Carnegie Mellon University
Research Showcase

Department of Philosophy | Dietrich College of Humanities and Social Sciences

3-6-2001

Transfer Principles in Nonstandard Intuitionistic Arithmetic

Jeremy Avigad
Carnegie Mellon University, avigad@cmu.edu

Jeffrey Helzner
Carnegie Mellon University

Follow this and additional works at: http://repository.cmu.edu/philosophy

Part of the Philosophy Commons

Recommended Citation
Avigad, Jeremy and Helzner, Jeffrey, "Transfer Principles in Nonstandard Intuitionistic Arithmetic" (2001). *Department of Philosophy.* Paper 49.
http://repository.cmu.edu/philosophy/49

This Article is brought to you for free and open access by the Dietrich College of Humanities and Social Sciences at Research Showcase. It has been accepted for inclusion in Department of Philosophy by an authorized administrator of Research Showcase. For more information, please contact research-showcase@andrew.cmu.edu.

Transfer principles in nonstandard intuitionistic arithmetic*

Jeremy Avigad and Jeffrey Helzner

March 6, 2001

Abstract

Using a slight generalization, due to Palmgren, of sheaf semantics, we present a term-model construction that assigns a model to any first-order intuitionistic theory. A modification of this construction then assigns a nonstandard model to any theory of arithmetic, enabling us to reproduce conservation results of Moerdijk and Palmgren for nonstandard Heyting arithmetic. Internalizing the construction allows us to strengthen these results with additional transfer rules; we then show that even trivial transfer axioms or minor strengthenings of these rules destroy conservativity over *HA*. The analysis also shows that nonstandard *HA* has neither the disjunction property nor the explicit definability property. Finally, careful attention to the complexity of our definitions allows us to show that a certain weak fragment of intuitionistic nonstandard arithmetic is conservative over primitive recursive arithmetic.

1 Introduction

Classical models and theories of nonstandard arithmetic and analysis have long been studied, but only recently have their intuitionistic analogues come under consideration. In [10], Moerdijk presents a model of nonstandard analysis, and in [11], Moerdijk and Palmgren use a variant of this construction to obtain conservation results for a nonstandard extension of first-order Heyting arithmetic. Some recent developments in intuitionistic nonstandard analysis are surveyed in [13].

When it comes to classical arithmetic, one can conservatively add a schema of transfer axioms, which asserts that any formula in the original language with standard parameters is equivalent to its relativization to the standard numbers. In [11], however, Moerdijk and Palmgren observe that in the intuitionistic setting the full schema needs to be avoided, since it implies the law of the excluded middle for formulae in the original language. Here we sharpen these results, and

*To appear in *The Archive for Mathematical Logic*.

determine exactly what forms of transfer one can add while maintaining conservativity over HA. We also extend the methods to show that a certain weak fragment of intuitionistic nonstandard arithmetic is conservative over primitive recursive arithmetic.

In Section 2, using a slight generalization, due to Palmgren [12], of sheaf semantics, we describe a general term-model construction that assigns to any intuitionistic first-order theory T a model, with the property that the sentences that are true in the model are exactly the ones provable in T. Our construction is a variant of ones that are well-known. From one point of view, it can be seen as an adaptation of the usual term-model constructions for Kripke or Beth semantics; in moving to the more general class of models afforded by Palmgren's semantics, one finds that the noncanonical elements of the construction — enumerating formulae, extending to prime theories, and so on — magically vanish. (See, for example, the construction of a Kripke model in [17], or Friedman's construction of a Beth model, described in [16, Chapter 13].) Alternatively, one can view our construction as a variant of the standard sheaf-theoretic constructions described in [12] and [4], obtained by replacing the provably functional relations of the underlying site by entailments with renamings of variables, and interpreting the sorts of the language by functors which map objects to sets of terms.

In Section 3, a modification of this construction — replacing formulae in the underlying site with sets of formulae (types), and modifying the notion of covering — allows us to assign a nonstandard model to any theory of arithmetic satisfying certain meager requirements. We show that this nonstandard model has most of the properties of the models constructed in [11] and [3], allowing us to duplicate the conservation results in these two papers and state them more generally. Our construction is further designed to use notions of low logical complexity, allowing us to extend the results to a weak fragment of intuitionistic arithmetic in Section 3. Most of the work described here was carried out without knowledge of [3], but we are grateful to Butz for pointing out that by associating types with their deductive closures, one can view the objects of our site as filters in the lattices of formulae with at most k variables (for various k), ordered under provability; this is to be compared with the filters on $\mathbb{N}^k$ in [10], and the provable filter bases of [11].

Towards the end of Section 3, we observe that one can restrict the objects of our site to types that can be represented with finite amounts of information in a fairly simple way. In Section 4, this enables us to internalize the construction, which is to say, express the forcing relation for the model in the language of arithmetic itself. In the case of Heyting arithmetic, HA, we can then augment the results of the previous section with an additional transfer rule: if nonstandard HA proves $\forall^{\mathrm{st}} x\ \varphi$, where φ is a formula in the original language, then HA proves $\forall x\ \varphi$. This immediately extends to formulae of the form $\forall^{\mathrm{st}} x\ \exists^{\mathrm{st}} y\ \varphi$, and we observe that conservation results for nonstandard classical arithmetic yield a transfer rule for negative formulae as well.

In Section 5, we show that these transfer principles are optimal, in the sense that the addition of even trivial transfer axioms, or minor strengthenings of

the transfer rules just described, destroy conservativity over *HA*. The discussion shows that nonstandard Heyting arithmetic does not have the disjunction property, which then implies that it also fails to have the explicit definability property, even if the existential quantifier is relativized to the standard numbers.

We would like to thank Steve Awodey, Carsten Butz, and the anonymous referee for comments and suggestions, and Harvey Friedman for supplying the proof of Lemma 5.5.

2 An algebraic term-model construction

We will take the basic intuitionistic logical symbols to be $\wedge$, $\vee$, $\rightarrow$, $\forall$, $\exists$, $\perp$, and $=$, with $\neg\varphi$ defined to be $\varphi \rightarrow \perp$ and $\top$ defined to be $\perp \rightarrow \perp$. Identify formulae that differ only in the names of their bound variables. When we introduce a formula $\varphi(x)$, with this notation we only mean to distinguish a particular variable x, and we do not exclude the possibility that φ may have other free variables as well; thereafter, $\varphi(t)$ denotes the result of substituting t for x in φ, renaming bound variables if necessary.

Let $\mathcal{C}$ be a category, and let K be a basis for a Grothendieck topology on $\mathcal{C}$ (providing a notion of "covering"). In the terminology of [12], a *presheaf interpretation* of a first-order language L over the site $(\mathcal{C}, K)$ consists of

- an interpretation of the universe of L by a presheaf F on $\mathcal{C}$, that is, a contravariant functor from $\mathcal{C}$ to the category of sets;
- an interpretation of each k-ary function symbol f of L by a natural transformation from F^k to F; and
- an interpretation of each l-ary relation symbol R of L (including equality) by a function $\hat{R}$ which assigns to each element c of the category an l-ary relation on $F(c)$, satisfying the following:
  1. Monotonicity: Given any object c in $\mathcal{C}$, and any sequence of elements $x_1, \ldots, x_l$ in $F(c)$, if $\hat{R}(c)(x_1, \ldots, x_l)$ and f is an arrow from d to c, then $\hat{R}(d)(F(f)(x_1), \ldots, F(f)(x_l))$.
  2. Covering (local character): Given any c in $\mathcal{C}$ and sequence of elements $x_1, \ldots, x_l$ in $F(c)$, if a set of arrows $\{c_i \xrightarrow{f_i} c\}$ covers c and $\hat{R}(c_i)(F(f_i)(x_1), \ldots, F(f_i)(x_l))$ holds for each i, then $\hat{R}(c)(x_1, \ldots, x_l)$.

Taken together the last two clauses assert that $\hat{R}$ is a closed subpresheaf of F. The more standard notion of a *sheaf interpretation* arises when F is a sheaf and the equality relation is interpreted as equality at each element of the site; in that case, the covering clause above simply says that $\hat{R}$ is a subsheaf of F. We will use the more general notion, and refer to such interpretations more simply as "models."

In the more general case where L is a many-sorted language, each sort S of L is interpreted by a presheaf F_S, and the clauses for the interpretations of

the function and relation symbols need to be modified accordingly. To keep the exposition simple, we will focus on languages with a single sort, but the extension to the more general case is immediate.

Just as with standard sheaf semantics, given a model of the kind described above, a notion of truth at each element of the underlying category is determined by an inductively presented forcing relation. In this section we will assume some familiarity with sheaf semantics and this forcing relation (see [8] or [12]), but readers without this background will be able to extract most of the relevant information from the the particular forcing relation described below. On the other hand, readers familiar with the usual syntactic construction of a sheaf model (see, e.g., [4, 8, 12]) may wish to skim to the beginning of Section 3. The key difference between the standard construction and ours is that instead of using arbitrary definable morphisms in the underlying site, we allow only entailments, up to renaming of variables; verifying that this works is otherwise routine.

Let T be an intuitionistic theory in a language L. Our goal in this section is simply to construct a model of T. We begin by defining the underlying category, $\mathcal{C}$.

If t is any term of L, a *renaming* of the variables of t is an injective map from the set of these variables to the set of all variables. If σ is such a renaming, t^σ denotes the result of replacing each variable x in t with $\sigma(x)$. Similarly, if σ is a renaming of the free variables of φ, then φ^σ denotes the result of replacing each free variable x in φ by $\sigma(x)$, changing the names of the bound variables, if necessary, to prevent collisions. If φ and ψ are formulae in L, we will write $\varphi \vdash_T \psi$ if $T \cup \{\varphi\}$ proves ψ in first-order intuitionistic logic with equality.

Let the objects of $\mathcal{C}$ be formulae in L, and let the arrows between objects φ and ψ be given by renamings σ such that

- σ maps each free variable of ψ to a free variable of φ, and
- $\varphi \vdash_T \psi^\sigma$.

The use of renamings has the result of abstracting away a particular choice of variables: think of the objects of $\mathcal{C}$ as formulae with "places," with σ specifying which places of ψ correspond to which places of φ.

