Australasian Journal of Logic

On Not Saying What We Shouldn't Have to Say

Shay Logan and Graham Leach-Krouse

Kansas State University

Abstract

In this paper we introduce a novel way of building arithmetics whose background logic is R. The purpose of doing this is to point in the direction of a novel family of systems that could be candidates for being the infamous $R^{\#\frac{1}{2}}$ that Meyer suggested we look for.

Introduction

In casual practice, we commonly allow the string "$\phi \wedge \psi \wedge \chi$" to indicate the same formula as the string "$(\phi \wedge \psi) \wedge \chi$". This observation about our casual practices is connected in subtle and interesting ways to a point that comes up in *The Consistency of Arithmetic*. The context is Meyer's discussion of his A13—the arithmetical axiom that 'says' distinct numbers have distinct successors. Meyer finds this axiom unpleasant. His complaint is as follows:

> The trouble is that this is not the sort of thing that we should *have to say*; if we think of the natural numbers as built up from 0 by applying the successor operation, it would seem that we could take for granted that each result of applying the operation produces a new unique number, unless a *specific assumption* is made to the contrary.

What Meyer seems to be getting at is this: it is part of the nature of basic arithmetical practice that distinct terms built up from successor and zero denote distinct numbers. This assumption should be in play unless it is somehow overridden (as in modular arithmetic, for example). But the fact that we feel the need to *explicitly assert* that things with identical successors are themselves identical suggests exactly the opposite. As a result, A13 smells like a post-hoc regimentation of practice, rather than a clear expression of a principle internal to arithmetic.[1]

Meyer's principle that distinct terms should denote distinct numbers can be put to work in the contrapositive: terms that transparently denote the same number should not be distinguished. In the system we build below, we will attempt to work this idea out but a bit more thoroughly, a bit less casually, and a bit more explicitly. In particular, we will construct our syntax so that many things that would traditionally be taken to be distinct expressions are in fact not. We will be adopting, that is, a large number of identities *at the level of syntax.*

There are two technical matters that will need settling. The first is to figure out what 'transparently denote the same number' means in its occurrence in 'terms that transparently denote the same number should not be distinguished.' We take a fairly liberal view on this matter. If e and e' are expressions and we can transform e into e' using primitive recursive (henceforth *p.r.*) means (in a sense to be specified below) then e and e' will simply not be distinguished. Thus, in particular, if the term t_1 can be transformed into the term t_2 using p.r. means, then t_1 and t_2 will in fact be the very same term. In typical presentations of arithmetic, of course, if the term t_1 can be transformed by p.r. means into the term t_2, then the sentence $t_1 = t_2$ will be provable. That will also be the case here, but for the simple reason that $t_1 = t_2$ will now be an instance of the trivial identity $t = t$. All that to say this: we adopt in this paper the view that 'transparently denote the same number' means 'can be seen, using only p.r. means, to denote the same number'.

The second technical matter is determining how to *enforce* identities at the level of syntax. We do so in the following way. First, we define a rather traditional sort of language. We then define a series of rewrite rules that

[1]A geometrical analogy might be helpful. There's a difference between the parallel postulate, and Hilbert's axiom of line completeness. The former is part of traditional geometrical practice, while the latter is something required to squeeze geometrical practice into the formal framework of Hilbert's Foundations of Geometry. For more on the ways in which axioms intended to impose logical rigor can in fact distort and obscure geometrical practice, see [16].

(as we show) capture the force of primitive recursion. The rules give rise in turn to an equivalence relation on the expressions of our base language. We then define our logic not on the base language, but on a language whose elements are the equivalence classes of expressions in the base language. Since the equivalence relation enforces the rewrite rules which in turn capture the force of primitive recursion, p.r. equivalent expressions in the language on which we define our logic end up being identical.

Since this paper will be part of a volume dedicated to some of Meyer's work, we think it's important to make the connections between our work and Meyer's as explicit as possible. One such connection is given by the quote mentioned above and our working out the idea of identity at the level of syntax. There's also another, more obvious connection: we, like Meyer, are interested in looking at arithmetic from a relevant perspective. On this front, we also offer something novel, which we will take a moment to get clear about.

At the end of [7], the basic problem haunting the relevant arithmetician is broached:

> Whither relevant arithmetic? This was our opening question. We leave it open. But $R^{\#}$ having failed and $R^{\#\#}$ being infinitary, the immediate task is to find $R^{\#\frac{1}{2}}$.

