsf{IA}^{\mathrm{st}}$: We will consider the equivalent external induction rule

$$\mathsf{IR}^{\mathrm{st}} : \quad \frac{\Phi(0),\ \forall^{\mathrm{st}} n^0(\Phi(n) \to \Phi(n + 1))}{\forall^{\mathrm{st}} n^0 \Phi(n)},$$

from which one can derive $\mathsf{IA}^{\mathrm{st}}$ by taking

$$\Phi(m^0) :\equiv \Psi(0) \wedge \forall^{\mathrm{st}} n^0(\Psi(n) \to \Psi(n + 1)) \to \Psi(m).$$

We assume that we have terms $\underline{T}_1$ and $\underline{T}_2, \underline{T}_3$ realizing the interpretations of $\Phi(0)$ and $\forall^{\mathrm{st}} n^0(\Phi(n) \to \Phi(n + 1))$, and we wish to obtain terms $\underline{T}_4$ realizing the interpretation of $\forall^{\mathrm{st}} n^0 \Phi(n)$. So with $\big(\Phi(n)\big)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}} \underline{x} \forall^{\mathrm{st}} \underline{y}\, \varphi(\underline{x}, \underline{y}, n, \underline{a})$ we have

$$\mathsf{E\text{-}HA}^{\omega *} + \Delta_{\mathrm{int}} \vdash \forall \underline{y}\, \varphi(\underline{T}_1, \underline{y}, 0, \underline{a})$$

and

$$\mathsf{E\text{-}HA}^{\omega *} + \Delta_{\mathrm{int}} \vdash \forall n^0, \underline{x}, \underline{y}'$$

$$\Big(\forall \underline{i} \in \underline{T}_3[n, \underline{x}, \underline{y}']\, \varphi(\underline{x}, \underline{i}, n, \underline{a}) \to \varphi(\underline{T}_2[n, \underline{x}], \underline{y}', n + 1, \underline{a})\Big),$$

and we want

$$\mathsf{E\text{-}HA}^{\omega *} + \Delta_{\mathrm{int}} \vdash \forall n^0, \underline{y}\, \varphi(\underline{T}_4[n], \underline{y}, n, \underline{a}). \tag{3}$$

By taking $\underline{T}_4 := \Lambda\, n^0 . \underline{\mathcal{R}}(n, \underline{T}_1, \underline{T}_4)$ we get

$$\mathsf{E\text{-}HA}^{\omega *} + \Delta_{\mathrm{int}} \quad \vdash \quad \underline{T}_4[0] =_{\rho} \underline{T}_1$$
$$\mathsf{E\text{-}HA}^{\omega *} + \Delta_{\mathrm{int}} \quad \vdash \quad \underline{T}_4[n + 1] =_{\rho} \underline{T}_2[n, \underline{T}_4[n]],$$

which suffices to establish (3).

6. The principles I, NCR, HAC, $\mathsf{HIP}_{\forall^{\mathrm{st}}}$, and $\mathsf{HGMP}^{\mathrm{st}}$:

(a) I: The D_{st}-interpretations of the premise and the conclusion of any instance of I are identical, and it is easy to show that $\Lambda \underline{x} . \langle \underline{x} \rangle$ realizes the interpretation of the whole implication, provably in $\mathsf{E\text{-}HA}^{\omega *}$.

(b) NCR: Suppose $\big(\Phi(x, y)\big)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}} \underline{u} \forall^{\mathrm{st}} \underline{v}\, \varphi(\underline{u}, \underline{v}, x, y)$. Then

$$\big(\forall x \exists^{\mathrm{st}} y\, \Phi(x, y)\big)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}} \underline{u}, y \forall^{\mathrm{st}} \underline{v} \forall x\, \exists i \in y\, \forall \underline{j} \in \underline{v}\, \varphi\big(\underline{u}, \underline{j}, x, i\big)$$

and

$$\big(\exists^{\mathrm{st}} y \forall x \exists k \in y\, \Phi(x, k)\big)^{D_{\mathrm{st}}} \equiv$$
$$\exists^{\mathrm{st}} \underline{u}, y \forall^{\mathrm{st}} \underline{v} \exists m \in y \forall \underline{n} \in \underline{v} \forall x \exists k \in m \forall \underline{l} \in \underline{n}\, \varphi\big(\underline{u}, \underline{l}, x, k\big),$$

and so $\left(\forall x \exists^{\mathrm{st}} y\,\Phi(x,y) \to \exists^{\mathrm{st}} y \forall x \exists k \in y\,\Phi(x,k)\right)^{D_{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}}\underline{U}, Y, \underline{V}\,\forall^{\mathrm{st}}\underline{u}, y, \underline{v}$$

$$\left(\forall \underline{i} \in \underline{V}[\underline{u}, y, \underline{v}]\,\forall x\,\exists j \in y \forall \underline{k} \in \underline{i}\,\varphi(\underline{u}, \underline{k}, x, j) \to\right.$$

$$\left.\exists m \in Y[\underline{u}, y]\,\forall \underline{n} \in \underline{v}\,\forall \tilde{x} \exists \tilde{k} \in m\,\forall \underline{l} \in \underline{n}\,\varphi(\underline{U}[\underline{u}, y], \underline{l}, \tilde{x}, \tilde{k})\right).$$

Thus we can take

$$\begin{aligned}
\underline{U} &:= \Lambda \underline{u}, y \,.\, \underline{u}, \\
Y &:= \Lambda \underline{u}, y \,.\, \langle y \rangle, \\
\underline{V} &:= \Lambda \underline{u}, y, \underline{v} \,.\, \underline{v}.
\end{aligned}$$

(c) HAC: Let $\left(\Phi(x,y)\right)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\varphi(\underline{u}, \underline{v}, x, y)$. Then $\left(\forall^{\mathrm{st}} x \exists^{\mathrm{st}} y\,\Phi(x,y)\right)^{D_{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}}\underline{U}, Y\forall^{\mathrm{st}} x, \underline{v}\,\exists i \in Y[x]\forall \underline{j} \in \underline{v}\,\varphi(\underline{U}[x], \underline{j}, x, i)$$

and $\left(\exists^{\mathrm{st}} F \forall^{\mathrm{st}} x \exists i \in F(x)\,\Phi(x,i)\right)^{D_{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}}\underline{\tilde{U}}, F\forall^{\mathrm{st}}\tilde{x}, \underline{\tilde{v}}\,\exists k \in F\forall l \in \tilde{x}\forall \underline{\tilde{m}} \in \underline{\tilde{v}}\,\exists i' \in k(l)\forall \underline{j}' \in \underline{\tilde{m}}\,\varphi(\underline{\tilde{U}}[l], \underline{j}', l, i'),$$

so $\left(\forall^{\mathrm{st}} x \exists^{\mathrm{st}} y\,\Phi(x,y) \to \exists^{\mathrm{st}} F \forall^{\mathrm{st}} x \exists i \in F(x)\Phi(x,i)\right)^{D_{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}}\underline{\tilde{U}}, F, X, \underline{V}\,\forall^{\mathrm{st}}\underline{U}, Y, \tilde{x}, \underline{\tilde{v}}$$

$$\left(\forall n \in X[\underline{U}, Y, \tilde{x}, \underline{\tilde{v}}]\forall \underline{n}' \in \underline{V}[\underline{U}, Y, \tilde{x}, \underline{\tilde{v}}]\exists i \in Y[n]\forall \underline{j} \in \underline{n}'\,\varphi(\underline{U}[n], \underline{j}, n, i) \to\right.$$

$$\left.\exists k \in F[\underline{U}, Y]\,\forall l \in \tilde{x}\,\forall \underline{\tilde{m}} \in \underline{\tilde{v}}\,\exists i' \in k(l)\,\forall \underline{j}' \in \underline{\tilde{m}}\,\varphi(\underline{\tilde{U}}[\underline{U}, Y][l], \underline{j}', l, i')\right).$$

Hence we can take

$$\begin{aligned}
\underline{\tilde{U}} &:= \Lambda \underline{U}, Y \,.\, \underline{U}, \\
F &:= \Lambda \underline{U}, Y \,.\, \langle \lambda x.Y[x] \rangle, \\
X &:= \Lambda \underline{U}, Y, \tilde{x}, \underline{\tilde{v}} \,.\, \tilde{x}, \\
\underline{V} &:= \Lambda \underline{U}, Y, \tilde{x}, \underline{\tilde{v}} \,.\, \underline{\tilde{v}}.
\end{aligned}$$