Lemma 2.1 *Suppose $\varphi \xrightarrow{\sigma} \psi$ is an arrow. Let i denote the identity map on the free variables of ψ. Then there is an isomorphism $\varphi' \xrightarrow{\tau} \varphi$ such that $\sigma\tau = i$.*

Proof. Let τ be a renaming of the variables of φ which takes x to $\sigma^{-1}(x)$ if x is in the range of σ, and an arbitrary "fresh" variable otherwise. $\square$

When we indicate an arrow $\varphi \to \psi$ without specifying the renaming, we mean, implicitly, the identity renaming. By Lemma 2.1, if $\varphi \xrightarrow{\sigma} \psi$ is any arrow, there is a φ' isomorphic to φ such that $\varphi' \to \psi$ is an arrow. To specify the pullbacks in this category, then, it is enough to consider the following special case.

Lemma 2.2 *Let $\varphi \to \psi$ and $\theta \to \psi$ be arrows. Let ρ be a renaming of the free variables of φ that fixes the variables that are free in ψ and maps the other free variables of φ to variables that are not free in θ. Then the pullback of $\varphi \to \psi$ and $\theta \to \psi$ is $\varphi^\rho \wedge \theta$, with arrows $\varphi^\rho \wedge \theta \xrightarrow{\rho} \varphi$ and $\varphi^\rho \wedge \theta \to \theta$.*

For example, if $\varphi(x, y)$, $\theta(x, y)$, and $\psi(y)$ have the free variables shown, the pullback of $\varphi(x, y) \to \psi(y)$ and $\theta(x, y) \to \psi(y)$ is $\varphi(x', y) \wedge \theta(x, y)$, where the variable x of $\varphi(x, y)$ has been renamed to x'. More generally, to find the pullback of arrows $\varphi \xrightarrow{\sigma} \psi$ and $\theta \xrightarrow{\tau} \psi$, use Lemma 2.1 to find $\varphi' \simeq \varphi$ and $\theta' \simeq \theta$ such that $\varphi' \to \psi$ and $\theta' \to \psi$ are arrows, and then apply Lemma 2.2. The proof of Lemma 2.2 is left to the reader.

The terminal object of $\mathcal{C}$ is given by the sentence $\top$.

We will say that a set of arrows $\{\varphi_0 \to \psi, \ldots, \varphi_k \to \psi\}$ *covers* ψ if

$$\psi \vdash_T \exists \vec{u} \bigvee_{i=0}^{k} \varphi_i,$$

where the sequence $\vec{u}$ consists of those variables that are free in one of the φ_i, but not ψ. More generally, a set of arrows $\{\varphi_0 \xrightarrow{\sigma_0} \psi, \ldots, \varphi_k \xrightarrow{\sigma_k} \psi\}$ covers ψ if the corresponding set of arrows $\{\varphi'_0 \to \psi, \ldots, \varphi'_k \to \psi\}$ guaranteed by Lemma 2.1 covers ψ. The following inductive characterization of "S covers ψ" makes it clear that this relation forms the basis for a Grothendieck topology.

Lemma 2.3 *The relation "S covers ψ" is the smallest relation (that is, the intersection of all relations between finite sets of arrows in $\mathcal{C}$ and elements of $\mathcal{C}$) satisfying the following:*

1. *If $\psi \vdash_T \varphi \vee \theta$, then $\{\psi \wedge \varphi \to \psi, \psi \wedge \theta \to \psi\}$ covers ψ.*
2. *If $\psi \vdash_T \exists x\ \varphi(x)$, and y is not a free variable of ψ or $\exists x\ \varphi(x)$, then $\{\psi \wedge \varphi(y) \to \psi\}$ covers ψ.*
3. *If $\varphi \xrightarrow{\sigma} \psi$ is an isomorphism, then $\{\varphi \xrightarrow{\sigma} \psi\}$ covers ψ.*
4. *If $\{\ldots \varphi_i \xrightarrow{\sigma_i} \psi \ldots\}$ covers ψ and $\theta \xrightarrow{\tau} \psi$ is any arrow, then the set of pullbacks $\{\ldots \varphi_i \times_\psi \theta \to \theta \ldots\}$ covers θ.*
5. *If $\{\ldots \varphi_i \xrightarrow{\sigma_i} \psi \ldots\}$ covers ψ, and for each i the set $\{\theta_{i1} \xrightarrow{\tau_{i1}} \varphi_i, \ldots, \theta_{ij_i} \xrightarrow{\tau_{ij_i}} \varphi_i\}$ covers φ_i, then the set $\{\ldots \theta_{ik} \xrightarrow{\sigma_i \tau_{ik}} \psi \ldots\}$ covers ψ.*

Alternatively, the relation can be characterized as the smallest basis for a Grothendieck topology satisfying 1 and 2.

Proof. The second statement follows from the first, since a basis for a Grothendieck topology is just a relation satisfying 3, 4, and 5.

It is not difficult to show that the covering relation satisfies clauses 1–5. Conversely, let R be any relation satisfying clauses 1–5, and suppose $\varphi_1 \to$

$\psi, \ldots, \varphi_k \to \psi$ are arrows such that $\psi \vdash_T \exists \vec{u} \bigvee_{i=0}^{k} \varphi_i$, with $\vec{u}$ as above. Applying clauses 1 and 2 repeatedly shows that the set $\{\varphi_1 \to \psi, \ldots, \varphi_k \to \psi\}$ bears the relation R to ψ. $\square$

To interpret the universe of L, define the functor $F : \mathcal{C}^{\mathrm{op}} \to \mathrm{Sets}$ as follows. If φ is an object of $\mathcal{C}$, then $F(\varphi)$ is the set of terms of L with free variables among those of φ. If $\varphi \xrightarrow{\sigma} \psi$ is an arrow, then $F(\sigma)$ is the function which maps a term $t \in F(\psi)$ to $t^\sigma \in F(\varphi)$.

Lemma 2.4 *Suppose $\{\varphi_0 \to \psi, \ldots, \varphi_k \to \psi\}$ covers ψ, and let η be any formula whose free variables are among those of ψ. Then $\psi \vdash_T \eta$ if and only if for each i, $\varphi_i \vdash_T \eta$.*

Proof. The forward direction is immediate, since for each i, $\varphi_i \to \psi$ is an arrow. For the converse direction, use induction on the characterization in Lemma 2.3. For example, consider the base case given by clause 1. Suppose $\psi \vdash_T \varphi \vee \theta$, $\psi \wedge \varphi \vdash_T \eta$, and $\psi \wedge \theta \vdash_T \eta$. Using intuitionistic logic, we have $\psi \wedge (\varphi \vee \theta) \vdash_T \eta$, and hence $\psi \vdash_T \eta$. $\square$

Now define the model $\mathcal{M}(T)$ by interpreting the language L as follows.

- If f is a k-ary function symbol of L, then $f^{\mathcal{M}(T)}$ is the natural transformation from F^k to F which, at an object φ, takes the sequence $t_1, \ldots, t_k$ in $F(\varphi)^k$ to $f(t_1, \ldots, t_k) \in F(\varphi)$.
- If R is a k-ary relation symbol of L or the (binary) equality symbol, then $R^{\mathcal{M}(T)}$ is defined by

$$R^{\mathcal{M}(T)}(\varphi)(t_1, \ldots, t_k) \text{ if and only if } \varphi \vdash_T R(t_1, \ldots, t_k).$$

In the more general case where L is a many-sorted language, interpret each sort of S of L with a presheaf F_S that maps each object φ to the set of terms of L of that sort, with free variables among those of φ; and then generalize the two clauses above in the natural way.

The following lemma asserts that the interpretation of the relation symbols satisfies the requisite conditions.

Lemma 2.5 *The following hold:*

1. *Stability under renaming: if $\varphi \xrightarrow{\sigma} \varphi'$ is an isomorphism, then we have $R^{\mathcal{M}(T)}(\varphi)(t_1, \ldots, t_k)$ if and only if $R^{\mathcal{M}(T)}(\varphi')(t_1^\sigma, \ldots, t_k^\sigma)$*
2. *Monotonicity: if $\varphi \to \psi$ is an arrow of $\mathcal{C}$ and $R^{\mathcal{M}(T)}(\psi)(t_1, \ldots, t_k)$ then $R^{\mathcal{M}(T)}(\varphi)(t_1, \ldots, t_k)$*
3. *Covering property (local character): if $\{\varphi_0 \to \psi, \ldots, \varphi_k \to \psi\}$ is a cover and for each i, $R^{\mathcal{M}(T)}(\varphi_i)(t_1, \ldots, t_k)$, then $R^{\mathcal{M}(T)}(\psi)(t_1, \ldots, t_k)$.*

Proof. The first two are immediate. For the third, use Lemma 2.4. $\square$

Now the forcing clauses presented in [8, Section VI.7] or [12, Section 4] inductively assign to each formula $\psi(x_1, \ldots, x_k)$ of L with the free variables shown (and listed in some preassigned ordering of the variables) a relation $\varphi \Vdash \psi[t_1, \ldots, t_k]$ between objects φ of $\mathcal{C}$ and sequences $t_1, \ldots, t_k$ of elements of $F(\varphi)$. In the current setting, the clauses amount to the ones specified below. When we write $\varphi \Vdash \psi[\vec{t}]$ on the left side of an equivalence, we are assuming that the free variables of ψ are given by a sequence $x_1, \ldots, x_k$, and $\vec{t}$ denotes a sequence of terms $t_1, \ldots, t_k$ of the same length, with the free variables of each t_i among those of φ. In clause 1, $s[\vec{t}/\vec{x}]$ denotes the result of substituting the terms t_i for the variables x_i simultaneously. In clauses 3, 4, and 5, which treat the binary connectives, it is possible that only a subset of the variables $x_1, \ldots, x_k$ occur in θ or η, in which case $\vec{t}$ should really be replaced by the subsequence of those terms corresponding to those variables.

1. $\varphi \Vdash R(s_1, \ldots, s_l)[\vec{t}]$ if and only if $\varphi \vdash_T R(s_1[\vec{t}/\vec{x}], \ldots, s_l[\vec{t}/\vec{x}])$, where R is a relation symbol of L or the equality symbol

2. $\varphi \Vdash \bot$ if and only if $\varphi \vdash_T \bot$

3. $\varphi \Vdash (\theta \wedge \eta)[\vec{t}]$ if and only if $\varphi \Vdash \theta[\vec{t}]$ and $\varphi \Vdash \eta[\vec{t}]$

4. $\varphi \Vdash (\theta \vee \eta)[\vec{t}]$ if and only if there is a cover $\{\nu_1 \to \varphi, \ldots, \nu_l \to \varphi\}$ such that for each i either $\nu_i \Vdash \theta[\vec{t}]$ or $\nu_i \Vdash \eta[\vec{t}]$

5. $\varphi \Vdash (\theta \to \eta)[\vec{t}]$ if and only if for every arrow $\nu \to \varphi$, if $\nu \Vdash \theta[\vec{t}]$, then $\nu \Vdash \eta[\vec{t}]$.