The failures Meyer mentions here are rather serious: earlier in the paper he and Friedman had shown that a formula they call QRF, which is a theorem of Peano arithmetic, is not provable in $R^{\#}$. On its own, the existence of a theorem of Peano arithmetic that is not a theorem of $R^{\#}$ seems tolerable, provided the theorem in question be sufficiently arithmetically unimportant (whatever that turns out to mean). But QRF is not just any old formula. Rather, QRF expresses the claim that for any odd prime p, there is a positive integer y that is not a quadratic residue mod p. This theorem was proved by Gauss. It's mentioned on the Wikipedia page for quadratic residues. It is, in short, a paradigmatic case of a theorem that any decent theory of arithmetic needs to capture. $R^{\#}$ doesn't. So $R^{\#}$ is indecent.

$R^{\#\#}$ rectifies this, and then some. In fact, $R^{\#\#}$ captures all formulas valid in the standard model of arithmetic. But this is because $R^{\#\#}$ sins in a different way: it uses an ω rule. This is again indecent—any decent formal arithmetic, after all, ought to restrict itself to finitary reasoning.

To our knowledge there's been relatively little progress in the nearly three decades since Meyer made the comment above on the 'immediate task' of finding $R^{\#\frac{1}{2}}$. This lack of progress is (we think) largely owed to the fact that it's hard to know where to look—there really just aren't many obvious candidates for the '$R^{\#\frac{1}{2}}$-role'.

One of the things we offer in this paper is a natural way of generating such candidates. We'll have more to say about this as we go along, but the basic idea is that we provide a technical apparatus that has a range of adjustable features (I picture them as dials that can be twisted, levers that can be pulled, buttons that can be pushed, etc.). By manipulating this machinery, one can build theories that differ in various ways from the theory we present. It's clear enough which 'settings' define systems that count as arithmetics. It's also fairly clear which settings define relevant systems. Thus, by fiddling with the controls we lay out, we're hopeful that the community of relevance logicians will be able to find at least a few nice-enough members of $\{R^{\#+\epsilon} : 0 < \epsilon < 1\}$, and that having such systems in hand will jumpstart the search for $R^{\#\frac{1}{2}}$. Unfortunately, due to constraints of time and space, this paper underdelivers on this front, as the one system we system we had time to in fact investigate ends up being weaker than we'd hoped. Be that as it may, the general idea remains on the table, and those looking for plausible candidate relevant arithmetics ought to be able to use our machinery to at least get a nose for some interesting candidates.

The organization of the paper is as follows: the first three sections all deal with the *terms* in our language. In Section 4, we turn from terms to the full language and define our logic. Section 5 gives a (perhaps surprisingly polemical) overview of the construction of our semantics. Section 6 lays the foundation for our construction of the canonical model, which is completed in Section 7. In Section 8 we give the full semantics and prove the standard metatheoretic results. In Section 9 we determine how much arithmetic we've captured and lay out directions for further research.

1 Linguistic Matters 1: Elementary Terms

We will have a stratified vocabulary. At the bottom level is what we call the *elementary* vocabulary—where here we emphasize the '*element*' in *elementary*. The point to be made is that the expressions in the elementary vocabulary are elements (in the set theoretic sense) of the equivalence classes that make up the language we define our logic on. That said, the elementary vocabulary consists of the following:

1. The unary function symbols $'$ and z, and for each $i > 0$ the symbol pr_i;
2. The constant symbol 0; and the variables $x_1, x_2, \ldots$;
3. The separator symbol '$|$';
4. The binary relation symbol '$=$';
5. The connectives t, $\sim$, $\wedge$, and $\rightarrow$;
6. The quantifier $\forall$; and
7. Set braces (which we will typically drop) and parentheses.

Some of the expressions in this vocabulary will be regarded as elementary terms; others as elementary function symbols. We will use these to construct elementary atomic formulas and, from there, arbitrary elementary formulas. Terms, function symbols, and formulas in general are then constructed as equivalence classes of elementary terms, elementary function symbols, and the like. But for the first three sections of the paper, we will be focused only on *terms*, since that's where all the arithmetic ends up happening.

A few terminological (no pun intended) notes are in order: If e is an (elementary) function symbol then we say e is n-adic if, intuitively, its inputs are n-tuples. All (elementary) function symbols in the system we give will have *arity* one in the sense that their outputs are one-tuples. However, for reasons of technical convenience, we allow *terms* to have any finite arity, with higher-arity terms generated from lower-arity terms by what is essentially just juxtaposition. Thus, what would in many contexts be thought of as a length-n sequence of terms will, for us, be thought of as a single n-ary term.