(d) HIP$_{\forall^{\mathrm{st}}}$: Let $\left(\Psi(y)\right)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\psi(\underline{u}, \underline{v}, y)$. Then $\left(\forall^{\mathrm{st}} x\,\varphi(x) \to \exists^{\mathrm{st}} y \Psi(y)\right)^{D_{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}}\underline{U}, Y, X\forall^{\mathrm{st}}\underline{v}\left(\forall k \in X[\underline{v}]\,\varphi(k) \to \exists i \in Y\forall \underline{j} \in \underline{v}\,\psi(\underline{U}, \underline{j}, i)\right)$$

and $\left(\exists^{\mathrm{st}} y\left(\forall^{\mathrm{st}} x\,\varphi(x) \to \exists i \in y\Psi(i)\right)\right)^{D_{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}}\underline{U}, X, y\forall^{\mathrm{st}}\underline{v}\exists n \in y\forall \underline{m} \in \underline{v}\left(\forall k \in X[\underline{m}]\,\varphi(k) \to \exists i \in n\,\forall \underline{j} \in \underline{m}\,\psi(\underline{U}, \underline{j}, i)\right),$$

so the D_{st}-interpretation of HIP$_{\forall^{\mathrm{st}}}$ is

$$\exists^{\mathrm{st}}\underline{\tilde{U}}, \tilde{X}, \tilde{Y}, \underline{V}\forall^{\mathrm{st}}\underline{U}, Y, X, \underline{\tilde{v}}$$

$$\left(\forall \underline{l} \in \underline{V}[\underline{U}, Y, X, \underline{\tilde{v}}]\left(\forall k \in X[\underline{l}]\,\varphi(k) \to \exists i \in Y\,\forall \underline{j} \in \underline{l}\,\psi(\underline{U}, \underline{j}, i)\right) \to\right.$$

$$\exists \tilde{n} \in \tilde{Y}[\underline{U}, Y, X]\forall \underline{\tilde{m}} \in \underline{\tilde{v}}\left(\forall \tilde{k} \in \tilde{X}[\underline{U}, Y, X][\underline{\tilde{m}}]\varphi(\tilde{k}) \to\right.$$

$$\left.\left.\exists \tilde{i} \in \tilde{n}\,\forall \underline{\tilde{j}} \in \underline{\tilde{m}}\,\psi(\underline{\tilde{U}}[\underline{U}, Y, X], \underline{\tilde{j}}, \tilde{i})\right)\right).$$

Hence we can take

$$\begin{aligned}
\underline{\tilde{U}} &:= \Lambda \underline{U}, Y, X \,.\, \underline{U}, \\
\tilde{X} &:= \Lambda \underline{U}, Y, X \,.\, X, \\
\tilde{Y} &:= \Lambda \underline{U}, Y, X \,.\, \langle Y \rangle, \\
\underline{V} &:= \Lambda \underline{U}, Y, X, \underline{\tilde{v}} \,.\, \underline{\tilde{v}}.
\end{aligned}$$

(e) HGMP$^{\mathrm{st}}$: The D_{st}-interpretation of HGMP$^{\mathrm{st}}$ is

$$\exists^{\mathrm{st}}X\forall^{\mathrm{st}}x\left(\left(\forall x'\in x\,\varphi(x')\to\psi\right)\to\exists j\in X[x]\big(\forall k\in j\,\varphi(k)\to\psi\big)\right),$$

and we can take $X:=\Lambda\,x\,.\,\langle x\rangle$.

$\square$

Theorem 5.5 implies the following conservation result, which improves on Moerdijk and Palmgren [37] and Avigad and Helzner [5].

Corollary 5.6. *The system*

$$\mathsf{E\text{-}HA}^{\omega *}_{\mathrm{st}}+\mathsf{I}+\mathsf{NCR}+\mathsf{HAC}+\mathsf{HGMP}^{\mathrm{st}}+\mathsf{HIP}_{\forall^{\mathrm{st}}}$$

is a conservative extension of $\mathsf{E\text{-}HA}^{\omega *}$ *and hence of* $\mathsf{E\text{-}HA}^{\omega}$.

Proof. By Theorem 5.5 and Lemma 5.3.

$\square$

Remark 5.7. We could define a system $\mathsf{E\text{-}HA}^{\omega *}_{\mathrm{nst}}$ by adding primitive predicates nst^{σ} ("nonstandard") to $\mathsf{E\text{-}HA}^{\omega *}_{\mathrm{st}}$ for each finite type σ, along with axioms

$$\forall x^{\sigma}\big(\mathrm{nst}(x)\leftrightarrow\neg\,\mathrm{st}(x)\big).$$

If we then extend the D_{st}-interpretation by

$$\big(\mathrm{nst}^{\sigma}(x^{\sigma})\big)^{D_{\mathrm{st}}}:\equiv\forall^{\mathrm{st}}y^{\sigma}y\neq_{\sigma}x),$$

we get an analogue of Theorem 5.5, since $\big(\mathrm{nst}(x)\to\neg\,\mathrm{st}(x)\big)^{D_{\mathrm{st}}}$ is provably equivalent to

$$\exists^{\mathrm{st}}Y\forall^{\mathrm{st}}z\,(\forall y\in Y[z](y\neq x)\to x\notin z)$$

and $\big(\neg\,\mathrm{st}(x)\to\mathrm{nst}(x)\big)^{D_{\mathrm{st}}}$ to

$$\exists^{\mathrm{st}}Z\forall^{\mathrm{st}}y\,(\forall z'\in Z[y]x\notin z'\to y\neq x)\,,$$

so that we can take $Y[z]:=z$ and $Z[y]:=\langle\langle y\rangle\rangle$ respectively .

5.2 The characteristic principles of the nonstandard functional interpretation

In this section we will prove that the characteristic principles of the nonstandard functional interpretation are I, NCR, HAC, HIP$_{\forall^{\mathrm{st}}}$, and HGMP$^{\mathrm{st}}$. For notational simplicity we will let

$$\mathsf{H}:=\mathsf{E\text{-}HA}^{\omega *}_{\mathrm{st}}+\mathsf{I}+\mathsf{NCR}+\mathsf{HAC}+\mathsf{HIP}_{\forall^{\mathrm{st}}}+\mathsf{HGMP}^{\mathrm{st}}.$$

Theorem 5.8 (Characterization theorem for the nonstandard functional interpretation)**.**

1. *For any formula Φ in the language of* $\mathsf{E\text{-}HA}^{\omega *}_{\mathrm{st}}$ *we have*

$$\mathsf{H}\vdash\Phi\leftrightarrow\Phi^{D_{\mathrm{st}}}.$$

2. *For any formula Ψ in the language of* $\mathsf{E\text{-}HA}^{\omega *}_{\mathrm{st}}$ *we have: If for all Φ in* $\mathcal{L}(\mathsf{E\text{-}HA}^{\omega *}_{\mathrm{st}})$ *(with $\Phi^{D_{\mathrm{st}}}\equiv\exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\varphi_{D_{\mathrm{st}}}\,(\underline{x},\underline{y})$) the implication*

$$\mathsf{H}+\Psi\vdash\Phi\quad\Longrightarrow\quad\text{there are closed terms }\underline{t}\in\mathcal{T}^*\text{ s.t. }\mathsf{E\text{-}HA}^{\omega}\vdash\forall\underline{y}\,\varphi_{D_{\mathrm{st}}}\,(\underline{t},\underline{y})\quad(4)$$

holds, then $\mathsf{H}\vdash\Psi$.

Proof.

1. We will prove item 1 by induction on the logical structure of Φ.

   (a) As induction start we note that for internal atomic φ we obviously have $\mathsf{H} \vdash \varphi \leftrightarrow \varphi^{D_{\mathrm{st}}}$, that
$$\mathsf{H} \vdash \mathrm{st}(u^\sigma) \to \exists^{\mathrm{st}} x^{\sigma^*} u \in x$$
   follows by taking $x := \langle u \rangle$, and that
$$\mathsf{H} \vdash \exists^{\mathrm{st}} x^{\sigma^*} u \in x \to \mathrm{st}(u^\sigma)$$
   is Lemma 2.11.

   (b) For the induction step involving Φ, Ψ we will use that (see 1.6.17 in [48]) via appropriate embeddings of tuples of types in a suitable common higher type and tuple coding of functionals with inverses (all given by terms in $\mathcal{T}^*$) there are $\varphi(x, y)$, $\psi(u, v)$ such that
$$\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*} \vdash \Phi^{D_{\mathrm{st}}} \leftrightarrow \exists^{\mathrm{st}} x \forall^{\mathrm{st}} y \, \varphi(x, y)$$
   and
$$\mathsf{E\text{-}HA}_{\mathrm{st}}^{\omega*} \vdash \Psi^{D_{\mathrm{st}}} \leftrightarrow \exists^{\mathrm{st}} u \forall^{\mathrm{st}} v \, \psi(u, v),$$
   and such that $\varphi(x, y)$ and $\psi(u, v)$ are provably upwards closed in x and u respectively.

      i. For $\wedge$ we must consider
$$\exists^{\mathrm{st}} x \forall^{\mathrm{st}} y \, \varphi(x, y) \wedge \exists^{\mathrm{st}} u \forall^{\mathrm{st}} v \, \psi(u, v) \leftrightarrow \exists^{\mathrm{st}} x, u \forall^{\mathrm{st}} y, v \, \big(\varphi(x, y) \wedge \psi(u, v)\big),$$
      which follows by intuitionistic logic. (We can assume that u, v do not appear in φ, and that x, y do not appear in ψ.)

      ii. For $\vee$ we note that
$$\exists^{\mathrm{st}} x \forall^{\mathrm{st}} y \, \varphi(x, y) \vee \exists^{\mathrm{st}} u \forall^{\mathrm{st}} v \, \psi(u, v) \leftrightarrow \exists^{\mathrm{st}} x, u \forall^{\mathrm{st}} y, v \, \big(\varphi(x, y) \vee \psi(u, v)\big)$$
      follows from $\mathsf{LLPO}^{\mathrm{st}}$, which by Proposition 3.4 follows from I.