6. $\varphi \Vdash (\forall x\ \theta(x))[\vec{t}]$ if and only if for every arrow $\nu \to \varphi$ and every $u \in F(\nu)$, $\nu \Vdash \theta(z)[\vec{t}, u]$, where z is the first variable after x_k in the preassigned ordering of the variables.

7. $\varphi \Vdash (\exists x\ \theta(x))[\vec{t}]$ if and only if there are a cover $\{\nu_1 \to \varphi, \ldots, \nu_l \to \varphi\}$ and a sequence of elements $u_1 \in F(\nu_1), \ldots, u_l \in F(\nu_l)$, such that for each i, $\nu_i \Vdash \theta(z)[\vec{t}, u_i]$, where z is the first variable after x_k in the preassigned ordering of the variables.

Lemma 2.5 then extends to arbitrary formulae $\psi(x_1, \ldots, x_k)$ in L, if we replace $R^{\mathcal{M}(T)}(\varphi)(t_1, \ldots, t_k)$ everywhere in the statement of the lemma by $\varphi \Vdash \psi[t_1, \ldots, t_k]$. Note that it is clause 1 of Lemma 2.5 that allows us to restrict our attention to identity arrows in the clauses above.

From now on we will use $\Vdash_{\mathcal{M}(T)}$ instead of $\Vdash$ to refer to the forcing relation for this particular model. If ψ is a sentence, we will say $\Vdash_{\mathcal{M}(T)} \psi$, or "$\psi$ is true in $\mathcal{M}(T)$," if $\top \Vdash_{\mathcal{M}(T)} \psi$. Then, as is the case for any presheaf interpretation, all the axioms of intuitionistic first-order logic (without equality) are true, and truth is maintained under the rules of inference. The following theorem shows that in this particular model, the axioms of equality as well as the sentences of T are also true.

Theorem 2.6 *Let $\psi(x_1, \ldots, x_k)$ be any formula of L with the free variables shown. If φ is any formula and $t_1, \ldots, t_k$ are terms whose free variables are among those of φ, then $\varphi \Vdash_{\mathcal{M}(T)} \psi[t_1, \ldots, t_k]$ if and only if $\varphi \vdash_T \psi(t_1, \ldots, t_k)$.*

Proof. By induction on ψ. The case where ψ is atomic is immediate from the definition of the forcing relation. By way of illustration, we will consider the cases where the outermost connective of ψ is either $\exists$ or $\rightarrow$. Out of laziness we will leave out the terms $t_1, \ldots, t_k$, but the diligent reader should feel free to fill them in.

$\exists$ case: Suppose $\varphi \Vdash_{\mathcal{M}(T)} \exists x\ \theta(x)$. Then there are a cover $\{\ldots \psi_i \rightarrow \varphi \ldots\}$ and terms u_i such that for each i, $\psi_i \Vdash_{\mathcal{M}(T)} \theta(z)[u_i]$, where z is the first term after the x_i's in the ordering on the variables. By the inductive hypothesis, for each i we have $\psi_i \vdash_T \theta(u_i)$, and hence $\psi_i \vdash_T \exists x\ \theta(x)$. By Lemma 2.4, $\varphi \vdash_T \exists x\ \theta(x)$.

Conversely, suppose $\varphi \vdash_T \exists x\ \theta(x)$. Then $\{\varphi \wedge \theta(y) \rightarrow \varphi\}$ covers φ, where y is not free in φ or $\exists x\ \theta(x)$. Since $\varphi \wedge \theta(y) \vdash_T \theta(y)$, by the inductive hypothesis we have $\varphi \wedge \theta(y) \Vdash_{\mathcal{M}(T)} \theta(z)[y]$. By the forcing clause for $\exists$, we have $\varphi \Vdash \exists x\ \theta(x)$.

$\rightarrow$ case: Suppose $\varphi \Vdash_{\mathcal{M}(T)} \theta \rightarrow \eta$. Since $\varphi \wedge \theta \vdash_T \theta$, by the inductive hypothesis we have $\varphi \wedge \theta \Vdash_{\mathcal{M}(T)} \theta$. By the forcing clause for implication, we then have $\varphi \wedge \theta \Vdash_{\mathcal{M}(T)} \eta$. By the inductive hypothesis again, we have $\varphi \wedge \theta \vdash_T \eta$, and hence $\varphi \vdash_T \theta \rightarrow \eta$.

Conversely, suppose $\varphi \vdash_T \theta \rightarrow \eta$. Suppose $\nu \rightarrow \varphi$ is any arrow such that $\nu \Vdash_{\mathcal{M}(T)} \theta$. By the inductive hypothesis, $\nu \vdash_T \theta$, and so $\nu \vdash_T \eta$. By the inductive hypothesis again, $\nu \Vdash_{\mathcal{M}(T)} \eta$. Hence $\varphi \Vdash_{\mathcal{M}(T)} \theta \rightarrow \eta$. $\square$

Corollary 2.7 *For any sentence ψ of L, we have $\Vdash_{\mathcal{M}(T)} \psi$ if and only if $\vdash_T \psi$.*

Notes. 1. With respect to conventional proof systems for intuitionistic logic, our semantics is in general sound only if one assumes that the presheaves interpreting the sorts are inhabited (see [12]). But Theorem 2.6 and Corollary 2.7 still hold even if the language L does not have any constant symbols, even though, for example, $F(\top)$ is the empty set. In that case, if one wants an inhabited presheaf, one can simply add a constant to the language before carrying out the construction.

2. Our interpretation of the universe is not, in general, a sheaf. For example, $F(\bot)$ is empty, whereas it is the one-point set in any sheaf. For another example, if L has a unary relation symbol R and no constant symbols, then $\{\exists x\ R(x) \wedge R(y) \rightarrow \exists x\ R(x)\}$ is a covering, y is a term in $F(\exists x\ R(x) \wedge R(y))$, but $F(\exists x\ R(x))$ is empty.

3. If for each formula φ we define the equivalence relation $t \equiv_\varphi t'$ to be $\varphi \vdash_T t = t'$, and then modify the definition of $\mathcal{M}(T)$ so that the elements of $F(\varphi)$ are equivalence classes $[t]$ of terms with free variables among those of φ, we can then interpret the equality symbol at each formula φ by equality in $F(\varphi)$.

4. Our category is not equivalent to the one used in the usual syntactic construction ([4, 8, 12]), but the corresponding sites are equivalent. To each

definable morphism $\sigma(\vec{x}, \vec{y})$ from $\varphi(\vec{x})$ to $\psi(\vec{y})$ in the syntactic category, associate the pair of arrows $\sigma(\vec{x}, \vec{y}) \to \varphi(\vec{x})$ and $\sigma(\vec{x}, \vec{y}) \to \psi(\vec{y})$ in ours. The first arrow is an isomorphism in the category of sheaves over our site, and the Grothendieck comparison lemma [8] does the rest.

We have shown

Corollary 2.8 *Palmgren's generalization of sheaf semantics is complete for first-order logic.*

We would again like to emphasize that this result is well known, following from the completeness of Kripke semantics, Beth semantics, or sheaf semantics. For more on Kripke semantics or Beth semantics, see [16], and for more on the relationship between first-order logic and sheaves, see, for example, [6] or [9].

3 Nonstandard models of arithmetic

Let us take the language of arithmetic to include symbols 0, 1, $+$, $\times$, and $<$, and possibly symbols for other primitive recursive functions and relations as well. If n is a natural number, we will use $\bar{n}$ to denote the corresponding numeral. In this section we will be concerned with theories T in a language L extending the language of arithmetic, satisfying the following:

- T evaluates the primitive recursive functions correctly; in other words, if f is one of the primitive recursive function symbols, T proves $f(\bar{n}_1, \ldots, \bar{n}_k) = \bar{m}$ if and only if $f(n_1, \ldots, n_k) = m$ is true in the standard model.
- T proves that $<$ is a decidable linear ordering; that is, T proves that $<$ is a linear ordering, satisfying $\forall x, y\ (x < y \vee \neg(x < y))$.
- For each natural number n, T proves $x \leq \bar{n} \leftrightarrow x = \bar{0} \vee \ldots \vee x = \bar{n}$.

Our goal is to modify the construction of the previous section to obtain a *nonstandard* model of T. Our construction has many of the properties enjoyed by the one used in [11], but Lemma 4.6 below exploits the fact that the notions involved are of lower logical complexity.

Let L^{st} denote the language L together with a new unary relation symbol St, intended to denote the "standard" numbers. Let (NS) denote the set of axioms consisting of the universal closures of the following:

1. $\mathrm{St}(x_1) \wedge \ldots \wedge \mathrm{St}(x_n) \to \mathrm{St}(f(x_1, \ldots, x_n))$ for any primitive recursive function symbol of L
2. $\neg\neg\mathrm{St}(x) \to \mathrm{St}(x)$
3. *External induction:* for each formula $\varphi(x)$ of L^{st},

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

4. *Overspill:* for each formula $\varphi(x)$ of L,

$$\forall^{\mathrm{st}} x\ \varphi(x) \to \exists x\ (\neg \mathrm{St}(x) \wedge \varphi(x)).$$

5. *Underspill:* for each formula $\varphi(x)$ of L,

$$\forall x\ (\neg \mathrm{St}(x) \to \varphi(x)) \to \exists^{\mathrm{st}} x\ \varphi(x).$$

The model of T that we construct will satisfy these axioms, as well as the nonclassical axiom (NC):

$$\neg \forall x\ (\mathrm{St}(x) \vee \neg \mathrm{St}(x)).$$

Applying overspill to the formula $\top$ implies that there is a nonstandard number, that is, $\exists x\ \neg \mathrm{St}(x)$. Using external induction we can show that the standard numbers form an initial segment, $\forall^{\mathrm{st}} x\ \forall y < x\ \mathrm{St}(y)$. The theory HAI of [11] consists of Heyting arithmetic, HA, together with (NS), without the underspill axiom; but it is noted implicitly in [13] and explicitly in [3] that underspill holds in their model as well. Below, we will use HAI to denote $HA + (NS)$.