Now to the details. We define the set of elementary function symbols and the set of elementary terms simultaneously as follows:

- $'$ and z are monadic elementary function symbols.
- $0 \mid \varnothing$ is a unary elementary term.
- If x is a variable, then $x \mid \{x\}$ is a unary elementary term.
- If $\sigma \mid X$ is an n-ary elementary term and $\tau \mid Y$ is an m-ary elementary term, then $\sigma\tau \mid X \cup Y$ is an $n+m$-ary elementary term.

- If f is an n-adic elementary function symbol and $\tau \mid X$ is an n-ary elementary term, then $f(\tau) \mid X$ is a unary elementary term.
- If $\tau \mid X$ is an n-ary elementary term and y is a variable, then $\tau \mid X \cup \{y\}$ is an n-ary elementary term
- If $\sigma \mid \{x_1, \ldots, x_m\}$ and $\tau \mid \{y_1, \ldots, y_k\}$ are k-ary elementary terms, then for each $1 \leq i \leq k$, $\mathsf{pr}_i[\sigma \mid X; \tau \mid Y]$ is an $m+1$-adic elementary function symbol.

We will discuss how to interpret z-, $'$-, and pr-expressions in §2. Before turning to that, we have to discuss a few other matters.

First, if $\tau \mid X$ is an elementary term, then we say that τ is its *untyped* part and X is its *typing* part. In a similar vein, if τ is a sequence of symbols and $\tau \mid X$ is an n-ary elementary term for some set of variables X, then τ is called an *untyped n-ary term.*

It's also worth explicitly noting that while $'$ and z are monadic elementary function symbols, pr_i is not an elementary function symbol at all. If we have to give a name to the type of thing pr_i is, I suppose it's best to call it an elementary function-symbol-building-symbol. That is, it is a symbol that can be used to build elementary function symbols. More to the point, when given a pair of elementary terms of the appropriate sort, pr_i returns an elementary function symbol whose adicity is determined by the cardinality of the typing part of the first elementary term we supply it.

As a convenient abbreviation, if $\sigma \mid \{x_1, \ldots, x_m\}$ and $\tau \mid \{y_1, \ldots, y_k\}$ are k-ary elementary terms, and η is an untyped $m+1$-ary elementary term, then

$$\mathsf{pr}[\sigma \mid X; \tau \mid Y](\eta) := \mathsf{pr}_1[\sigma \mid X; \tau \mid Y](\eta) \ldots \mathsf{pr}_k[\sigma \mid X; \tau \mid Y](\eta).$$

In the other direction, if σ is an untyped n-ary elementary term, and $1 \leq i \leq n$, then by σ_i we will understand the ith component of σ.

It will be necessary to talk quite a bit about substitutions. For this purpose, we introduce the following notation:

Definition 1.1.

1. if τ, σ, and η are constants or variables, then $\tau(\sigma/\eta)$ is η if $\tau = \sigma$, and otherwise τ. Read this as "replace σ with η"

2. If σ and η are untyped n-ary terms— say $\sigma_1 \ldots \sigma_n$ and $\eta_1 \ldots \eta_n$, respectively— then (σ/η) abbreviates a simultaneous replacement of each σ_i with η_i.

3. (Now getting ahead of ourselves a bit) $\phi(\sigma/\eta)$ is the result of replacing each free occurrence of an untyped term τ in ϕ by an occurrence of $\tau(\sigma/\eta)$.

4. If X is a set of variables, then $\overline{X}$ abbreviates the sequence of those variables taken in increasing order (by their subscript indices).

5. If X and Y are sets of variables with $\mathrm{card}(X) = \mathrm{card}(Y)$, then we call a replacement of the form $\phi(\overline{Y}/\overline{X})$ a change of variables.

Note that as we use the term 'change of variables', a change of variables is always *monotonic*: the lowest-subscripted variable in Y replaces the lowest-subscripted variable in X, the second-lowest-subscripted variable in Y replaces the second-lowest-subscripted variable in X, etc.

2 Linguistic Matters 2: Terms

We call the following family of three transformations the *elementary rewrite rules*. They specify that certain strings of symbols may replace certain other strings of symbols no matter where the strings occur within an expression.

(i) $\mathsf{z}(\tau) \rightharpoonup 0$;

(ii) $\mathsf{pr}_i[\sigma \mid X; \tau \mid Y](\eta 0) \rightharpoonup \sigma_i \left(\overline{X}/\eta\right)$; and

(iii) $\mathsf{pr}_i[\sigma \mid X; \tau \mid Y](\eta\rho') \rightharpoonup \tau_i \left(\overline{Y}/\, \mathsf{pr}[\sigma \mid X; \tau \mid Y](\eta\rho)\right)$.