      iii. For implication we use that $\psi(u, v)$ is upwards closed in u to conclude

$$
\begin{array}{ll}
\exists^{\mathrm{st}} x \forall^{\mathrm{st}} y \, \varphi(x, y) \to \exists^{\mathrm{st}} u \forall^{\mathrm{st}} v \, \psi(u, v) & \leftrightarrow \\[4pt]
\forall^{\mathrm{st}} x \, \big(\forall^{\mathrm{st}} y \, \varphi(x, y) \to \exists^{\mathrm{st}} u \forall^{\mathrm{st}} v \, \psi(u, v)\big) & \overset{\mathsf{HIP}_{\forall^{\mathrm{st}}}}{\leftrightarrow} \\[4pt]
\forall^{\mathrm{st}} x \exists^{\mathrm{st}} u \, \big(\forall^{\mathrm{st}} y \, \varphi(x, y) \to \forall^{\mathrm{st}} v \, \psi(u, v)\big) & \leftrightarrow \\[4pt]
\forall^{\mathrm{st}} x \exists^{\mathrm{st}} u \forall^{\mathrm{st}} v \, \big(\forall^{\mathrm{st}} y \, \varphi(x, y) \to \psi(u, v)\big) & \overset{\mathsf{HGMP}^{\mathrm{st}}}{\leftrightarrow} \\[4pt]
\forall^{\mathrm{st}} x \exists^{\mathrm{st}} u \forall^{\mathrm{st}} v \exists^{\mathrm{st}} y \, \big(\forall i \in y \, \varphi(x, i) \to \psi(u, v)\big) & \overset{\mathsf{HAC}}{\leftrightarrow} \\[4pt]
\exists^{\mathrm{st}} U \forall^{\mathrm{st}} x \forall^{\mathrm{st}} v \exists^{\mathrm{st}} y \, \big(\forall i \in y \, \varphi(x, i) \to \psi(U(x), v)\big) & \overset{\mathsf{HAC}+\text{ coding } x, v \text{ into one}}{\leftrightarrow} \\[4pt]
\exists^{\mathrm{st}} U, Y \forall^{\mathrm{st}} x, v \, \big(\forall i \in Y(x, v) \, \varphi(x, i) \to \psi(U(x), v)\big) & \leftrightarrow \\[4pt]
\exists^{\mathrm{st}} U, Y \forall^{\mathrm{st}} x, v \, \big(\forall i \in Y[x, v] \, \varphi(x, i) \to \psi(U[x], v)\big). &
\end{array}
$$

      iv. For $\forall$ we use NCR and that $\varphi(x, y, z)$ is upwards closed in x to get
$$\forall z \exists^{\mathrm{st}} x \forall^{\mathrm{st}} y \, \varphi(x, y, z) \overset{\mathsf{NCR}}{\leftrightarrow} \exists^{\mathrm{st}} x \forall z \forall^{\mathrm{st}} y \, \varphi(x, y, z) \leftrightarrow \exists^{\mathrm{st}} x \forall^{\mathrm{st}} y \forall z \, \varphi(x, y, z).$$

      v. For $\exists$ we use I:
$$
\begin{array}{ll}
\exists z \exists^{\mathrm{st}} x \forall^{\mathrm{st}} y \, \varphi(x, y, z) & \leftrightarrow \\[4pt]
\exists^{\mathrm{st}} x \exists z \forall^{\mathrm{st}} y \, \varphi(x, y, z) & \overset{\mathsf{I}}{\leftrightarrow} \\[4pt]
\exists^{\mathrm{st}} x \forall^{\mathrm{st}} y \exists z \forall y' \in y \, \varphi(x, y', z). &
\end{array}
$$

vi. For $\forall^{\mathrm{st}}$ we use HAC and that $\varphi(x,y,z)$ is upwards closed in x:

$$\forall^{\mathrm{st}}z\exists^{\mathrm{st}}x\forall^{\mathrm{st}}y\,\varphi(x,y,z) \overset{\mathsf{HAC}}{\leftrightarrow} \exists^{\mathrm{st}}X\forall^{\mathrm{st}}z,y\,\varphi(X(z),y,z) \leftrightarrow \exists^{\mathrm{st}}X\forall^{\mathrm{st}}z,y\,\varphi(X[z],y,z).$$

vii. For $\exists^{\mathrm{st}}$ we will again use I:

$$
\begin{aligned}
\exists^{\mathrm{st}}z\big(\exists^{\mathrm{st}}x\forall^{\mathrm{st}}y\,\varphi(x,y,z)\big) &\leftrightarrow \\
\exists^{\mathrm{st}}z\exists z'\in z\big(\exists^{\mathrm{st}}x\forall^{\mathrm{st}}y\,\varphi(x,y,z')\big) &\leftrightarrow \\
\exists^{\mathrm{st}}z\,\exists^{\mathrm{st}}x\,\exists z'\in z\,\forall^{\mathrm{st}}y\,\varphi(x,y,z') &\overset{\mathsf{I}}{\leftrightarrow} \\
\exists^{\mathrm{st}}z\,\exists^{\mathrm{st}}x\,\forall^{\mathrm{st}}y\,\exists z'\in z\,\forall y'\in y\,\varphi(x,y',z'). &
\end{aligned}
$$

2. Assume that Ψ is a formula of $\mathcal{L}(\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}})$ such that implication (4) holds, and let $\Psi^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\psi_{D_{\mathrm{st}}}(\underline{x},\underline{y})$. Then because of item 1 we have

$$
\begin{gathered}
\mathsf{H} + \Psi \vdash \Psi \\
\Longrightarrow \\
\exists\underline{t}\in\mathcal{T}^* \text{ s.t. } \mathsf{E\text{-}HA}^{\omega*}\vdash \forall\underline{y}\,\psi_{D_{\mathrm{st}}}(\underline{t},\underline{y}) \\
\Longrightarrow \\
\exists\underline{t}\in\mathcal{T}^* \text{ s.t. } \mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}\vdash \forall\underline{y}\,\psi_{D_{\mathrm{st}}}(\underline{t},\underline{y}) \\
\Longrightarrow \\
\exists\underline{t}\in\mathcal{T}^* \text{ s.t. } \mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}\vdash \forall^{\mathrm{st}}\underline{y}\,\psi_{D_{\mathrm{st}}}(\underline{t},\underline{y}) \\
\Longrightarrow \\
\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}\vdash \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\psi_{D_{\mathrm{st}}}(\underline{x},\underline{y}) \\
\Longrightarrow \\
\mathsf{H} \vdash \exists^{\mathrm{st}}\underline{x}\forall^{\mathrm{st}}\underline{y}\,\psi_{D_{\mathrm{st}}}(\underline{x},\underline{y}) \\
\Longrightarrow \\
\mathsf{H} \vdash \Psi.
\end{gathered}
$$

$\square$

Theorem 5.5 allows us to extract a finite sequence of candidates for the existential quantifier in formulas of the form $\forall^{\mathrm{st}}x\,\exists^{\mathrm{st}}y\,\varphi(x,y)$, in the following sense:

Theorem 5.9 (Main theorem on program extraction by the D_{st}-interpretation)**.** *Let $\forall^{\mathrm{st}}x\exists^{\mathrm{st}}y\,\varphi(x,y)$ be a sentence of $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ with $\varphi(x,y)$ an internal formula, and let Δ_{int} be a set of internal sentences. If*

$$\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} + \mathsf{I} + \mathsf{NCR} + \mathsf{HAC} + \mathsf{HGMP}^{\mathrm{st}} + \mathsf{HIP}_{\forall^{\mathrm{st}}} + \Delta_{\mathsf{int}} \vdash \forall^{\mathrm{st}}x\,\exists^{\mathrm{st}}y\,\varphi(x,y),$$

then from the proof we can extract a closed term t in $\mathcal{T}^*$ such that

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathsf{int}} \vdash \forall x\,\exists y\in t(x)\,\varphi(x,y).$$

Proof. Since
$$\big(\forall^{\mathrm{st}}x\exists^{\mathrm{st}}y\,\varphi(x,y)\big)^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}Y\forall^{\mathrm{st}}x\exists y\in Y[x]\,\varphi(x,y)$$
it follows from the soundness theorem of the D_{st}-interpretation that there is a closed term s such that
$$\mathsf{E\text{-}HA}^{\omega*} + \Delta_{\mathsf{int}} \vdash \forall x\exists y\in s[x]\,\varphi(x,y),$$
and so we can let $t := \lambda x\,.\,s[x]$. $\square$

Remark 5.10. Probably it is clear by now that our functional interpretation has some striking similarities with the bounded functional interpretation due to Ferreira and Oliva [12]. Also there the authors work with two types of quantifiers, get an interpretation whose matrix is upwards closed in the first component (albeit with respect to a different ordering), interpret implications *à la* Diller-Nahm and have some similar looking characteristic principles, like a monotone axiom of choice. But, still, the precise relationship is not entirely clear to us, because we are now comparing the external quantifiers with the unbounded quantifiers in the bounded functional interpretation, which is not very natural. These issues deserve to be further investigated.