To construct a model of T, we again begin by describing the underlying category. If Γ is a set of formulae in L, by the "free variables of Γ" we mean the set of variables free in some formula in Γ. Such a set Γ is said to be an *n-type* if it has n free variables. We will say that Γ is a *type* if it is an n-type, for some n. (Here "type" is used in the model-theoretic sense, rather than in the type-theoretic sense.) If Γ and Δ are types, we will use $\Gamma \vdash_T \Delta$ to indicate that $T \cup \Gamma$ proves each formula in Δ intuitionistically. If σ is a renaming of the free variables of Γ, Γ^σ is defined to be $\{\varphi^\sigma \mid \varphi \in \Gamma\}$.

Given T, define the category $\mathcal{D}$ to consist of types in the language of T, with an arrow $\Gamma \xrightarrow{\sigma} \Delta$ if

- σ maps each free variable of Δ to a free variable of Γ, and
- $\Gamma \vdash_T \Delta^\sigma$.

$\mathcal{D}$ is not very different from the category $\mathcal{C}$ of Section 2: for example, the corresponding version of Lemma 2.1 still holds, and pullbacks are constructed in the analogous way. The terminal object is the empty type, $\emptyset$.

Define the relation "S covers Γ," where S is a finite set of arrows to Γ, to be the smallest basis for a Grothendieck topology satisfying the following two clauses:

1. If $\Gamma \vdash \varphi \vee \theta$, then $\{\Gamma \cup \{\varphi\} \to \Gamma, \Gamma \cup \{\theta\} \to \Gamma\}$ covers Γ.
2. If $\Gamma \vdash \exists x\ \varphi(x)$, and y is not free in Γ or $\exists x\ \varphi(x)$, then the singleton

$$\{\Gamma \cup \{\varphi(y)\} \cup \{y \geq \bar{n} \mid \Gamma \vdash_T \varphi(\bar{n})\} \to \Gamma\}$$

covers Γ.

Alternatively, the relation "S covers Γ" can be characterized as the smallest relation satisfying these two clauses as well as the analogues of clauses 3–5 in Lemma 2.3. It will become clear below that the second clause is designed to ensure that overspill holds in the model we are building.[1]

In analogy to Lemma 2.4, we have:

Lemma 3.1 *Suppose $\{\Gamma_0 \to \Delta, \ldots, \Gamma_k \to \Delta\}$ covers Δ, and let η be any formula whose free variables are among those of Δ. Then $\Delta \vdash_T \eta$ if and only if for each i, $\Gamma_i \vdash_T \eta$.*

Proof. The only novelty here is that the clause for the existential quantifier in Lemma 2.3 has been replaced by clause 2 above. But suppose $\Gamma \vdash_T \exists x\ \varphi(x)$ and $\Gamma \cup \{\varphi(y)\} \cup \{y \geq \bar{n} \mid \Gamma \vdash_T \varphi(\bar{n})\} \vdash_T \eta$, where y is not free in Γ or $\exists x\ \varphi(x)$. Then either $\Gamma \cup \{\varphi(y)\} \vdash_T \eta$, or $\Gamma \cup \{\varphi(y), y \geq \bar{n}\} \vdash_T \eta$ for some n such that $\Gamma \vdash_T \varphi(\bar{n})$. In the first case, $\Gamma \cup \{\exists x\ \varphi(x)\} \vdash_T \eta$ by the laws of intuitionistic logic, and hence $\Gamma \vdash_T \eta$. In the second case, substituting $\bar{n}$ for y yields $\Gamma \cup \{\varphi(\bar{n}), \bar{n} \geq \bar{n}\} \vdash_T \eta$, and hence $\Gamma \vdash_T \eta$. $\square$

Define the presheaf G in analogy to the presheaf F of Section 2; that is, let G map each type Γ to the set of terms whose free variables are among those in Γ. Then interpret the function symbols and relation symbols of L as before, and, temporarily, call the resulting model $\mathcal{N}(T)$. (We will extend the interpretation to L^{st} in just a moment.) Clauses describing the forcing relation in $\mathcal{N}(T)$ are obtained from the those of $\mathcal{M}(T)$ by replacing elements φ of the category $\mathcal{C}$ by elements Γ of category $\mathcal{D}$. The relevant lemmata then carry over, mutatis mutandis:

Lemma 3.2 *The stability, monotonicity, and covering properties of Lemma 2.5 hold for $\Vdash_{\mathcal{N}(T)}$ as well.*

Lemma 3.3 *Let $\psi(x_1, \ldots, x_k)$ be any formula of L with the free variables shown. If Γ is any type and $t_1, \ldots, t_k$ are terms whose free variables are among those of Γ, then $\Gamma \Vdash_{\mathcal{N}(T)} \psi[t_1, \ldots, t_k]$ if and only if $\Gamma \vdash_T \psi(t_1, \ldots, t_k)$.*

Lemma 3.3 guarantees that $\mathcal{N}(T)$, like $\mathcal{M}(T)$, is a minimal model of T, in the sense that the sentences true in the model are exactly the ones provable in T. Now extend $\mathcal{N}(T)$ to L^{st} by declaring that at each type Γ, $\mathrm{St}^{\mathcal{N}(T)}(\Gamma)(t)$ if and only if the free variables of t are among those of Γ and there is an n such that $\Gamma \vdash_T t \leq \bar{n}$. In other words, to interpret the larger language, we extend the forcing relation for $\mathcal{N}(T)$ with the following clause:

$$\Gamma \Vdash_{\mathcal{N}(T)} \mathrm{St}(s)[\vec{t}] \text{ if and only if for some } n,\ \Gamma \vdash_T s[\vec{t}/\vec{x}] \leq \bar{n}.$$

[1] Were we not concerned with formalizing the argument later on, we could more generally say that Γ' covers Γ if for every finite subset $S \subseteq \Gamma'$, $\Gamma \vdash_T \exists \vec{y} \bigwedge S$, where $\vec{y}$ are the free variables of Γ' that are not free in Γ. This would result in a model with corresponding "saturation" properties.

It is not difficult to verify that this interpretation of St also satisfies the stability, monotonicity, and covering properties of Lemma 2.5.

We will now show that the model $\mathcal{N}(T)$ shares most of the properties of the nonstandard models constructed in [11] and [3]. Take $\forall^{\mathrm{st}} x\ \varphi$ and $\exists^{\mathrm{st}} x\ \varphi$ to abbreviate $\forall x\ (\mathrm{St}(x) \to \varphi)$ and $\exists x\ (\mathrm{St}(x) \wedge \varphi)$ respectively.

The following lemma guarantees that the axioms of equality still hold in the expanded version of $\mathcal{N}(T)$.

Lemma 3.4 *We have* $\Vdash_{\mathcal{N}(T)} \forall x, y\ (\mathrm{St}(x) \wedge x = y \to \mathrm{St}(y))$.

Proof. Left to the reader. □

Lemma 3.5 *Suppose* Γ *is a type,* $\varphi(x_1, \ldots, x_k, y)$ *is any formula of* L^{st} *with the free variables shown, and* $t_1, \ldots, t_k$ *are elements of* $G(\Gamma)$*. Then* $\Gamma \Vdash_{\mathcal{N}(T)} \forall^{\mathrm{st}} y\ \varphi(\vec{x}, y)[\vec{t}]$ *if and only if for every natural number* n*,* $\Gamma \Vdash_{\mathcal{N}(T)} \varphi(\vec{x}, y)[\vec{t}, \bar{n}]$*.*

Proof. For convenience we will leave out the variables $\vec{x}$ and the terms $\vec{t}$. The forward direction is immediate, since for every natural number n we have $\Vdash_{\mathcal{N}(T)} \mathrm{St}(y)[\bar{n}]$.

Conversely, suppose $\Gamma \Vdash_{\mathcal{N}(T)} \varphi(y)[\bar{n}]$ for every n, $\Gamma' \to \Gamma$ is an arrow, and $u \in G(\Gamma)$. We need to show $\Gamma' \Vdash_{\mathcal{N}(T)} (\mathrm{St}(y) \to \varphi(y))[u]$. So let us further assume that $\Gamma'' \to \Gamma'$ is an arrow and $\Gamma'' \Vdash_{\mathcal{N}(T)} \mathrm{St}(y)[u]$, and let us show $\Gamma'' \Vdash_{\mathcal{N}(T)} \varphi(y)[u]$.

The assumption $\Gamma'' \Vdash_{\mathcal{N}(T)} \mathrm{St}(y)[u]$ means that for some m, $\Gamma'' \vdash_T u \leq \bar{m}$, and hence $\Gamma'' \vdash_T u = \bar{0} \vee \ldots \vee u = \bar{m}$. This implies that the set of arrows

$$\{\Gamma'' \cup \{u = \bar{0}\} \to \Gamma'', \ldots, \Gamma'' \cup \{u = \bar{m}\} \to \Gamma''\}$$

covers Γ''. For each i, $\Gamma'' \cup \{u = \bar{i}\} \Vdash_{\mathcal{N}(T)} (y = \bar{i})[u]$. By the assumption on Γ and monotonicity, $\Gamma'' \cup \{y = \bar{i}\} \Vdash_{\mathcal{N}(T)} \varphi(\bar{i})$. Since the axioms of equality are forced, $\Gamma'' \cup \{u = \bar{i}\} \Vdash_{\mathcal{N}(T)} \varphi(y)[u]$, again for each i. By the covering property, $\Gamma'' \Vdash_{\mathcal{N}(T)} \varphi(y)[u]$, as required. □

Lemma 3.6 *Suppose* u *is any term of* L*. Then if* Γ *is any type and* $\vec{t}$ *is a sequence of terms in* $G(\Gamma)$ *corresponding to the free variables of* u*, then* $\Gamma \Vdash_{\mathcal{N}(T)} (\neg\mathrm{St}(u))[\vec{t}]$ *if and only if for every* n*,* $\Gamma \vdash_T u[\vec{t}/\vec{x}] > \bar{n}$*.*

Proof. Suppose $\Gamma \Vdash_{\mathcal{N}(T)} (\neg\mathrm{St}(u))[\vec{t}]$. Since for every n, $\Gamma \cup \{u[\vec{t}/\vec{x}] \leq \bar{n}\} \Vdash_{\mathcal{N}(T)} \mathrm{St}(u)[\vec{t}]$, we have that for every n, $\Gamma \cup \{u[\vec{t}/\vec{x}] \leq \bar{n}\} \Vdash_{\mathcal{N}(T)} \bot$. By Lemma 3.3, we have $\Gamma \cup \{u[\vec{t}/\vec{x}] \leq \bar{n}\} \vdash_T \bot$, and so $\Gamma \vdash_T u[\vec{t}/\vec{x}] > \bar{n}$, as desired.