We write $\overset{*}{\rightharpoonup}$ for the reflexive-transitive closure of $\rightharpoonup$, and $\leftrightharpoons$ for the smallest equivalence relation containing $\rightharpoonup$.

Before we discuss how to *use* the rewrite rules, we should comment on the rules themselves. The first, we hope you'll agree, ensures that z behaves like the constant zero function. The second and third ensure that pr behaves somewhat like *primitive recursion*. Roughly, (ii) 'says' that to 'evaluate' $\mathsf{pr}_i[\sigma \mid X; \tau \mid Y]$ at the 'argument' $(\eta 0)$, what we do is evaluate $\sigma_i \mid X$ at the argument η. Similarly, (iii) says that to evaluate $\mathsf{pr}_i[\sigma \mid X; \tau \mid Y]$ at the argument $\eta\rho'$ what we do is evaluate $\tau_i \mid Y$ at the prior case; that is, at $\mathsf{pr}[\sigma \mid X; \tau \mid Y](\eta\rho)$.

In a little more detail: Recall how primitive recursion usually works. Given an n-adic function f and an $n+2$-adic function g, we define the function $pr(f,g)$ by the two equations

$$pr(f,g)(x_1,\ldots,x_n,0) = f(x_1,\ldots x_n)$$
$$pr(f,g)(x_1,\ldots,x_n,y') = g(pr(f,g)(x_1,\ldots,x_n,y),x_1,\ldots,x_n,y)$$

At a glance, it might seem that our scheme is less expressive than primitive recursion, since if we use 1-ary terms f,g in setting up the recursion, then g is forced to be 1-adic. So, letting $X = x_1\ldots x_n$ and continuing the analogy, we get

$$\mathsf{pr}_1[f \mid X; g \mid x](x_1\ldots x_n 0) \rightharpoonup f(x_1,\ldots,x_n)$$
$$\mathsf{pr}_1[f \mid X; g \mid x](x_1\ldots x_n y') \rightharpoonup g(\mathsf{pr}_1[f \mid X; g \mid x](x_1\ldots x_n y))$$

Which is mere iteration of g, not proper recursion.

However, by being a little more creative, we can recover the expressive power of full primitive recursion, in spite of the apparent simplicity of our rewrite rules. Suppose X is $x_1\ldots x_n$ and Y is $x_1\ldots x_{n+2}$. Let $\tilde{f} \mid X$ be the $n+2$-ary, n-adic term

$$f(x_1,\ldots,x_n)x_1x_2x_3\ldots x_n0 \mid X$$

and let $\tilde{g} \mid Y$ be the $n+2$-ary, $n+2$-adic term

$$g(x_1,\ldots,x_{n+2})x_2x_3\ldots x_nx_{n+1}x'_{n+2} \mid Y$$

We can then consider the $n+2$ distinct $n+1$-adic function symbols $\mathsf{pr}_i(\tilde{f} \mid X; \tilde{g} \mid X)$.

If $1 < i < n+2$, we get:

$$\mathsf{pr}_i(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1\ldots x_n 0) \rightharpoonup \tilde{f}_i\left(\overline{X}/\overline{X}\right) = x_{i-1}\left(\overline{X}/\overline{X}\right) = x_{i-1}$$
$$\begin{aligned}\mathsf{pr}_i(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1\ldots x_n y') &\rightharpoonup \tilde{g}_i\left(\overline{Y}/\,\mathsf{pr}(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1\ldots x_n y)\right)\\ &= x_i\left(\overline{Y}/\,\mathsf{pr}(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1\ldots x_n y)\right)\\ &= \mathsf{pr}_i(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1\ldots x_n y)\end{aligned}$$

so evidently, in this case, $\mathsf{pr}_i(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1 \dots x_n y) \overset{*}{\rightharpoonup} x_{i-1}$. Similarly,

$$\begin{aligned}\mathsf{pr}_{n+2}(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1 \dots x_n 0) &\rightharpoonup 0\\ \mathsf{pr}_{n+2}(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1 \dots x_n y') &\rightharpoonup x'_{n+2}\left(\overline{Y}/\,\mathsf{pr}(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1 \dots x_n y)\right)\\ &= \mathsf{pr}_{n+2}(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1 \dots x_n y)'\end{aligned}$$

so evidently if τ is a numeral, then $\mathsf{pr}_{n+2}(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1 \dots x_n \tau) \overset{*}{\rightharpoonup} \tau$. Thus,

$$\begin{aligned}\mathsf{pr}_1(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1 \dots x_n 0) &\rightharpoonup \tilde{f}_1\left(\overline{X}/\overline{X}\right) = f(x_1, \dots, x_n)\\ \mathsf{pr}_1(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1 \dots x_n \tau') &\rightharpoonup g\left(\overline{Y}/\,\mathsf{pr}(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1 \dots x_n \tau)\right)\\ &= g(\mathsf{pr}_1(\tilde{f} \mid X; \tilde{g} \mid Y)(x_1 \dots x_n y), x_1, \dots, x_n, \tau).\end{aligned}$$