5.3 Discussion

It follows from the soundness of the D_{st}-interpretation (Theorem 5.5) that it can be used to eliminate nonstandard principles, like overspill, realization and idealization, from proofs. It also allows one to eliminate underspill, since we have the following result (recall that R is the realization principle from Section 4.1):

Proposition 5.11. *We have*

$$\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} + \mathsf{R} + \mathsf{HGMP}^{\mathrm{st}} \vdash \mathsf{US},$$

and therefore the underspill principle US *is eliminated by the D_{st}-interpretation.*

Proof. We reason in $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} + \mathsf{R} + \mathsf{HGMP}^{\mathrm{st}}$. Assume $\forall x^\sigma (\neg \mathrm{st}(x) \to \varphi(x))$. Our aim is to find a standard x such that $\varphi(x)$ holds.

$\forall x^\sigma (\neg \mathrm{st}(x) \to \varphi(x))$ is equivalent to $\forall x^\sigma (\forall^{\mathrm{st}} y^\sigma (y \neq x) \to \varphi(x))$, which, using $\mathsf{HGMP}^{\mathrm{st}}$, we can rewrite as

$$\forall x^\sigma \exists^{\mathrm{st}} y^{\sigma*} (\forall y' \in y(y' \neq x) \to \varphi(x)).$$

It now follows from R that there is a standard y of type σ^* such that

$$\forall x^\sigma (\forall y' \in y(y' \neq x) \to \varphi(x)).$$

So if we choose x to be a standard element of type σ different from all $y' \in y$, then $\varphi(x)$ will hold. $\qquad\square$

We also have:

Proposition 5.12. *The system* $\mathsf{H} :\equiv \mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} + \mathsf{I} + \mathsf{NCR} + \mathsf{HAC} + \mathsf{HGMP}^{\mathrm{st}} + \mathsf{HIP}_{\forall^{\mathrm{st}}}$ *is closed under both transfer rules,* $\mathsf{TR}_\forall$ *and* $\mathsf{TR}_\exists$.

Proof. If $\mathsf{H} \vdash \forall^{\mathrm{st}} x\, \varphi(x)$, then $\mathsf{E\text{-}HA}^{\omega*} \vdash \forall x\, \varphi(x)$ by soundness of the D_{st}-interpretation. Since $\mathsf{E\text{-}HA}^{\omega*}$ is a subsystem of H, it follows that $\mathsf{H} \vdash \forall x\, \varphi(x)$. This shows that H is closed under $\mathsf{TR}_\forall$.

If $\mathsf{H} \vdash \exists x\, \varphi(x)$, then $\mathsf{E\text{-}HA}^{\omega*} \vdash \exists x\, \varphi(x)$ by conservativity of H over $\mathsf{E\text{-}HA}^{\omega*}$. Since $\mathsf{E\text{-}HA}^{\omega*}$ has the existence property (the usual argument for the existence property of $\mathsf{E\text{-}HA}^{\omega}$, as in [25, Corollary 5.24], for instance, carries over to $\mathsf{E\text{-}HA}^{\omega*}$), it follows that there is a term t in $\mathcal{T}^*$ such that $\mathsf{E\text{-}HA}^{\omega*} \vdash \varphi(t)$. Again, because $\mathsf{E\text{-}HA}^{\omega*}$ is a subsystem of H, we have $\mathsf{H} \vdash \varphi(t)$ as well, and since all terms from $\mathcal{T}^*$ are provably standard in H, we have $\mathsf{H} \vdash \exists^{\mathrm{st}} x\, \varphi(x)$. This shows that H is closed under $\mathsf{TR}_\exists$ as well. $\qquad\square$

Therefore our functional interpretation D_{st} meets all the benchmarks that we discussed in Section 4.

6 The system $\mathbf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$ and negative translation

By combining the functional interpretation from the previous section with negative translation we can obtain conservation and term extraction results for classical systems as well. We will work out the details in this and the next section.

First, we need to set up a suitable classical system $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$. It will be an extension of $\mathsf{E\text{-}PA}^{\omega*}$, which is $\mathsf{E\text{-}HA}^{\omega*}$ with the law of excluded middle added for all formulas. When working with classical systems, we will often take the logical connectives $\neg, \vee, \forall$ as primitive and regard the others as defined. In a similar spirit, the language of $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$ will be that of $\mathsf{E\text{-}PA}^{\omega*}$ extended just with unary predicates st^{σ} for every type $\sigma \in \boldsymbol{T}^*$; the external quantifiers $\forall^{\mathrm{st}}, \exists^{\mathrm{st}}$ are regarded as abbreviations:

$$
\begin{aligned}
\forall^{\mathrm{st}} x\, \Phi(x) &\quad:\equiv\quad \forall x(\, \mathrm{st}(x) \to \Phi(x)\,), \\
\exists^{\mathrm{st}} x\, \Phi(x) &\quad:\equiv\quad \exists x(\, \mathrm{st}(x) \wedge \Phi(x)\,).
\end{aligned}
$$

Definition 6.1 ($\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$)**.** The system $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$ is

$$
\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}} := \mathsf{E\text{-}PA}^{\omega*} + \mathcal{T}^*_{\mathrm{st}} + \mathsf{IA}^{\mathrm{st}}
$$

where

- $\mathcal{T}^*_{\mathrm{st}}$ consists of:

    1. the schema $\mathrm{st}(x) \wedge x = y \to \mathrm{st}(y)$,
    2. a schema providing for each closed term t in $\mathcal{T}^*$ the axiom $\mathrm{st}(t)$,
    3. the schema $\mathrm{st}(f) \wedge \mathrm{st}(x) \to \mathrm{st}(fx)$.

- $\mathsf{IA}^{\mathrm{st}}$ is the external induction axiom:

$$
\mathsf{IA}^{\mathrm{st}} \quad : \quad \big(\Phi(0) \wedge \forall^{\mathrm{st}} n^0(\Phi(n) \to \Phi(n+1))\big) \to \forall^{\mathrm{st}} n^0 \Phi(n).
$$

Again we warn the reader that the induction axiom from $\mathsf{E\text{-}PA}^{\omega*}$

$$
\big(\varphi(0) \wedge \forall n^0(\varphi(n) \to \varphi(n+1))\big) \to \forall n^0 \varphi(n)
$$

is supposed to apply to internal formulas φ only.

As for $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$, we have:

Proposition 6.2. *If a formula Φ is provable in $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$, then its internalization Φ^{int} is provable in $\mathsf{E\text{-}PA}^{\omega*}$. Hence $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$ is a conservative extension of $\mathsf{E\text{-}PA}^{\omega*}$ and $\mathsf{E\text{-}PA}^{\omega}$.*

We will now show how negative translation provides an interpretation of $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$ in $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$. Various negative translations exist, with the one due to Gödel and Gentzen being the most well-known. Here, we work with two variants, the first of which is due to Kuroda [33].

Definition 6.3 (Kuroda's negative translation for $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$)**.** For an arbitrary formula Φ in the language of $\mathsf{E\text{-}PA}^{\omega}_{\mathrm{st}}$, we define its Kuroda negative translation in $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ as

$$
\Phi^{\mathsf{Ku}} \quad:\equiv\quad \neg\neg\Phi_{\mathsf{Ku}},
$$

where Φ_{Ku} is defined inductively on the structure of Φ as follows:

$$
\begin{aligned}
\Phi_{\mathsf{Ku}} &:\equiv \Phi \quad \text{for atomic formulas } \Phi, \\
\big(\neg\Phi\big)_{\mathsf{Ku}} &:\equiv \neg\Phi_{\mathsf{Ku}}, \\
\big(\Phi \vee \Psi\big)_{\mathsf{Ku}} &:\equiv \Phi_{\mathsf{Ku}} \vee \Psi_{\mathsf{Ku}}, \\
\big(\forall x\, \Phi(x)\big)_{\mathsf{Ku}} &:\equiv \forall x\, \neg\neg\Phi_{\mathsf{Ku}}(x).
\end{aligned}
$$

Theorem 6.4. $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}} \vdash \Phi \leftrightarrow \Phi^{\mathsf{Ku}}$ *and if* $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}} + \Delta \vdash \Phi$ *then* $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}} + \Delta^{\mathsf{Ku}} \vdash \Phi^{\mathsf{Ku}}$.

Proof. It is clear that, classically, Φ, Φ_{Ku} and Φ^{Ku} are all equivalent. The second statement is proved by induction on the proof of $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}} + \Delta \vdash \Phi$. For the cases of the axioms and rules of classical logic and $\mathsf{E\text{-}PA}^{\omega*}$, see, for instance, [25, Proposition 10.3]. As the Kuroda negative translation of every instance of $\mathcal{T}^*_{\mathrm{st}}$ or $\mathsf{IA}^{\mathrm{st}}$ is provable in $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{st}}$ using the same instance of $\mathcal{T}^*_{\mathrm{st}}$ or $\mathsf{IA}^{\mathrm{st}}$, the statement is proved. $\qquad\square$

It will turn out to be convenient to introduce a second negative translation, extracted from the work of Krivine by Streicher and Reus (see [32, 47, 46]). This translation will interpret $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$ into $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{nst}}$ (see Remark 5.7).