Conversely, suppose that for every n, $\Gamma \vdash_T u[\vec{t}/\vec{x}] > \bar{n}$. Let $\Gamma' \to \Gamma$ be an arrow such that $\Gamma' \Vdash_{\mathcal{N}(T)} \mathrm{St}(u)[\vec{t}]$. Then for some m, we have $\Gamma' \vdash_T u[\vec{t}/\vec{x}] \leq \bar{m}$. Since $\Gamma \vdash_T u[\vec{t}/\vec{x}] > \bar{m}$ and $\Gamma' \vdash_T \Gamma$, we have $\Gamma' \vdash_T \bot$. This implies $\Gamma' \Vdash_{\mathcal{N}(T)} \bot$, so $\Gamma \Vdash_{\mathcal{N}(T)} (\neg\mathrm{St}(u))[\vec{t}]$. □

Lemma 3.7 *Each axiom of (NS) is true in $\mathcal{N}(T)$.*

Proof. The fact that the first axiom is forced follows from the assumption that T evaluates the primitive recursive function symbols correctly.

For the second axiom, suppose Γ is a type, u is in $G(\Gamma)$, and $\Gamma \Vdash_{\mathcal{N}(T)} \neg\neg\mathrm{St}(x)[u]$. Consider the arrow $\Gamma \cup \{u > \bar{n} \mid n \in \mathbb{N}\} \to \Gamma$. By Lemma 3.6, $\Gamma \cup \{u > \bar{n} \mid n \in \mathbb{N}\} \Vdash_{\mathcal{N}(T)} \neg\mathrm{St}(x)[u]$, so $\Gamma \cup \{u > \bar{n} \mid n \in \mathbb{N}\} \Vdash_{\mathcal{N}(T)} \bot$. By Lemma 3.3, $\Gamma \cup \{u > \bar{n} \mid n \in \mathbb{N}\} \vdash_T \bot$. Then there is some m such that $\Gamma \cup \{u > \bar{m}\} \vdash_T \bot$, and so $\Gamma \vdash_T u \leq \bar{m}$. Hence $\Gamma \Vdash_{\mathcal{N}(T)} \mathrm{St}(x)[u]$, as required.

For external induction, suppose Γ is a type, $\varphi(x)$ is a formula of L^{st}, $\vec{t}$ is a sequence of terms in $G(\Gamma)$ corresponding to the free variables of $\forall^{\mathrm{st}} x\ \varphi(x)$, and $\Gamma \Vdash_{\mathcal{N}(T)} \varphi(0)[\vec{t}]$ and $\Gamma \Vdash_{\mathcal{N}(T)} (\forall^{\mathrm{st}} x\ \varphi(x) \to \varphi(x+1))[\vec{t}]$. Then induction in the metatheory shows that for each natural number n, $\Gamma \Vdash_{\mathcal{N}(T)} \varphi(z)[\vec{t}, \bar{n}]$, where z is the first variable after the free variables of φ. By Lemma 3.5, $\Gamma \Vdash \forall^{\mathrm{st}} x\ \varphi(x)$, as required.

For overspill, suppose Γ is a type, $\varphi(x)$ is a formula of L, $\vec{t}$ is a sequence of terms in $G(\Gamma)$ corresponding to the free variables of $\forall^{\mathrm{st}} x\ \varphi(x)$, and $\Gamma \Vdash_{\mathcal{N}(T)} \forall^{\mathrm{st}} x\ \varphi(x)[\vec{t}]$. We need to show $\Gamma \Vdash_{\mathcal{N}(T)} (\exists x\ (\neg\mathrm{St}(x) \wedge \varphi(x)))[\vec{t}]$. For the rest of the argument, we will suppress the terms $\vec{t}$.

By Lemma 3.5, we have $\Gamma \Vdash_{\mathcal{N}(T)} \varphi(z)[\bar{n}]$ for every n. By Lemma 3.3, we have $\Gamma \vdash_T \varphi(\bar{n})$, for each n. Let y be a new variable, and let Γ' be $\Gamma \cup \{\varphi(y)\} \cup \{y \geq n \mid n \in \mathbb{N}\}$. Then $\{\Gamma' \to \Gamma\}$ is a cover of Γ, $\Gamma' \Vdash_{\mathcal{N}(T)} \varphi(z)[y]$, and by Lemma 3.6 $\Gamma' \Vdash_{\mathcal{N}(T)} \neg\mathrm{St}(z)[y]$. Hence $\Gamma \Vdash_{\mathcal{N}(T)} \exists x\ (\neg\mathrm{St}(x) \wedge \varphi(x))$, as desired.

Finally, for underspill, suppose $\Gamma \Vdash_{\mathcal{N}(T)} \forall x\ (\neg\mathrm{St}(x) \to \varphi(x))$. (Again, we will suppress the sequence of terms $\vec{t}$.) Let y be a variable not occurring in Γ. By Lemma 3.6, we have $\Gamma \cup \{y > \bar{n} \mid n \in \mathbb{N}\} \Vdash_{\mathcal{N}(T)} \neg\mathrm{St}(z)[y]$, so $\Gamma \cup \{y > \bar{n} \mid n \in \mathbb{N}\} \Vdash_{\mathcal{N}(T)} \varphi(z)[y]$. By Lemma 3.3, $\Gamma \cup \{y > \bar{n} \mid n \in \mathbb{N}\} \vdash_T \varphi(y)$. In particular, there is a single value of n such that $\Gamma \cup \{y > \bar{n}\} \vdash_T \varphi(y)$. So $\Gamma \vdash_T \varphi(\overline{n+1})$, and hence $\Gamma \Vdash_{\mathcal{N}(T)} \exists^{\mathrm{st}} x\ \varphi(x)$. □

Lemma 3.8 *Axiom (NC) is true in $\mathcal{N}(T)$.*

Proof. Suppose $\Gamma \Vdash_{\mathcal{N}(T)} \forall x\ (\mathrm{St}(x) \vee \neg\mathrm{St}(x))$; we need to show $\Gamma \Vdash_{\mathcal{N}(T)} \bot$. Let y be a variable that does not occur in Γ; then $\Gamma \cup \{y = y\} \Vdash_{\mathcal{N}(T)} (\mathrm{St}(z) \vee \neg\mathrm{St}(z))[y]$. This means that there is a covering $\Gamma_1, \ldots, \Gamma_k, \Delta_1, \ldots, \Delta_l$ of $\Gamma \cup \{y = y\}$ such that for each i, $\Gamma_i \Vdash_{\mathcal{N}(T)} \mathrm{St}(z)[y]$, and for each j, $\Delta_j \Vdash_{\mathcal{N}(T)} \neg\mathrm{St}(z)[y]$. The first fact implies that there is an n such that for each i, $\Gamma_i \Vdash_{\mathcal{N}(T)} (z \leq \bar{n})[y]$. The second fact and Lemma 3.6 implies that for every m, $\Delta_j \Vdash_{\mathcal{N}(T)} (z > \bar{m})[y]$, and, in particular, $\Delta_j \Vdash_{\mathcal{N}(T)} (z > \overline{n+1})[y]$. So each element of the covering proves $y \leq \bar{n} \vee y > \overline{n+1}$, and so Γ proves $y \leq \bar{n} \vee y > \overline{n+1}$. Substituting $\overline{n+1}$ for y shows that $\Gamma \vdash_T \bot$, and hence $\Gamma \Vdash_{\mathcal{N}(T)} \bot$. □

Let ($\forall^{\mathrm{st}}$-ω) be the following infinitary rule:

From $\Gamma \vdash \varphi(\bar{0})$, $\Gamma \vdash \varphi(\bar{1})$, …, conclude $\Gamma \vdash \forall^{\mathrm{st}} x\ \varphi(x)$

where $\varphi(x)$ is any formula of L^{st}.

Theorem 3.9 *$T + (NS) + (NC) + (\forall^{\mathrm{st}}\text{-}\omega)$ is a conservative extension of T.*

Proof. By Lemmata 3.3, 3.7, and 3.8, every sentence of $T + (NS) + (NC)$ is true in $\mathcal{N}(T)$. By Lemma 3.5, truth in $\mathcal{N}(T)$ is closed under the rule $(\forall^{\mathrm{st}}\text{-}\omega)$. Finally, by Lemma 3.3 anything in the language of T that is true in $\mathcal{N}(T)$ is provable in T. $\square$

Taking T to be Heyting arithmetic, *HA*, yields Corollary 4.11 of Moerdijk and Palmgren [11]:

Corollary 3.10 *$HAI + (NC) + (\forall^{\mathrm{st}}\text{-}\omega)$ is a conservative extension of HA.*

Relying on arbitrary types in the category $\mathcal{D}$ is somewhat lavish. The construction still works if we restrict the category to computable types; in fact, we can restrict the category to types that can be represented with finite amounts of information in a simple way, as follows.

Let the set of *benign* types be given by the following inductive definition:

- If Γ is a finite set of formulae, it is benign.
- If Γ and Δ are benign types, $\varphi(x)$ is a formula, and t is a term, then $\Delta \cup \{t \geq \bar{n} \mid \Gamma \vdash_T \varphi(\bar{n})\}$ is benign.