Of course, comparing all this to ordinary primitive recursion, one glaring oddity still stands out: primitive recursion is something one does with *functions*, but we've applied pr_i to *terms*. Our rewrite rules are ultimately devices for defining a relation on a set of syntactic entities—our elementary terms—and only indirectly a method of defining functions. So, while primitive recursive functions are one analogue of our pr terms, a better analogy might be to something like programs written in a very restrictive programing language, with the reduction steps analogous to the steps of execution of the program.[2] The programming language, in fact, is sufficiently restrictive that every program it can construct can be guaranteed to terminate.

Lemma 2.1.

1. *The rewrite rules are* locally confluent*: if $t \rightharpoonup a$ and $t \rightharpoonup b$ then there's some c such that $a \overset{*}{\rightharpoonup} c$ and $b \overset{*}{\rightharpoonup} c$*

2. *The rewrite rules are* strongly normalizing*: the rewrite relation is well-founded.*

[2] How to make sense of free variables on this analogy? One might imagine a crudely optimizing compiler performing a bit of computation while compiling the program, so that the program is faster at runtime. Given a program containing a term like $\mathrm{Pr}(1 \mid \varnothing; x' \mid x)(3)$ the compiler would optimize by replacing this term with $1''' = 4$ once and for all, rather than performing these reduction steps every time the program is executed. Given a term containing a free variable, the optimization must stop since the value of that variable would presumably be supplied at runtime.

Proof. (1) is clear upon inspection. The rest of the proof requires a little terminology. Say that a *reduction sequence* is a sequence of one or more elementary terms, with a first item, such that each item of the sequence is either first or is the result of applying a rewrite rule to the previous item. So, $a \xrightharpoonup{*} b$ if there's a reduction sequence beginning in a and ending in b. A term is *strongly normalizing* if it is not the first item of any infinite reduction sequence, and the rules are strongly normalizing if every term is. Say that a term is *irreducible* if it is not the first item of a reduction sequence of length greater than one.

We can then reduce the proof of (ii) to something slightly simpler. First of all, we can assume without loss of generality that we are dealing with a unary term t. Second, we can safely assume that arguments of t (its maximal proper subterms) are strongly normalizing. Since variables and constants are trivially strongly normalizing, knowing that every term with strongly normalizing arguments is strongly normalizing would give us (ii) by a simple induction on term complexity.

So, we need only to show that every term whose arguments are strongly normalizing is itself strongly normalizing. We note that because of (1), any strongly normalizing term t in fact has a *unique* irreducible c such that $t \xrightharpoonup{*} c$ (by Newman's lemma). We call this c the normal form of t.

We define the H-complexity of a unary elementary term to be the number of pr symbols occurring in its main functional expression, so that e.g. the H-complexity of $0''$ is zero and the H-complexity of the doubling function, $\mathsf{pr}[0 \mid \varnothing; \mathsf{pr}[0 \mid \varnothing; x'' \mid x](x) \mid x]$ applied to any argument, is two. We define the T-complexity of a unary elementary term be the total number of occurrences of successor that are contained in the normal forms of its arguments. We let the overall complexity of a term be measured by a pair (h, t) consisting of its H and T-complexity, where overall complexities are ordered lexically, with h lexically prior.

We argue by induction on overall complexity. For our basis case, we note that terms of complexity (0,0) must be either 0 or a variable, both of which are irreducible and a fortiori strongly normalizing.

So, for the induction step, suppose (call this assumption IH1) that all the arguments to a term t are strongly normalizing, and that any term of lesser overall complexity whose arguments are strongly normalizing is itself strongly normalizing.