Definition 6.5 (Krivine's negative translation for $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$)**.** For an arbitrary formula Φ in the language of $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$, we define its Krivine negative translation in $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{nst}}$ as

$$\Phi^{\mathsf{Kr}} :\equiv \neg\Phi_{\mathsf{Kr}},$$

where Φ_{Kr} is defined inductively on the structure of Φ as follows

$$\begin{aligned} \varphi_{\mathsf{Kr}} &:\equiv \neg\varphi \quad \text{for an internal atomic formula } \varphi, \\ \mathrm{st}(x)_{\mathsf{Kr}} &:\equiv \mathrm{nst}(x), \\ \left(\neg\Phi\right)_{\mathsf{Kr}} &:\equiv \neg\Phi_{\mathsf{Kr}}, \\ \left(\Phi \vee \Psi\right)_{\mathsf{Kr}} &:\equiv \Phi_{\mathsf{Kr}} \wedge \Psi_{\mathsf{Kr}}, \\ \left(\forall x\,\Phi(x)\right)_{\mathsf{Kr}} &:\equiv \exists x\,\Phi_{\mathsf{Kr}}(x). \end{aligned}$$

Theorem 6.6. *For every formula Φ in the language of $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$, we have:*

1. $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{nst}} \vdash \Phi^{\mathsf{Kr}} \leftrightarrow \Phi^{\mathsf{Ku}}$.

2. *If $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}} + \Delta \vdash \Phi$, then $\mathsf{E\text{-}HA}^{\omega*}_{\mathrm{nst}} + \Delta^{\mathsf{Kr}} \vdash \Phi^{\mathsf{Kr}}$.*

Proof. Item 1 is easily proved by induction on the structure of Φ. Item 2 follows from item 1 and Theorem 6.4. $\qquad\square$

7   A functional interpretation for $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$

We will now combine negative translation and our functional interpretation D_{st} to obtain a functional interpretation of the classical system $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$.

7.1   The interpretation

Definition 7.1. (S_{st}-interpretation for $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$.) To each formula $\Phi(\underline{a})$ with free variables $\underline{a}$ in the language of $\mathsf{E\text{-}PA}^{\omega*}_{\mathrm{st}}$ we associate its S_{st}-interpretation

$$\Phi^{S_{\mathrm{st}}}(\underline{a}) :\equiv \forall^{\mathrm{st}}\underline{x}\,\exists^{\mathrm{st}}\underline{y}\,\varphi_S(\underline{x}, \underline{y}, \underline{a}),$$

where φ_S is an internal formula. Moreover, $\underline{x}$ and $\underline{y}$ are tuples of variables whose length and types depend only on the logical structure of Φ. The interpretation of the formula is defined inductively on its structure. If

$$\Phi^{S_{\mathrm{st}}}(\underline{a}) :\equiv \forall^{\mathrm{st}}\underline{x}\,\exists^{\mathrm{st}}\underline{y}\,\varphi_S(\underline{x}, \underline{y}, \underline{a}) \quad \text{and} \quad \Psi^{S_{\mathrm{st}}}(\underline{b}) :\equiv \forall^{\mathrm{st}}\underline{u}\,\exists^{\mathrm{st}}\underline{v}\,\psi_S(\underline{u}, \underline{v}, \underline{b}),$$

then

   (i) $\varphi^{S_{\mathrm{st}}} :\equiv \varphi$ for atomic internal $\varphi(\underline{a})$,

   (ii) $\left(\mathrm{st}(z)\right)^{S_{\mathrm{st}}} :\equiv \exists^{\mathrm{st}}x\,(z = x)$,

(iii) $(\neg\Phi)^{S_{\mathrm{st}}} :\equiv \forall^{\mathrm{st}}\underline{Y}\exists^{\mathrm{st}}\underline{x}\,\forall\underline{y}\in\underline{Y}[\underline{x}]\neg\varphi_S(\underline{x},\underline{y},\underline{a})$,

(iv) $(\Phi\vee\Psi)^{S_{\mathrm{st}}} :\equiv \forall^{\mathrm{st}}\underline{x},\underline{u}\exists^{\mathrm{st}}\underline{y},\underline{v}\,\big(\varphi_S(\underline{x},\underline{y},\underline{a})\vee\psi_S(\underline{u},\underline{v},\underline{b})\big)$,

(v) $(\forall z\,\varphi)^{S_{\mathrm{st}}} :\equiv \forall^{\mathrm{st}}\underline{x}\exists^{\mathrm{st}}\underline{y}\forall z\exists\underline{y}'\in\underline{y}\,\varphi_S(\underline{x},\underline{y}',z)$.

Theorem 7.2. (Soundness of the S_{st}-interpretation.) *Let $\Phi(\underline{a})$ be a formula in the language of* $\mathsf{E\text{-}PA}^{\omega *}_{\mathrm{st}}$ *and suppose* $\Phi(\underline{a})^{S_{\mathrm{st}}} \equiv \forall^{\mathrm{st}}\underline{x}\,\exists^{\mathrm{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$. *If* Δ_{int} *is a collection of internal formulas and*

$$\mathsf{E\text{-}PA}^{\omega *}_{\mathrm{st}} + \Delta_{\mathsf{int}} \vdash \Phi(\underline{a}),$$

then one can extract from the formal proof a sequence of closed terms $\underline{t}$ in $\mathcal{T}^*$ such that

$$\mathsf{E\text{-}PA}^{\omega *} + \Delta_{\mathsf{int}} \vdash \ \forall\underline{x}\exists\underline{y}\in\underline{t}(\underline{x})\ \varphi(\underline{x},\underline{y},\underline{a}).$$

Our proof of this theorem relies on the following lemma:

Lemma 7.3. *Let $\Phi(\underline{a})$ be a formula in the language of* $\mathsf{E\text{-}PA}^{\omega *}_{\mathrm{st}}$ *and assume*

$$\begin{aligned}
\Phi^{S_{\mathrm{st}}} &\equiv \forall^{\mathrm{st}}\underline{x}\exists^{\mathrm{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a}) \quad and \\
(\Phi_{\mathsf{Kr}})^{D_{\mathrm{st}}} &\equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\theta(\underline{u},\underline{v},\underline{a}).
\end{aligned}$$

Then the tuples $\underline{x}$ and $\underline{u}$ have the same length and the variables they contain have the same types. The same applies to $\underline{y}$ and $\underline{v}$. In addition, we have

$$\mathsf{E\text{-}PA}^{\omega *} \ \vdash \ \varphi(\underline{x},\underline{y},\underline{a}) \leftrightarrow \neg\theta(\underline{x},\underline{y},\underline{a}).$$

Proof. The proof is by induction on the structure of Φ.

(i) If $\Phi \equiv \psi$, an internal and atomic formula, then $\varphi \equiv \psi$ and $\theta \equiv \neg\psi$, so $\mathsf{E\text{-}PA}^{\omega *} \vdash \varphi \leftrightarrow \neg\theta$.

(ii) If $\Phi \equiv \mathrm{st}(z)$, then $\varphi \equiv y = z$ and $\theta \equiv y \neq z$, so $\mathsf{E\text{-}PA}^{\omega *} \vdash \varphi \leftrightarrow \neg\theta$.

(iii) If $\Phi \equiv \neg\Phi'$ with $(\Phi')^{S_{\mathrm{st}}} \equiv \forall^{\mathrm{st}}\underline{x}\exists^{\mathrm{st}}\underline{y}\,\varphi'(\underline{x},\underline{y},\underline{a})$ and $(\Phi'_{\mathsf{Kr}})^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\theta'(\underline{u},\underline{v},\underline{a})$, then $\varphi \equiv \forall\underline{y}' \in \underline{Y}[\underline{x}]\ \neg\varphi'(\underline{x},\underline{y}')$ and $\theta \equiv \neg\forall\underline{i} \in \underline{Y}[\underline{x}]\ \theta'(\underline{x},\underline{i})$. Since $\mathsf{E\text{-}PA}^{\omega *} \vdash \varphi' \leftrightarrow \neg\theta'$ by induction hypothesis, also $\mathsf{E\text{-}PA}^{\omega *} \vdash \varphi \leftrightarrow \neg\theta$.

(iv) If $\Phi \equiv \Phi_0 \vee \Phi_1$ with

$$\Phi_i^{S_{\mathrm{st}}} \equiv \forall^{\mathrm{st}}\underline{x}\exists^{\mathrm{st}}\underline{y}\,\varphi_i(\underline{x},\underline{y},\underline{a})$$

and

$$((\Phi_i)_{\mathsf{Kr}})^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\theta_i(\underline{u},\underline{v},\underline{a}),$$

then $\varphi \equiv \varphi_0 \vee \varphi_1$ and $\theta \equiv \theta_0 \wedge \theta_1$. Since $\mathsf{E\text{-}PA}^{\omega *} \vdash \varphi_i \leftrightarrow \neg\theta_i$ by induction hypothesis, also $\mathsf{E\text{-}PA}^{\omega *} \vdash \varphi \leftrightarrow \neg\theta$.

(v) If $\Phi \equiv \forall z\,\Phi'$ with

$$(\Phi')^{S_{\mathrm{st}}} \equiv \forall^{\mathrm{st}}\underline{x}\exists^{\mathrm{st}}\underline{y}\,\varphi'(\underline{x},\underline{y},z,\underline{a})$$

and

$$(\Phi'_{\mathsf{Kr}})^{D_{\mathrm{st}}} \equiv \exists^{\mathrm{st}}\underline{u}\forall^{\mathrm{st}}\underline{v}\,\theta'(\underline{u},\underline{v},z,\underline{a}),$$

then $\varphi \equiv \forall z\exists\underline{y}' \in \underline{y}\varphi'(\underline{x},\underline{y}',z,\underline{a})$ and $\theta \equiv \exists z\forall\underline{y}' \in \underline{y}\ \theta'(\underline{x},\underline{y}',z,\underline{a})$. Since $\mathsf{E\text{-}PA}^{\omega *} \vdash \varphi' \leftrightarrow \neg\theta'$ by induction hypothesis, also $\mathsf{E\text{-}PA}^{\omega *} \vdash \varphi \leftrightarrow \neg\theta$.