It is not hard to see that the benign types are closed under finite unions. By varying the choices of Γ, Δ, φ, and t we can obtain the various types used in the lemmata above; for example, taking $\Delta = \Gamma \cup \{\varphi(y)\}$ and $t = y$, one has the set referred to in the second clause of our definition of covering, and taking $\Delta = \emptyset$ and $\varphi(x) = \top$, we have the set $\{t \geq \bar{n} \mid n \in \mathbb{N}\}$. As a result, we have the following:

Lemma 3.11 *The lemmata in this section still hold even if one restricts the objects of the category $\mathcal{D}$ to benign types.*

In fact, we can restrict our attention to finite *presentations* of benign types. Define the set of such presentations inductively, as follows:

- If $\varphi_0, \ldots, \varphi_k$ are formulae of L, $c_0, \ldots, c_l$ are presentations, $\psi_0(x_0), \ldots, \psi_l(x_l)$ are formulae of L with distinguished variables $x_0, \ldots, x_l$, and $t_1, \ldots, t_l$ are terms, then

$$\langle\langle \varphi_0, \ldots, \varphi_k \rangle, \langle c_0, \psi_0(x_0), t_0 \rangle, \ldots, \langle c_l, \psi_l(x_l), t_l \rangle\rangle$$

  is a presentation.

Here l is allowed to be 0, getting the definition off the ground. The idea is that each presentation d of the form just displayed represents the type $Tp(d)$, given by

$$\{\varphi_0, \ldots, \varphi_k\} \cup \{t_0 \geq \bar{n} \mid Tp(c_0) \vdash_T \psi_0(\bar{n})\} \cup \ldots \cup \{t_l \geq \bar{n} \mid Tp(c_l) \vdash_T \psi_l(\bar{n})\}.$$

If d' is another presentation

$$\langle\langle\varphi'_0,\ldots,\varphi'_{k'}\rangle,\langle c'_0,\psi'_0(x'_0),t'_0\rangle,\ldots,\langle c'_{l'},\psi'_{l'}(x'_{l'}),t'_{l'}\rangle\rangle,$$

say d *formally includes* d' if

$$\{\varphi_0,\ldots,\varphi_k\}\supseteq\{\varphi'_0,\ldots,\varphi'_{k'}\}$$

and

$$\{\langle c_0,\psi_0(x_0),t_0\rangle,\ldots,\langle c_l,\psi_l(x_l),t_l\rangle\}\supseteq\{\langle c'_0,\psi'_0(x'_0),t'_0\rangle,\ldots,\langle c'_{l'},\psi'_{l'}(x'_{l'}),t'_l\rangle\}.$$

Clearly, if d formally includes d' then $Tp(d)$ includes $Tp(d')$, but the converse does not always hold. The set of free variables of a presentation is defined in the obvious way, and if σ is a renaming of the free variables of d, d^σ is what one would expect it to be. Also, if d is a presentation, say $d\vdash_T\varphi$ if $Tp(d)\vdash_T\varphi$. Then we have

Lemma 3.12 *The lemmata in this section still hold even if one takes the objects of the category $\mathcal{D}$ to be presentations of types, with arrows $d\xrightarrow{\sigma}e$ whenever d formally includes e^σ.*

Note that $\langle\langle\rangle\rangle$ is a terminal object in the category.

4 Internalizing the construction

In this section and the next it will be convenient to restrict our attention to theories T that include primitive recursive arithmetic. In other words, we will take the language of arithmetic to include symbols for the primitive recursive functions and relations, and assume that T includes their defining axioms and induction for atomic formulae. A formula is then said to be Σ_1 if it is of the form $\exists\vec{x}\ R(\vec{x},\vec{y})$, where R is primitive recursive. Since sequences of numbers can be coded as single numbers in a primitive recursive way, every Σ_1 formula is equivalent to one with a single existential quantifier. We will take Heyting arithmetic, *HA*, to be the intuitionistic theory extending primitive recursive arithmetic with the full schema of induction. $I\Sigma^i_1$ denotes the theory obtained from *HA* by restricting the induction schema to Σ_1 formulae.[2]

There is something odd about Theorem 3.9. Suppose T is a weak theory of arithmetic. In a sense, $T+(NS)$ is strong, since it interprets Heyting arithmetic: the axioms of *HA* hold in $T+(NS)$ when they are relativized to the standard numbers, so whenever *HA* proves φ, T proves its relativization, φ^{st}. In another sense, $T+(NS)$ is weak; Theorem 3.9 shows that it proves the same statements as T in the common language. In fact, the infinitary rule ($\forall^{\mathrm{st}}$-ω) allows us to add axioms of the form $\forall^{\mathrm{st}}x\ \varphi$, where $\forall x\ \varphi$ is an arbitrary true Π_1 sentence, without affecting provability in the original language.

[2] *HA* and $I\Sigma^i_1$ are often presented in a language with only $0,1,+,\times$, and $<$. Since the primitive recursive functions can be introduced in a definitional extension of this latter version of $I\Sigma^i_1$, the difference is minor.

What is at play here is the lack of "transfer principles" that relate the standard and nonstandard universes. Moerdijk and Palmgren show in [11] that even mild forms of transfer imply new instances of the law of the excluded middle, so in general these need to be avoided. Nonetheless, in this section we will show that in certain cases on *can* safely add restricted forms of transfer.

Let $(\forall^{\mathrm{st}}\text{-}TR)$ be the following transfer rule:

$$\text{If} \vdash \forall^{\mathrm{st}} x\ \varphi \text{ then} \vdash \forall x\ \varphi \qquad (\forall^{\mathrm{st}}\text{-}TR)$$

where φ is a formula of L, which is to say, φ does not mention the predicate St. Note that $(\forall^{\mathrm{st}}\text{-}TR)$ concerns provability without hypotheses, but it extends to provability with hypotheses in L: if $\{\psi_1, \ldots, \psi_k\} \vdash \forall^{\mathrm{st}} x\ \varphi$, then $\vdash \forall^{\mathrm{st}} x\ (\psi_1 \wedge \ldots \psi_k \rightarrow \varphi)$; and from $\forall x\ (\psi_1 \wedge \ldots \psi_k \rightarrow \varphi)$ one can conclude $\psi_1 \wedge \ldots \psi_k \rightarrow \forall x\ \varphi$. $(\forall^{\mathrm{st}}\text{-}TR)$ also yields transfer for provable sentences of the form $\forall^{\mathrm{st}} x\ \exists^{\mathrm{st}} y\ R(x, y)$, since $\exists^{\mathrm{st}} y\ R(x, y)$ implies $\exists y\ R(x, y)$. As a result, adding it to a nonstandard theory guarantees that theory some strength:

Proposition 4.1 *Let T satisfy the requirements above. Then whenever HA proves a Π_2 sentence $\forall x\ \exists y\ R(x, y)$, so does $T + (NS) + (\forall^{\mathrm{st}}\text{-}TR)$.*

Proof. Since the relativization of each axiom of HA to the standard numbers is provable in $T + (NS)$, if HA proves $\forall x\ \exists y\ R(x, y)$, then $T + (NS)$ proves $\forall^{\mathrm{st}} x\ \exists^{\mathrm{st}} y\ R(x, y)$. The transfer rule then yields $\forall x\ \exists y\ R(x, y)$. □

We will show that in the case where T is HA, $(\forall^{\mathrm{st}}\text{-}TR)$ is in fact a derived rule.

Theorem 4.2 *Suppose HAI proves $\forall^{\mathrm{st}} x\ \varphi$, where φ is a formula of L. Then HA proves $\forall x\ \varphi$.*

This yields

Corollary 4.3 *$HAI + (\forall^{\mathrm{st}}\text{-}TR)$ is a conservative extension of HA.*

The proof of Theorem 4.2 involves "internalizing" the construction of Section 3, and using the fact that HA proves uniform reflection principles for each of its finitely axiomatized fragments. Paying careful attention to the complexity of defined concepts allows us to extend the results to weaker theories. Let (NS') consist of the axioms of (NS), with induction restricted to formulae $\theta(\vec{w}, x)$ of the form $\exists \vec{y}\ \exists^{\mathrm{st}} \vec{z}\ R(\vec{y}, \vec{z}, \vec{w}, x)$, where R is primitive recursive. Let $(\forall^{\mathrm{st}}\text{-}TR')$ denote the restriction of $(\forall^{\mathrm{st}}\text{-}TR)$ to Σ_1 formulae φ (in this case, one can allow primitive recursive hypotheses). Finally, let PRA denote primitive recursive arithmetic. Then we have

Theorem 4.4 *Suppose $PRA + (NS') + (\forall^{\mathrm{st}}\text{-}TR')$ proves $\forall^{\mathrm{st}} x\ \varphi$, where φ is a Σ_1 formula of L. Then PRA proves $\forall x\ \varphi$.*

Corollary 4.5 *$PRA + (NS') + (\forall^{\mathrm{st}}\text{-}TR')$ is a Π_2 conservative extension of PRA. In other words, if the former theory proves $\forall x\ \exists y\ R(x, y)$, where R is primitive recursive, then the latter theory proves it as well.*

Fix reasonable (primitive recursive) encodings of logical syntax, ordered pairs, finite sequences, finite sets, etc. To internalize the argument from the previous section, we need to know that the relevant notions can be represented in the language of arithmetic in an appropriate way.

Lemma 4.6 *The following are primitive recursive:*

1. *the unary relation, "c is a presentation of a benign type"*
2. *the function $FV(c)$ which returns the set of free variables of a presentation c*
3. *assuming c and d are presentations, the relation "σ is an injective function from $FV(c)$ to $FV(d)$"*
4. *the relation "c formally includes d"*

Assuming the axioms of T are represented by a Σ_1 formula, the following are defined by Σ_1 formulae:

1. *the relation "$c \vdash_T \varphi$" between presentations c and (codes for) formulae φ*
2. *the relation "$c_1 \xrightarrow{\sigma_1} d, \ldots, c_k \xrightarrow{\sigma_k} d$ covers d," between finite sets of arrows and presentations*

These notions can be represented in such a way that the relevant properties are provable in PRA.

Proof. We leave it to the reader to verify that the functions and relations in the first four clauses are primitive recursive.

The next two relations are described by inductive definitions, where the inductive clauses are Σ_1. By a formalized version of the recursion theorem, the relations themselves are Σ_1. Alternatively, one can provide explicit Σ_1 definitions: for example, "$c \vdash_T \varphi$" holds if and only if there is a finite sequence of pairs $\langle d_i, \psi_i \rangle$ of presentations and formulae, such that the sequence includes $\langle c, \varphi \rangle$, and such that for each i, if d_i is of the form

$$\langle \langle \theta_0, \ldots, \theta_k \rangle, \langle e_0, \eta_0(x_0), t_0 \rangle, \ldots, \langle e_l, \eta_l(x_l), t_l \rangle \rangle$$

then one of the following holds:

- ψ_i is one of the θ_j, or
- ψ_i if of the form $t_j \geq \bar{n}$, and $\langle e_j, \eta_j(\bar{n}) \rangle$ occurs earlier in the sequence
- ψ_i is an axiom of intuitionistic logic

- ψ_i follows from formulae $\nu_1, \ldots, \nu_m$ by a rule of inference of intuitionistic logic, and $\langle d_i, \nu_1 \rangle$, ..., $\langle d_i, \nu_m \rangle$ occur earlier in the sequence.