Suppose towards a reductio that there is an infinite reduction sequence beginning with t. Since the induced reduction sequences on the arguments of t must be finite, at some point, we will need to stop applying rewrite rules to subterms, and apply a rewrite rule to a complete term $t' \overset{*}{\leftharpoondown} t$. But because we've only applied rewrites to subterms so far, t' has the same main functional expression as t (hence the same H-complexity) and its arguments have the same normal forms as the arguments to t (hence t' has the same T-complexity as t). So t' has the same overall complexity as t. We now argue by cases

If the rewrite $t \overset{*}{\rightharpoondown} t'$ is

$$\mathsf{z}(\tau) \rightharpoondown 0$$

the reduction terminates, for a contradiction. If the rewrite is

$$\mathsf{pr}_i[\sigma \mid X; \tau \mid Y](\eta 0) \rightharpoondown \sigma_i\left(\overline{X}/\eta\right)$$

then we argue by structural induction that all subterms of $\sigma_i\left(\overline{X}/\eta\right)$ (including the entire term) are strongly normalizing. For our basis case, we know that the terms η replacing $\overline{X}$ are all strongly normalizing. For the induction step, consider a (non-variable) subterm σ' of $\sigma_i\left(\overline{X}/\eta\right)$, whose arguments are all strongly normalizing. We know that σ' is $\sigma''\left(\overline{X}/\eta\right)$ for some subterm σ'' of σ_i. But σ'' must then have H-complexity lower than t. So, the same is true of σ'. And so by the prior inductive hypothesis (IH1), σ' strongly normalizes. So every subterm of $\sigma_i\left(\overline{X}/\eta\right)$ is strongly normalizing. But this contradicts our assumption that $t \overset{*}{\rightharpoondown} t'$ is part of an infinite reduction sequence.

If the rewrite is

$$\mathsf{pr}_i[\sigma \mid X; \tau \mid Y](\eta\rho') \rightharpoondown \tau_i\left(\overline{Y}/\,\mathsf{pr}[\sigma \mid X; \tau \mid Y](\eta\rho)\right)$$

Then the terms $\mathsf{pr}[\sigma \mid X; \tau \mid Y](\eta\rho)$ on the right hand side are strongly normalizing: each one has the same H-complexity as t, but lower T-complexity, and has strongly normalizing arguments. Since every subterm of τ_i must have a lower H-complexity than t, we can then repeat the structural induction of the previous paragraph to establish that $\tau_i\left(\overline{Y}/\,\mathsf{pr}[\sigma \mid X; \tau \mid Y](\eta\rho)\right)$ strongly normalizes, for a contradiction, ending the reductio. □

We say that $\sigma \mid X$ and $\tau \mid Y$ are *directly equivalent* when either (a) $X = Y$ and $\sigma \leftrightharpoons \tau$, or (b) $\sigma \mid X = \tau(\overline{Y}/\overline{X}) \mid X$. We take *equivalence* simpliciter to be the transitive closure of direct equivalence, and we write $\langle \tau \mid X \rangle$ for the equivalence class of $\tau \mid X$. We will call $\langle \tau \mid X \rangle$ a *term*—recall that the members of $\langle \tau \mid X \rangle$ are, in contrast, called *elementary terms*.

Summary and Reminder

Before moving on, let's remember what we're up to, since it would be easy to have lost the thread by now. Our plan was to build a system in which p.r. equivalent terms are in fact identical. What we've done so far is build a family of elementary terms and an equivalence relation that identifies elementary terms that are p.r. equivalent. The goal, eventually, is to build an entire language in which we make similar identifications on syntactic objects of all types. Before turning to that, however, we first pause to study the *algebraic* structure of the class of terms.

3 Linguistic Matters 3: A Category of Types and Terms

Suppose $\tau \mid y_1, \ldots, y_{n_2}$ is an n_1-ary elementary term. Given any n_2-ary untyped term $\sigma_1 \ldots \sigma_{n_2}$, we can simultaneously substitute the σ_i's for the y_i's to construct an untyped elementary term $\tau(y_1/\sigma_1, \ldots, y_{n_2}/\sigma_{n_2})$. In general, of course, $\tau(y_1/\sigma_1, \ldots, y_{n_2}/\sigma_{n_2}) \mid y_1, \ldots, y_{n_2}$ won't be well-formed because the σ_i's may contain variables other than the y_j's. But this is easily rectified. Rather than beginning with a heathenous *untyped* elementary term we should simply help ourselves to honest-to-god and *properly typed* elementary term—say $\sigma \mid X$. It's then clear enough how to end up somewhere well-formed—we just let the typing part of the term being 'substituted in' come along for the ride. That is, while in general $\tau(y_1/\sigma_1, \ldots, y_{n_2}/\sigma_{n_2}) \mid y_1, \ldots, y_{n_2}$ will *not* be well-formed, the fact that $\sigma \mid X$ is well-formed guarantees that $\tau(y_1/\sigma_1, \ldots, y_{n_2}/\sigma_{n_2}) \mid X$ always *will* be. Thus, there is a natural way to define composition of elementary terms that we specify as follows:

Definition 3.1. If $\text{card}(X) = m$ and $\sigma \mid Y$ is m-ary, we define

$$\tau \mid X \circ_e \sigma \mid Y = \tau\left(\overline{X}/\sigma\right) \mid Y$$

In order to import this from elementary terms to terms, we need the following fact:

Fact 3.2. *If $\tau \mid Y$ and $\tau' \mid Y'$ are representatives of the same term, and $\sigma \mid X$ and $\sigma' \mid X'$ are also representatives of the same term then $\sigma \mid X \circ_e \tau \mid Y$ and $\sigma \mid X \circ_e \tau' \mid Y'$ and $\sigma' \mid X' \circ_e \tau \mid Y$ are all, if defined, also equivalent.*

Definition 3.3. If $\mathrm{card}(X) = m$ and $\sigma \mid Y$ is m-ary, we define

$$\langle \tau \mid X \rangle \circ \langle \sigma \mid Y \rangle = \langle \tau \left(\overline{X}/\sigma \right) \mid Y \rangle$$

By inspection, one easily verifies that terms of the form $\left\langle \overline{X} \mid X \right\rangle$ act as two-sided identities for $\circ$. It's also trivial but tedious to check that $\circ$ is associative. Altogether, then, the terms can be viewed as arrows in a category where $\circ$ is the composition.

A final note: the elementary terms contained in the class of terms $\langle \overline{X} \mid X \rangle$ all have the form $\overline{Y} \mid Y$ where $\mathrm{card}(Y) = \mathrm{card}(X)$. Since the identity arrows and the objects of a category are basically the same thing, it follows that the objects in this category can be identified with the cardinalities of the sets X.

We call the category thus defined $\mathcal{B}$. We call the objects of $\mathcal{B}$ types and let T_n be the type corresponding to the identity arrow $\langle \overline{X_1^n} \mid X_1^n \rangle$. By inspection, we see that the term ($\mathcal{B}$-arrow) $\langle \tau \mid X \rangle$ has as its domain the type $T_{\mathrm{card}(X)}$ and as its codomain the type $T_{\mathrm{len}(\tau)}$.

3.1 Examples

Before we move on, we give three examples that will hopefully serve to demonstrate the usefulness of the machinery developed here.

Example 3.4. We will write $\underline{n}$ for the numeral $0^{\prime\cdots\prime}$ with n primes. We claim that $\langle \mathsf{pr}[x_1 \mid x_1; x_2' \mid x_2](\underline{n}\ \underline{m}) \mid \varnothing \rangle = \langle \underline{n+m} \mid \varnothing \rangle$. Thus $\mathsf{pr}[x_1 \mid x_1; x_2' \mid x_2]$ represents (in a sense we won't try to get clearer on) the addition function. Note one important consequence of this is that addition of *numerals* is both commutative and associative. Also note that this is an equality at the level of *syntax*, and explicitly tells us nothing about whether formulas of the form $\mathsf{pr}[x_1 \mid x_1; x_2' \mid x_2](\underline{n}\ \underline{m}) = \underline{n+m} \mid \varnothing$ are theorems—indeed, we have yet to even *define* the class of formulas or the notion of theoremhood![3]

At any rate, we prove our claim by induction in the metalanguage. For the base case we rewrite as follows:

$$\begin{aligned} \mathsf{pr}[x_1 \mid x_1; x_2' \mid x_2](\underline{n}\ 0) \mid \varnothing &\rightharpoonup x_1(x_1/\underline{n}) \mid \varnothing \\ &\rightharpoonup \underline{n} \mid \varnothing \\ &\rightharpoonup \underline{n+0} \mid \varnothing. \end{aligned}$$

[3]Of course, as it turns out, all such formulas *are* theorems. In fact, all such formulas are axioms, since by our syntactic rules, all such formulas are in fact of the form $\tau = \tau \mid \varnothing$.

Which establishes that $\langle \mathsf{pr}[x_1 \mid x_1; x_2' \mid x_2](\underline{n}\ 0) \mid \varnothing\rangle = \langle \underline{n+0} \mid \varnothing\rangle$.

For the successor case, we rewrite as follows:

$$\begin{aligned}\mathsf{pr}[x_1 \mid x_1; x_2' \mid x_2](\underline{n}\ \underline{m}') \mid \varnothing &\rightharpoonup x_2'(x_2/\ \mathsf{pr}[x_1 \mid x_1; x_2' \mid x_2](\underline{n}\ \underline{m})) \mid \varnothing\\ &\rightharpoonup \mathsf{pr}[x_1 \mid x_1; x_2' \mid x_2](\underline{n}\ \underline{m})' \mid \varnothing\\ &\rightharpoonup \underline{n+m'} \mid \varnothing\\ &\rightharpoonup \underline{n+m+1} \mid \varnothing\end{aligned}$$

Which demonstrates that $\langle \mathsf{pr}[x_1 \mid x_1; x_2' \mid x_2](\underline{n}\ \underline{m}') \mid \varnothing\rangle = \langle \underline{n+m+1} \mid \varnothing\rangle$.