$\square$

Remark 7.4. This lemma is the reason why we introduced the system $\mathsf{E\text{-}HA}^{\omega *}_{\mathsf{nst}}$ in Remark 5.7: it would fail if we would let the Krivine negative translation land directly in $\mathsf{E\text{-}HA}^{\omega *}_{\mathrm{st}}$ with $\mathrm{st}(z)_{\mathsf{Kr}} = \neg\,\mathrm{st}(z)$. As it is, this lemma yields a quick proof of the soundness of the S_{st}-interpretation.

Proof. (Of the soundness of the S_{st}-interpretation, Theorem 7.2.) Let $\Phi(\underline{a})$ be a formula in the language of $\mathsf{E\text{-}PA}^{\omega*}_{st}$ and let φ and θ be such that

$$
\begin{aligned}
\Phi^{S_{st}} &\equiv \forall^{st}\underline{x}\exists^{st}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a}), \\
(\Phi_{Kr})^{D_{st}} &\equiv \exists^{st}\underline{x}\forall^{st}\underline{y}\,\theta(\underline{x},\underline{y},\underline{a})
\end{aligned}
$$

and $\mathsf{E\text{-}PA}^{\omega*} \vdash \varphi \leftrightarrow \neg\theta$, as in Lemma 7.3.

Now, suppose that Δ_{int} is a set of internal formulas and $\Phi(\underline{a})$ is a formula provable in $\mathsf{E\text{-}PA}^{\omega*}_{st}$ from Δ_{int}. We first apply soundness of the Krivine negative translation (Theorem 6.6) to see that

$$\mathsf{E\text{-}HA}^{\omega*}_{nst} + \Delta^{Kr}_{int} \vdash \Phi^{Kr},$$

where $\Phi^{Kr} \equiv \neg\Phi_{Kr}$. So if $(\Phi_{Kr})^{D_{st}} \equiv \exists^{st}\underline{x}\forall^{st}\underline{y}\,\theta(\underline{x},\underline{y},\underline{a})$, then

$$(\Phi^{Kr})^{D_{st}} \equiv \exists^{st}\underline{Y}\forall^{st}\underline{x}\exists\underline{y}\in\underline{Y}[\underline{x}]\neg\theta(\underline{x},\underline{y},\underline{a}).$$

It follows from the soundness theorem for D_{st} (Theorem 5.5) and Remark 5.7 that there is a sequence of closed terms $\underline{s}$ from $\mathcal{T}^*$ such that

$$\mathsf{E\text{-}HA}^{\omega*} + \Delta^{Kr}_{int} \vdash \forall\underline{x}\exists\underline{y}\in\underline{s}[\underline{x}]\neg\theta(\underline{x},\underline{y},\underline{a}).$$

Since $\mathsf{E\text{-}PA}^{\omega*} \vdash \Delta^{Kr}_{int} \leftrightarrow \Delta_{int}$ and $\mathsf{E\text{-}PA}^{\omega*} \vdash \varphi \leftrightarrow \neg\theta$ we have

$$\mathsf{E\text{-}PA}^{\omega*} + \Delta_{int} \vdash \forall\underline{x}\exists\underline{y}\in\underline{t}(\underline{x})\,\varphi(\underline{x},\underline{y},\underline{a}),$$

with $\underline{t} \equiv \lambda\underline{x}.\underline{s}[\underline{x}]$. $\qquad\square$

7.2 Characteristic principles

The characteristic principles of our functional interpretation for classical arithmetic are idealization I (or, equivalently, R: see Section 4.1) and HAC_{int}

$$\forall^{st}x\exists^{st}y\,\varphi(x,y) \rightarrow \exists^{st}F\forall^{st}x\exists y\in F(x)\,\varphi(x,y),$$

which is the choice scheme HAC restricted to internal formulas. To see this, note first of all that we have:

Proposition 7.5. *For any formula Φ in the language of $\mathsf{E\text{-}PA}^{\omega*}_{st}$ one has:*

$$\mathsf{E\text{-}PA}^{\omega*}_{st} + \mathsf{I} + \mathsf{HAC}_{int} \vdash \Phi \leftrightarrow \Phi^{S_{st}}.$$

Proof. An easy proof by induction on the structure of Φ, using HAC_{int} for the case of negation and I (or rather R) in the case of internal universal quantification. $\qquad\square$

For the purpose of showing that I and HAC_{int} are interpreted, it will be convenient to consider the "hybrid" system $\mathsf{E\text{-}HA}^{\omega*}_{nst} + \mathsf{LEM}_{int}$, where LEM_{int} is the law of excluded middle for internal formulas. For this hybrid system we have the following easy lemma, whose proof we omit:

Lemma 7.6. *We have:*

1. *$\mathsf{E\text{-}HA}^{\omega*}_{nst} + \mathsf{LEM}_{int} \vdash \varphi^{Ku} \leftrightarrow \varphi$, if φ is an internal formula in the the language of $\mathsf{E\text{-}PA}^{\omega*}_{st}$.*

2. *$\mathsf{E\text{-}HA}^{\omega*}_{nst} + \mathsf{LEM}_{int} + \mathsf{I} \vdash \mathsf{I}^{Ku}$.*

3. *$\mathsf{E\text{-}HA}^{\omega*}_{nst} + \mathsf{LEM}_{int} + \mathsf{HAC}_{int} + \mathsf{HGMP}^{st} \vdash \mathsf{HAC}^{Ku}_{int}$.*

This means we can strengthen Theorem 7.2 to:

Theorem 7.7. (Soundness of the S_{st}-interpretation, full version.) *Let $\Phi(\underline{a})$ be a formula in the language of* $\mathsf{E\text{-}PA}^{\omega*}_{\text{st}}$ *and suppose* $\Phi(\underline{a})^{S_{\text{st}}} \equiv \forall^{\text{st}}\underline{x}\,\exists^{\text{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$. *If* Δ_{int} *is a collection of internal formulas and*

$$\mathsf{E\text{-}PA}^{\omega*}_{\text{st}} + \mathsf{I} + \mathsf{HAC}_{\text{int}} + \Delta_{\text{int}} \vdash \Phi(\underline{a}),$$

then one can extract from the formal proof a sequence of closed terms $\underline{t}$ in $\mathcal{T}^*$ such that

$$\mathsf{E\text{-}PA}^{\omega*} + \Delta_{\text{int}} \vdash \ \forall \underline{x}\,\exists \underline{y} \in \underline{t}(\underline{x})\ \varphi(\underline{x},\underline{y},\underline{a}).$$

Proof. The argument is a slight extension of the proof of Theorem 7.2. So, once again, let $\Phi(\underline{a})$ be a formula in the language of $\mathsf{E\text{-}PA}^{\omega*}_{\text{st}}$ and φ and θ be such that

$$
\begin{aligned}
\Phi^{S_{\text{st}}} &\equiv \ \forall^{\text{st}}\underline{x}\,\exists^{\text{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a}), \\
(\Phi_{\mathsf{Kr}})^{D_{\text{st}}} &\equiv \ \exists^{\text{st}}\underline{x}\,\forall^{\text{st}}\underline{y}\,\theta(\underline{x},\underline{y},\underline{a})
\end{aligned}
$$

and $\mathsf{E\text{-}PA}^{\omega*} \vdash \varphi \leftrightarrow \neg\theta$, as in Lemma 7.3.

This time we suppose Δ_{int} is a set of internal formulas and $\Phi(\underline{a})$ is a formula provable in $\mathsf{E\text{-}PA}^{\omega*}_{\text{st}}$ from $\mathsf{I} + \mathsf{HAC}_{\text{int}} + \Delta_{\text{int}}$. We first apply soundness of the Kuroda negative translation (Theorem 6.4), which yields:

$$\mathsf{E\text{-}HA}^{\omega*}_{\text{nst}} + \mathsf{I}^{\mathsf{Ku}} + \mathsf{HAC}^{\mathsf{Ku}}_{\text{int}} + \Delta^{\mathsf{Ku}}_{\text{int}} \vdash \Phi^{\mathsf{Ku}}.$$

Then the previous lemma implies that:

$$\mathsf{E\text{-}HA}^{\omega*}_{\text{nst}} + \mathsf{LEM}_{\text{int}} + \mathsf{I} + \mathsf{HAC}_{\text{int}} + \mathsf{HGMP}^{\text{st}} + \Delta^{\mathsf{Ku}}_{\text{int}} \vdash \Phi^{\mathsf{Ku}}.$$

Note that $\mathsf{E\text{-}HA}^{\omega*}_{\text{nst}} \vdash \Phi^{\mathsf{Ku}} \leftrightarrow \Phi^{\mathsf{Kr}}$, $\Phi^{\mathsf{Kr}} \equiv \neg\Phi_{\mathsf{Kr}}$ and

$$(\Phi^{\mathsf{Kr}})^{D_{\text{st}}} \equiv \exists^{\text{st}}\underline{Y}\,\forall^{\text{st}}\underline{x}\,\exists \underline{y} \in \underline{Y}[\underline{x}]\neg\theta(\underline{x},\underline{y},\underline{a}).$$