We leave it to the reader to verify that this description is Σ_1, and that the corresponding closure properties are provable in *PRA*. $\square$

Our construction of a nonstandard model was carried out with an eye towards keeping the complexity of defined notions low. For example, according to Lemma 4.6 the set of morphisms in our underlying category is primitive recursive, whereas in the category used in [11] it is inherently Π_2.

Given a theory T represented by a Σ_1 formula, the inductive clauses of the forcing relation assign to each fixed formula $\psi(x_1, \ldots, x_k)$ with the free variables shown a formula $c \Vdash_{\mathcal{N}(T)} \psi[t_1, \ldots, t_k]$, depending on (codes for) a presentation c and the sequence of (codes for) terms $t_1, \ldots, t_k$. Also, for each fixed formula $\psi(x_1, \ldots, x_k)$, using Lemma 4.6 and a primitive recursive substitution function we have that the relation $c \vdash_T \psi(t_1, \ldots, t_k)$ is represented by a Σ_1 formula in $c, t_1, \ldots, t_k$ as well. Assuming our basic assumptions on T are provable in *PRA*, internalizing the arguments behind the proofs of Lemmata 3.1, 3.3, and 3.7 yields the following:

Lemma 4.7 *For each formula $\psi(x_1, \ldots, x_k)$ of L with the free variables shown, PRA proves the following: if $\{c_1 \to d, \ldots, c_l \to d\}$ is any covering of d and $t_1, \ldots, t_k$ are any terms whose free variables are among those of d, then $d \vdash_T \psi(t_1, \ldots, t_k)$ if and only if for each i, $c_i \vdash_T \psi(t_1, \ldots, t_k)$.*

Lemma 4.8 *Let $\psi(x_1, \ldots, x_k)$ be any formula of L with the free variables shown. Then PRA proves the following: if c is a presentation and $t_1, \ldots, t_k$ are terms whose free variables are among those of c, then $c \Vdash_{\mathcal{N}(T)} \psi[t_1, \ldots, t_k]$ if and only if $c \vdash_T \psi(t_1, \ldots, t_k)$.*

Lemma 4.9 *If ψ is any axiom of (NS), then HA proves $\Vdash_{\mathcal{N}(T)} \psi$. If ψ is any axiom of (NS′), then $I\Sigma_1^i$ proves $\Vdash_{\mathcal{N}(T)} \psi$.*

Proof. The proof of the last lemma involves a straightforward formalization of the proof of Lemma 3.7. Note that if ψ is an instance of external induction, one can prove $\Vdash_{\mathcal{N}(T)} \psi$ using an instance of ordinary induction in *HA*. For the second claim, we need to know that if $\theta(\vec{w}, x)$ is of the form $\exists \vec{y}\, \exists^{\mathrm{st}} \vec{z}\, R(\vec{y}, \vec{z}, \vec{w}, x)$, c is a presentation, and $\vec{t}$ is a sequence of terms corresponding to $\vec{w}$, then $c \Vdash_{\mathcal{N}(T)} \theta(\vec{w}, x)[\vec{t}, \bar{n}]$ is equivalent to a Σ_1 formula in c, $\vec{t}$, and n. But unwinding definitions shows that this is equivalent to saying "there exists a cover $\{d_1 \to c, \ldots, d_k \to c\}$ of c, sequences of terms $\vec{u}_i \in G(d_i)$, and sequences of numerals $\vec{m}_i$, such that for each $i \leq k$, $d_i \vdash_T R(\vec{u}_i, \vec{m}_i, \vec{t}, \bar{n})$." Using Σ_1 collection, provable in $I\Sigma_1^i$, to move the existential quantifier in $d_i \vdash_T R(\vec{u}_i, \vec{m}_i, \vec{t}, \bar{n})$ through the bounded universal quantifier, this is equivalent to a Σ_1 formula. $\square$

We can now prove Theorems 4.2 and 4.4.

Proof of Theorem 4.2. Suppose HAI proves $\forall^{\mathrm{st}} x\ \varphi(x)$, where $\varphi(x)$ is a formula in L. Let Γ be the finite set of axioms of HA (not (NS)) that are used in the proof, and let T be the theory axiomatized by Γ.

Now consider the forcing relation associated with $\mathcal{N}(T)$ in HA. By induction on the length of proof, we can show that if $T + (NS)$ proves a formula ψ, then HA proves that for every sequence $\vec{t}$ of terms corresponding to the free variables of ψ, $\Vdash_{\mathcal{N}(T)} \psi[\vec{t}]$. If ψ is one of the axioms of T, this follows from Lemma 4.8; if ψ is an axiom in (NS), this follows from Lemma 4.9; and otherwise the claim follows from the inductive hypothesis and the fact that the semantics is sound for intuitionistic logic.

As a result, HA proves $\Vdash_{\mathcal{N}(T)} \forall^{\mathrm{st}} x\ \varphi(x)$. Hence it also proves that for every natural number x, $\Vdash_{\mathcal{N}(T)} \varphi(z)[\bar{x}]$. By Lemma 4.8, HA proves that for every x, $\vdash_T \varphi(\bar{x})$. But since HA proves the uniform reflection principle

$$\forall x\ (Prov_T(\ulcorner \varphi(\bar{x}) \urcorner) \rightarrow \varphi(\bar{x}))$$

for every finite subtheory T (see, for example, [15]), it proves $\forall x\ \varphi(x)$. $\square$

Proof of Theorem 4.4. As in the last argument, if $PRA + (NS')$ proves $\forall^{\mathrm{st}} x\ \varphi(x)$, there is a finite subtheory T of PRA such that $I\Sigma_1^i$ proves that for every x, $\vdash_T \varphi(\bar{x})$. Since $I\Sigma_1^i$ proves the uniform reflection principle for any finite subtheory of PRA, it proves $\forall x\ \varphi(x)$. The desired conclusion follows from the well-known result that $I\Sigma_1^i$ (as well as the classical version, $I\Sigma_1$) is conservative over PRA for Π_2 sentences. (See, for example, [2].) $\square$

In the case of HA, we can add yet another form of transfer. A formula in the language of arithmetic is said to be *negative* if there are no occurrences of $\vee$ or $\exists$. It is well known that using the double-negation translation one can show that classical arithmetic, PA, is conservative over HA for negative sentences. On the other hand, by carrying out the usual model-theoretic construction of a nonstandard model of arithmetic inside PA (more precisely, describing nonstandard Σ_k-elementary extensions of the natural numbers, for arbitrarily large k), one can show that $PA + (NS)$ together with the full schema of transfer axioms is conservative over PA.[3] Putting these facts together, one has

Theorem 4.10 *If φ is a negative formula in L and HAI proves φ^{st}, then HA proves φ.*

5 Transfer principles

We have shown that if HAI proves $\forall^{\mathrm{st}} x\ \exists^{\mathrm{st}} y\ \varphi(x, y)$, where φ is any formula in the language of arithmetic, then HA proves $\forall x\ \exists y\ \varphi(x, y)$; and similarly, if HAI proves φ^{st}, where φ is any negative formula in the language of arithmetic,

[3] In this context, overspill and underspill are redundant. This conservation result was obtained by Friedman in the 1960's, but was never published. For a proof of this theorem using, instead, and internalized ultrapower construction, see [14].

then HA proves φ. This means what we can conservatively augment HAI with the corresponding transfer rules. In this section we will show that even minor strengthenings of these rules destroy conservativity over HA, as do even trivial instances of the transfer axioms.

In the discussion which follows, we will use PRA^* to denote the fragment of HAI axiomatized by the axioms of PRA, an axiom asserting the existence of a nonstandard number, an axiom asserting that the standard numbers form an initial segment closed under successor, and the axiom $\neg\neg\mathrm{St}(x) \to \mathrm{St}(x)$. The proofs below use the fact that PRA proves the law of the excluded middle and the least element principle for primitive recursive formulae, and the fact that the primitive recursive relations are closed under boolean operations and bounded quantification.

Lemma 5.1 *PRA^* proves Markov's principle relativized to the standard numbers, for primitive recursive predicates, possibly with nonstandard parameters. In other words, PRA^* proves*

$$\neg\forall^{\mathrm{st}} x\ R(x, \vec{y}) \to \exists^{\mathrm{st}} x\ \neg R(x, \vec{y})$$

for any primitive recursive relation R.

Proof. Argue in PRA^*. For convenience we will leave off the parameters $\vec{y}$. Suppose $\neg\forall^{\mathrm{st}} x\ R(x)$, and let w be nonstandard. Then $\neg\forall x < w\ R(x)$. By the law of the excluded middle for primitive recursive relations, $\exists x < w \neg R(x)$. Let x be the least value satisfying $R(x)$. We claim x is standard, witnessing the statement $\exists^{\mathrm{st}} x\ \neg R(x)$.

Suppose otherwise. Then by the leastness of x, we have $\forall y < x\ R(y)$, and hence $\forall^{\mathrm{st}} y\ R(y)$, contradicting the original assumption. So we have $\neg\neg\mathrm{St}(x)$, and hence $\mathrm{St}(x)$. □

Lemma 5.2 *There is a unary primitive recursive relation $A(x)$ such that HA does not prove $\neg\forall x\ A(x) \to \exists x\ \neg A(x)$.*

Proof. Section 1.11.5 of [15] gives an example, taking $\forall x\ A(x)$ to be a Rosser sentence. □

Proposition 5.3 *There is a unary primitive recursive relation $A(x)$ such that HAI together with the transfer axiom $\forall^{\mathrm{st}} x\ A(x) \to \forall x\ A(x)$ is not conservative over HA. There is also a unary primitive recursive relation $B(x)$ such that HAI together with the transfer axiom $\exists x\ B(x) \to \exists^{\mathrm{st}} x\ B(x)$ is not conservative over HA.*

Proof. For the first statement, let A be as in the previous lemma. One the one hand, $\neg\forall x\ A(x) \to \exists x\ \neg A(x)$ is not provable in (HA). On the other hand, with the additional axiom and Lemma 5.1, it is provable in HAI: the extra

axiom implies $\neg\forall x\ A(x) \to \neg\forall^{\mathrm{st}} x\ A(x)$; using Lemma 5.1, $\neg\forall^{\mathrm{st}} x\ A(x)$ implies $\exists^{\mathrm{st}} x\ \neg A(x)$, which implies $\exists x\ \neg A(x)$.