Therefore the soundness theorem for D_{st} (Theorem 5.5), in combination with Remark 5.7 and the fact that the axiom scheme $\mathsf{LEM}_{\text{int}}$ is internal, implies that there is a sequence of closed terms $\underline{s}$ from $\mathcal{T}^*$ such that

$$\mathsf{E\text{-}HA}^{\omega*} + \mathsf{LEM} + \Delta^{\mathsf{Ku}}_{\text{int}} \vdash \forall \underline{x}\,\exists \underline{y} \in \underline{s}[\underline{x}]\neg\theta(\underline{x},\underline{y},\underline{a}).$$

Since $\mathsf{E\text{-}PA}^{\omega*} \vdash \mathsf{LEM}$, $\mathsf{E\text{-}PA}^{\omega*} \vdash \Delta^{\mathsf{Ku}}_{\text{int}} \leftrightarrow \Delta_{\text{int}}$ and $\mathsf{E\text{-}PA}^{\omega*} \vdash \varphi \leftrightarrow \neg\theta$, we have

$$\mathsf{E\text{-}PA}^{\omega*} + \Delta_{\text{int}} \vdash \ \forall \underline{x}\,\exists \underline{y} \in \underline{t}(\underline{x})\ \varphi(\underline{x},\underline{y},\underline{a})$$

with $\underline{t} \equiv \lambda\underline{x}.\underline{s}[\underline{x}]$. $\qquad\qquad\square$

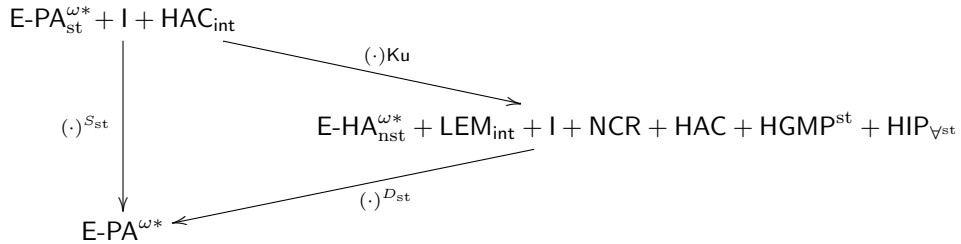The following picture depicts the relation between the various interpretations we have established:

Figure 1: The Shoenfield and negative Dialectica interpretations.

7.3 Conservation results and the transfer principle

Theorem 7.7 immediately gives us the following conservation result:

Corollary 7.8. $\mathsf{E\text{-}PA}^{\omega*}_{\mathsf{st}} + \mathsf{I} + \mathsf{HAC}_{\mathsf{int}}$ *is a conservative extension of* $\mathsf{E\text{-}PA}^{\omega*}$ *and hence of* $\mathsf{E\text{-}PA}^{\omega}$.

We conjecture that this result is not the best possible and that $\mathsf{E\text{-}PA}^{\omega*}_{\mathsf{st}} + \mathsf{I} + \mathsf{HAC}_{\mathsf{int}} + \mathsf{TP}_\forall$ is also conservative over $\mathsf{E\text{-}PA}^{\omega*}$. This would follow from:

Conjecture 7.9. *Let* $\Phi(\underline{a})$ *be a formula in the language of* $\mathsf{E\text{-}PA}^{\omega*}_{\mathsf{st}}$ *and suppose* $\Phi(\underline{a})^{S_{\mathsf{st}}} \equiv \forall^{\mathsf{st}}\underline{x}\,\exists^{\mathsf{st}}\underline{y}\,\varphi(\underline{x},\underline{y},\underline{a})$. *If* Δ_{int} *is a collection of internal formulas and*

$$\mathsf{E\text{-}PA}^{\omega*}_{\mathsf{st}} + \mathsf{I} + \mathsf{HAC}_{\mathsf{int}} + \mathsf{TP}_\forall + \Delta_{\mathsf{int}} \vdash \Phi(\underline{a}),$$

then

$$\mathsf{E\text{-}PA}^{\omega*} + \Delta_{\mathsf{int}} \vdash \forall\underline{x}\exists\underline{y}\varphi(\underline{x},\underline{y},\underline{a}).$$

Unfortunately, one cannot prove this by showing that one can strengthen the hypothesis of Theorem 7.7 from $\mathsf{E\text{-}PA}^{\omega*}_{\mathsf{st}} + \mathsf{I} + \mathsf{HAC}_{\mathsf{int}} + \Delta_{\mathsf{int}} \vdash \Phi(\underline{a})$ to $\mathsf{E\text{-}PA}^{\omega*}_{\mathsf{st}} + \mathsf{I} + \mathsf{HAC}_{\mathsf{int}} + \mathsf{TP}_\forall + \Delta_{\mathsf{int}} \vdash \Phi(\underline{a})$, for this strengthened version fails. To see why, note that the S_{st}-interpretation of $\mathsf{TP}_\forall$

$$\forall^{\mathsf{st}}\underline{t}\,(\,\forall^{\mathsf{st}}x\,\varphi(x,\underline{t}) \to \forall x\,\varphi(x,\underline{t})\,)$$

is provably equivalent to

$$\forall^{\mathsf{st}}\underline{t}\,\exists^{\mathsf{st}}y\,(\,\varphi(y,\underline{t}) \to \forall x\,\varphi(x,\underline{t})\,).$$

Therefore such a strengthened version of Theorem 7.7 would imply that for any formula in the language of $\mathsf{E\text{-}PA}^{\omega*}$ without parameters $\varphi(x)$ there are terms $t_1,\ldots,t_n$ such that

$$\mathsf{E\text{-}PA}^{\omega*} \vdash \bigwedge_i \varphi(t_i) \to \forall x\,\varphi(x).$$

To refute this general statement, it suffices to consider a quantifier-free formula $\varphi(x)$ such that $\forall x\,\varphi(x)$ is true, but not provable in $\mathsf{E\text{-}PA}^{\omega}$ (such as "x is not the Gödel number of a proof in $\mathsf{E\text{-}PA}^{\omega}$ of $\bot$"). This last argument does not refute the conjecture, for the statement

$$\forall\underline{t}\,\exists y\,(\,\varphi(y,\underline{t}) \to \forall x\,\varphi(x,\underline{t})\,)$$

is a tautology and hence provable in $\mathsf{E\text{-}PA}^{\omega*}$.

Still, we expect that adding $\mathsf{TP}_\forall$ to $\mathsf{E\text{-}PA}^{\omega*}_{\mathsf{st}} + \mathsf{I} + \mathsf{HAC}_{\mathsf{int}}$ does not destroy conservativity over $\mathsf{E\text{-}PA}^{\omega*}$, because transfer is part of many similar nonstandard systems that have been shown to be conservative over classical base theories (see [41] and [39]). One natural way to attack this problem would be to try to prove Conjecture 7.9 along the lines of [39]. We plan to take up these issues in future work.

8 Conclusion and plans for future work

We hope this paper lays the groundwork for future uses of functional interpretations to analyse nonstandard arguments and systems. There are many directions, both theoretical and applied, in which one could further develop this research topic. We conclude this paper by mentioning a few possibilities which we would like to take up in future research.

First of all, we would like to see if the interpretations that we have developed in this paper could be used to "unwind" or "proof-mine" nonstandard arguments. Nonstandard arguments have been used in areas where proof-mining techniques have also

been successful, such as metric fixed point theory (for methods of nonstandard analysis applied to metric fixed point theory, see [1, 22]; for application of proof-mining to metric fixed point theory, see [8, 14, 23, 27, 29, 34]) and ergodic theory (for a nonstandard proof of an ergodic theorem, see [20]; for applications of proof-mining to ergodic theory, see [3, 4, 15, 16, 26, 28, 44]), therefore this looks quite promising. For the former type of applications to work in full generality, one would have to extend our functional interpretation to include types for abstract metric spaces, as in [17, 24].

But there are also a number of theoretical questions which still need to be answered. Several have been mentioned already: for example, mapping the precise relationships between the nonstandard principles that we have introduced. Another question was whether $\mathsf{E}\text{-}\mathsf{PA}^{\omega*}_{\mathrm{st}} + \mathsf{I} + \mathsf{HAC}_{\mathrm{int}} + \mathsf{TP}_\forall$ is conservative over $\mathsf{E}\text{-}\mathsf{PA}^{\omega*}$. Another question is whether our methods allow one to prove conservativity results over $\mathsf{WE}\text{-}\mathsf{HA}^\omega$ and $\mathsf{WE}\text{-}\mathsf{PA}^\omega$ as well: this will be important if one wishes to combine the results presented here with the proof-mining techniques from [25].

In addition, we would also like to understand the use of saturation principles in nonstandard arguments. These are of particular interest for two reasons: first, they are used in the construction of Loeb measures, which belong to one of the most successful nonstandard techniques. Secondly, for certain systems it has turned out that extending them with saturation principles has resulted in an increase in proof-theoretic strength (see [18, 21]).

The general saturation principle is

$$\mathsf{SAT}: \quad \forall^{\mathrm{st}} x^\sigma \, \exists y^\tau \, \Phi(x, y) \to \exists f^{\sigma \to \tau} \, \forall^{\mathrm{st}} x^\sigma \, \Phi(x, f(x)).$$

Whether this principle has a D_{st}-interpretation within Gödel's $\mathcal{T}^*$, we do not know; but

$$\mathsf{CSAT}: \quad \forall^{\mathrm{st}} n^0 \, \exists y^\tau \, \Phi(n, y) \to \exists f^{0 \to \tau} \, \forall^{\mathrm{st}} n^0 \, \Phi(n, f(n))$$

has and that seems to be sufficient for the construction of Loeb measures. Interpreting CSAT and SAT in the classical context using the S_{st}-interpretation is probably quite difficult and it is possible that they require some form of bar recursion. We hope to be able to clarify this in future work.

References

[1] A.G. Aksoy and M.A. Khamsi. *Nonstandard methods in fixed point theory*. Universitext. Springer-Verlag, New York, 1990.

[2] J. Avigad. Weak theories of nonstandard arithmetic and analysis. In *Reverse mathematics 2001*, volume 21 of *Lect. Notes Log.*, pages 19–46. Assoc. Symbol. Logic, La Jolla, CA, 2005.

[3] J. Avigad. The metamathematics of ergodic theory. *Ann. Pure Appl. Logic*, 157(2-3):64–76, 2009.

[4] J. Avigad, P. Gerhardy, and H. Towsner. Local stability of ergodic averages. *Trans. Amer. Math. Soc.*, 362(1):261–288, 2010.

[5] J. Avigad and J. Helzner. Transfer principles in nonstandard intuitionistic arithmetic. *Arch. Math. Logic*, 41(6):581–602, 2002.

[6] B. van den Berg. The Herbrand topos. arXiv:1112.3837, 2011.

[7] U. Berger. Uniform Heyting arithmetic. *Ann. Pure Appl. Logic*, 133(1-3):125–148, 2005.

[8] E. M. Briseid. Logical aspects of rates of convergence in metric spaces. *J. Symbolic Logic*, 74(4):1401–1428, 2009.

[9] J. Diller and W. Nahm. Eine Variante zur Dialectica-Interpretation der Heyting-Arithmetik endlicher Typen. *Arch. Math. Logik Grundlagenforsch.*, 16:49–66, 1974.

[10] F. Ferreira. Injecting uniformities into Peano arithmetic. *Ann. Pure Appl. Logic*, 157(2-3):122–129, 2009.

[11] F. Ferreira and A. Nunes. Bounded modified realizability. *J. Symbolic Logic*, 71(1):329–346, 2006.

[12] F. Ferreira and P. Oliva. Bounded functional interpretation. *Ann. Pure Appl. Logic*, 135(1-3):73–112, 2005.

[13] J. Gaspar. Factorization of the Shoenfield-like bounded functional interpretation. *Notre Dame J. Form. Log.*, 50(1):53–60, 2009.

[14] P. Gerhardy. A quantitative version of Kirk's fixed point theorem for asymptotic contractions. *J. Math. Anal. Appl.*, 316(1):339–345, 2006.

[15] P. Gerhardy. Proof mining in topological dynamics. *Notre Dame J. Form. Log.*, 49(4):431–446, 2008.

[16] P. Gerhardy. Proof mining in practice. In *Logic Colloquium 2007*, volume 35 of *Lect. Notes Log.*, pages 82–91. Assoc. Symbol. Logic, La Jolla, CA, 2010.

[17] P. Gerhardy and U. Kohlenbach. General logical metatheorems for functional analysis. *Trans. Amer. Math. Soc.*, 360(5):2615–2660, 2008.

[18] C.W. Henson and H.J. Keisler. On the strength of nonstandard analysis. *J. Symbolic Logic*, 51(2):377–386, 1986.

[19] M.-D. Hernest. Light functional interpretation: an optimization of Gödel's technique towards the extraction of (more) efficient programs from (classical) proofs. In *Computer science logic*, volume 3634 of *Lecture Notes in Comput. Sci.*, pages 477–492. Springer, Berlin, 2005.

[20] T. Kamae. A simple proof of the ergodic theorem using nonstandard analysis. *Israel J. Math.*, 42(4):284–290, 1982.

[21] H.J. Keisler. The strength of nonstandard analysis. In *The strength of nonstandard analysis*, pages 3–26. SpringerWienNewYork, Vienna, 2007.

[22] W. A. Kirk. Fixed points of asymptotic contractions. *J. Math. Anal. Appl.*, 277(2):645–650, 2003.

[23] U. Kohlenbach. Some computational aspects of metric fixed-point theory. *Nonlinear Anal.*, 61(5):823–837, 2005.

[24] U. Kohlenbach. Some logical metatheorems with applications in functional analysis. *Trans. Amer. Math. Soc.*, 357(1):89–128 (electronic), 2005.

[25] U. Kohlenbach. *Applied proof theory: proof interpretations and their use in mathematics.* Springer Monographs in Mathematics. Springer-Verlag, Berlin, 2008.

[26] U. Kohlenbach. On quantitative versions of theorems due to F. E. Browder and R. Wittmann. *Adv. Math.*, 226(3):2764–2795, 2011.

[27] U. Kohlenbach and L. Leuştean. Mann iterates of directionally nonexpansive mappings in hyperbolic spaces. *Abstr. Appl. Anal.*, (8):449–477, 2003.

[28] U. Kohlenbach and L. Leuştean. A quantitative mean ergodic theorem for uniformly convex Banach spaces. *Ergodic Theory Dynam. Systems*, 29(6):1907–1915, 2009.

[29] U. Kohlenbach and L. Leuştean. Asymptotically nonexpansive mappings in uniformly convex hyperbolic spaces. *J. Eur. Math. Soc. (JEMS)*, 12(1):71–92, 2010.

[30] G. Kreisel. Informal rigour and completeness proofs. In I. Lakatos, editor, *Problems in the Philosophy of Mathmematics*, pages 138–186. North-Holland, Amsterdam, 1967.

[31] G. Kreisel. Axiomatizations of nonstandard analysis that are conservative extensions of formal systems for classical standard analysis. In *Applications of Model Theory to Algebra, Analysis, and Probability (Internat. Sympos., Pasadena, Calif., 1967)*, pages 93–106. Holt, Rinehart and Winston, New York, 1969.

[32] J.-L. Krivine. Opérateurs de mise en mémoire et traduction de Gödel. *Arch. Math. Logic*, 30(4):241–267, 1990.

[33] S. Kuroda. Intuitionistische Untersuchungen der formalistischen Logik. *Nagoya Math. J.*, 2:35–47, 1951.

[34] L. Leustean. A quadratic rate of asymptotic regularity for CAT(0)-spaces. *J. Math. Anal. Appl.*, 325(1):386–399, 2007.

[35] V. Lifschitz. Calculable natural numbers. In *Intensional mathematics*, volume 113 of *Stud. Logic Found. Math.*, pages 173–190. North-Holland, Amsterdam, 1985.

[36] I. Moerdijk. A model for intuitionistic non-standard arithmetic. *Ann. Pure Appl. Logic*, 73(1):37–51, 1995.

[37] I. Moerdijk and E. Palmgren. Minimal models of Heyting arithmetic. *J. Symbolic Logic*, 62(4):1448–1460, 1997.

[38] E. Nelson. Internal set theory: a new approach to nonstandard analysis. *Bull. Amer. Math. Soc.*, 83(6):1165–1198, 1977.

[39] E. Nelson. The syntax of nonstandard analysis. *Ann. Pure Appl. Logic*, 38(2):123–134, 1988.

[40] E. Palmgren. Developments in constructive nonstandard analysis. *Bull. Symbolic Logic*, 4(3):233–272, 1998.

[41] E. Palmgren. An effective conservation result for nonstandard arithmetic. *MLQ Math. Log. Q.*, 46(1):17–23, 2000.

[42] A. Robinson. Metamathematical problems. *J. Symbolic Logic*, 38:500–516, 1973.

[43] A. Robinson. *Non-standard analysis*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 1996. Reprint of the second (1974) edition, With a foreword by Wilhelmus A. J. Luxemburg.

[44] P. Safarik. A quantitive nonlinear strong ergodic theorem for Hilbert spaces. Submitted for publication. Available from the author's homepage.

[45] J.R. Shoenfield. *Mathematical logic*. Association for Symbolic Logic, Urbana, IL, 2001. Reprint of the 1973 second printing.

[46] T. Streicher and U. Kohlenbach. Shoenfield is Gödel after Krivine. *Mathematical Logic Quarterly*, 53(2):176–179, 2007.

[47] T. Streicher and B. Reus. Classical logic, continuation semantics and abstract machines. *Journal of Functional Programming*, 8(06):543–572, 1998.

[48] A. S. Troelstra, editor. *Metamathematical investigation of intuitionistic arithmetic and analysis*. Lecture Notes in Mathematics, Vol. 344. Springer-Verlag, Berlin, 1973.

[49] A. S. Troelstra and D. van Dalen. *Constructivism in mathematics. Vol. II*, volume 123 of *Studies in Logic and the Foundations of Mathematics*. North-Holland Publishing Co., Amsterdam, 1988. An introduction.