For the second statement, let $B(x)$ be $\neg A(x)$. Then the second axiom implies its contrapositive which, using the law of the excluded middle for primitive recursive relations, is equivalent to the first axiom. $\square$

This shows that one cannot add even trivial transfer axioms to *HAI* without violating conservativity. Now let us show that one cannot strengthen the transfer rules. Consider the schema

$$\forall x, y\ (S(x, \vec{z}) \vee T(y, \vec{z})) \to \forall x\ S(x, \vec{y}) \vee \forall y\ T(y, \vec{z})$$

for primitive recursive relations S and T. This schema is considered by van Oosten in [18].

Lemma 5.4 *PRA^* proves the schema above with the quantifiers relativized to standard numbers, where $\vec{z}$ may be nonstandard. In other words, PRA^* proves*

$$\forall^{\mathrm{st}} x, y\ (S(x, \vec{z}) \vee T(y, \vec{z})) \to \forall^{\mathrm{st}} x\ S(x, \vec{z}) \vee \forall^{\mathrm{st}} y\ T(y, \vec{z}).$$

Proof. For convenience, again we will leave off the parameters $\vec{z}$. Arguing in PRA^*, suppose $\forall^{\mathrm{st}} x, y\ (S(x) \vee T(y))$, and let w be nonstandard. By the law of the excluded middle for primitive recursive relations, we have either $\forall u < w\ (S(u) \wedge T(u))$ or $\exists u < w\ (\neg S(u) \vee \neg T(u))$. In the first case, we have both $\forall^{\mathrm{st}} x\ S(x)$ and $\forall^{\mathrm{st}} y\ T(y)$, each of which implies the desired conclusion. So let us focus on the second case.

Let u be the least value satisfying $\neg S(u) \vee \neg T(u)$. In the first subcase, we have $\neg S(u)$. We claim that this implies $\forall^{\mathrm{st}} y\ T(y)$. To see this, let y be standard. If $\neg T(y)$, then by the leastness of u, we have that u is less than or equal to y, and hence standard as well. But then we have $\neg S(u) \wedge \neg T(y)$, contrary to our assumption. Since y was an arbitrary standard number, this implies $\forall^{\mathrm{st}} y\ T(y)$.

Symmetrically, if $\neg T(u)$, then $\forall^{\mathrm{st}} x\ S(x)$. Either way, we have the desired conclusion. $\square$

Lemma 5.5 *There are primitive recursive relations $C(x)$ and $D(x)$ such that HA proves $\forall x, y\ (C(x) \vee D(y))$, but HA does not prove $\forall x\ C(x) \vee \forall y\ D(y)$.*

Proof. This proof is due to Friedman. Using the fact that classical arithmetic, *PA*, is Π_2 conservative over *HA*, and the fact that *HA* has the disjunction property, it suffices to find C and D such that *PA* proves $\forall x, y\ (C(x) \vee D(y))$, but proves neither $\forall x\ C(x)$ nor $\forall y\ D(y)$.

Using the fixed-point lemma, let φ be a formula equivalent to the statement $\forall x\ C(x)$, where $C(x)$ says

> if x is a proof of φ in *PA*, there is a shorter proof of $\neg\varphi$.

Here "shorter" refers to a reasonable Gödel numbering of proofs. Then let ψ be of the form $\forall y\ D(y)$, where $D(y)$ says

if y is a proof of $\neg\varphi$ in PA, then there is a shorter proof of φ.

It is not hard to show that neither φ nor ψ is provable in PA. But PA does prove $\forall x, y\ (C(x) \vee D(y))$. To see this, argue in PA. Suppose $\neg C(x)$. Then x is a proof of φ in PA, and there is no shorter proof of $\neg\varphi$. We need to show $D(y)$. To that end, suppose y is a proof of $\neg\varphi$ in PA; then we have $y > x$, which means x is a shorter proof of φ. □

The next three propositions give the precise sense in which the transfer rules of Section 4 are optimal.

Proposition 5.6 *There are primitive recursive relations $C(x)$ and $D(y)$ such that HAI proves $\forall^{\mathrm{st}} x\ C(x) \vee \forall^{\mathrm{st}} y\ D(y)$, but HA does not prove $\forall x\ C(x) \vee \forall y\ D(y)$.*

Proof. Use the C and D from Lemma 5.5. Then HAI proves $\forall x, y\ (C(x) \vee D(y))$, and hence $\forall^{\mathrm{st}} x, y\ (C(x) \vee D(y))$. By Lemma 5.4 it proves $\forall^{\mathrm{st}} x\ C(x) \vee \forall^{\mathrm{st}} y\ D(y)$. On the other hand, by Lemma 5.5, $\forall x\ C(x) \vee \forall y\ D(y)$ is not provable in HA. □

Proposition 5.7 *There is a primitive recursive relation $E(z, x)$ such that HAI proves $\exists^{\mathrm{st}} z\ \forall^{\mathrm{st}} x\ E(z, x)$, but HA does not prove $\exists z\ \forall x\ E(z, x)$.*

Proof. Let $E(z, x)$ be $((z = 0 \wedge C(x)) \vee (z = 1 \wedge D(y))$, with C and D as above. □

Proposition 5.8 *There are a negative sentence of arithmetic φ and a primitive recursive relation $F(x)$, such that HAI proves $\varphi^{\mathrm{st}} \to \exists^{\mathrm{st}} x\ F(x)$, but HA does not prove $\varphi \to \exists x\ F(x)$.*

Proof. Let φ be the $\neg\forall x\ A(x)$, where $A(x)$ is as in Lemma 5.2, and let $F(x)$ be $\neg A(x)$. □

With no extra work, we have the following two corollaries.

Corollary 5.9 *HAI does not have the disjunction property.*

Proof. By Proposition 5.6, HAI proves $\forall^{\mathrm{st}} x\ C(x) \vee \forall^{\mathrm{st}} y\ D(y)$, but by Theorem 4.2, it does not prove either disjunct. □

Corollary 5.10 *HAI does not have the explicit definability property. In fact, there is a formula ψ such that HAI proves $\exists^{\mathrm{st}} x\ \psi(x)$ but there is no term t such that it proves $\psi(t)$.*

Proof. The first statement follows from the simple observation that HAI proves $\exists x\ \neg\mathrm{St}(x)$, but doesn't prove $\neg\mathrm{St}(t)$ for any term t. But the second, stronger statement follows from Proposition 5.7 and Theorem 4.2. $\square$

If one is only interested in Π_2 statements, however, the situation is entirely different. On the one hand, classical arithmetic, PA, is conservative over HA for such sentences (see [7]). On the other hand, we have already noted that $PA + (NS)$ together with the full schema of transfer axioms is conservative over PA. The net effect is that classical nonstandard arithmetic with full transfer is Π_2 conservative over HA.

Similarly, classical $I\Sigma_1$ is Π_2 conservative over $I\Sigma_1^i$ (see [1] or [5] for a direct translation). In this context, however, we do not know how much transfer one can conservatively add to a nonstandard version of $I\Sigma_1$. In particular, we do not know if Theorem 4.4 remains true if one replaces the transfer rule with the restricted transfer axioms of Proposition 5.3.

References

[1] Jeremy Avigad. Interpreting classical theories in constructive ones. *Journal of Symbolic Logic*, 65:1785–1812, 2000.

[2] Samuel Buss. First-order proof theory of arithmetic. In Samual Buss, editor, *The Handbook of Proof Theory*, 79–147. North-Holland, 1998.

[3] Carsten Butz. The filter construction revisited. Preprint, 1999.

[4] Carsten Butz and Peter Johnstone. Classifying toposes for first-order theories. *Annals of Pure and Applied Logic*, 91:33–58, 1998.

[5] Thierry Coquand and Martin Hofmann. A new method for establishing conservativity of classical systems over their intuitionistic version. *Mathematical Structures in Computer Science*, 9:323–333, 1999.

[6] Michael Fourman and Dana Scott. Sheaves and logic. In Fourman et al., editor, *Applications of Sheaves*, Lecture Notes in Mathematics #753, pages 302–401. Springer, 1977.

[7] Harvey Friedman. Classically and intuitionistically provable functions. In H. Müller and D. Scott, editors, *Higher Set Theory*, Lecture Notes in Mathematics #669, pages 21–27. Springer, 1978.

[8] Saunders Mac Lane and Ieke Moerdijk. *Sheaves in Geometry and Logic*. Springer, 1992.

[9] Michael Makkai and Gonzalo Reyes. *First Order Categorical Logic*. Lecture Notes in Mathematics #611. Springer, 1977.

[10] Ieke Moerdijk. A model for intuitionistic non-standard arithmetic. *Annals of Pure and Applied Logic*, 73:37–51, 1995.

[11] Ieke Moerdijk and Erik Palmgren. Minimal models of Heyting arithmetic. *Journal of Symbolic Logic*, 62:1448–1460, 1997.

[12] Erik Palmgren. Constructive sheaf semantics. *Mathematical Logic Quarterly*, 43:321–327, 1997.

[13] Erik Palmgren. Developments in constructive nonstandard analysis. *Bulletin of Symbolic Logic*, 4:223–272, 1998.

[14] Erik Palmgren. An effective conservation result for nonstandard arithmetic. *Mathematical Logic Quarterly*, 46:17–23, 2000.

[15] A. S. Troelstra. *Metamathematical Investigation of Intuitionistic Arithmetic and Analysis*. Lecture Notes in Mathematics #344. Springer, 1973.

[16] A. S. Troelstra and Dirk van Dalen. *Constructivism in Mathematics: An Introduction*. North-Holland, 1988.

[17] Dirk van Dalen. *Logic and Structure*. Springer, third edition, 1997.

[18] Jaap van Oosten. Two remarks on the Lifschitz realizability topos. *Journal of Symbolic Logic*, 61:70–79, 1996.