In light of these results, we will write $\mathsf{plus}(x, y)$ for $\mathsf{pr}[x_1 \mid x_1; x_2' \mid x_2](xy)$, in which case we can abbreviate the result above as $\langle \mathsf{plus}(\underline{n}, \underline{m}) \mid \varnothing\rangle = \langle \underline{n+m} \mid \varnothing\rangle$.

Example 3.5. In the same way that $\mathsf{pr}[x_1 \mid x_1; x_2' \mid x_2]$ represents the addition function, we claim that $\mathsf{pr}_1[0x_10 \mid x_1; \mathsf{plus}(x_1, x_2)x_2x_3' \mid x_1, x_2, x_3](\underline{n}\ \underline{m}) \mid \varnothing$ represents the multiplication function. It's useful to work through this example in order to better understand our demonstration from the previous section that the pr-machinery in fact captures full primitive recursion.

To that end, we first examine the pr_2 and pr_3 parts. Here we sometimes perform several reductions at once, in order to save space. Also, for readability, we will simply write $\mathsf{pr}_i(\underline{n}\ \underline{m})$ instead of $\mathsf{pr}_i[0x_10 \mid x_1; \mathsf{plus}(x_1, x_2)x_2x_3' \mid x_1, x_2, x_3](\underline{n}\ \underline{m})$.

For pr_2, things are fairly easy. For the zero case we have $\mathsf{pr}_2(\underline{n}0) \rightharpoonup x_1(x_1/\underline{n}) = \underline{n}$; for the successor case we have that $\mathsf{pr}_2(\underline{n}\ \underline{m}') \rightharpoonup \mathsf{pr}_2(\underline{n}\ \underline{m})$. Thus, inductively it's clear that $\mathsf{pr}_2(\underline{n}\ \underline{m}) \overset{*}{\rightharpoonup} \underline{n}$.

pr_3 turns out not to matter, but we go through it anyways, since it's good for the soul to do so. For the zero case we have $\mathsf{pr}_3(\underline{n}0) \rightharpoonup 0$; for the successor case we get $\mathsf{pr}_3(\underline{n}\ \underline{m}') \rightharpoonup \mathsf{pr}_3(\underline{n}\ \underline{m})' \overset{*}{\rightharpoonup} \underline{m}'$. Thus in general $\mathsf{pr}_3(\underline{n}\ \underline{m}) \overset{*}{\rightharpoonup} \underline{m}$.

Now we turn to pr_1. The zero case is immediate: $\mathsf{pr}_1(\underline{n}0) \rightharpoonup 0 = \underline{n \times 0}$. For the successor case, we reduce as follows:

$$\begin{aligned}\mathsf{pr}_1(\underline{n}\ \underline{m}') &\rightharpoonup \mathsf{plus}(x_1, x_2)(x_1x_2x_3/\ \mathsf{pr}_1(\underline{n}\ \underline{m})\ \mathsf{pr}_2(\underline{n}\ \underline{m})\ \mathsf{pr}_3(\underline{n}\ \underline{m}))\\ &= \mathsf{plus}(\mathsf{pr}_1(\underline{n}\ \underline{m}), \underline{n})\\ &\overset{*}{\rightharpoonup} \mathsf{plus}(\underline{n \times m}, \underline{n})\\ &\overset{*}{\rightharpoonup} \underline{(n \times m) + n}\end{aligned}$$

Here the step from the second line to the third follows by the inductive hypothesis and the step from the third line to the fourth follows from the previous example. It follows that, as we claimed, $\langle \mathsf{pr}_1[0x_10 \mid x_1; \mathsf{plus}(x_1, x_2)x_2x_3' \mid$

$x_1, x_2, x_3](\underline{n}\ \underline{m}) \mid \varnothing\rangle$ represents the multiplication function in the sense that

$$\langle \mathsf{pr}_1[0x_10 \mid x_1; \mathsf{plus}(x_1, x_2)x_2x_3' \mid x_1, x_2, x_3](\underline{n}\ \underline{m}) \mid \varnothing\rangle = \langle \underline{n \times m} \mid \varnothing\rangle$$

Note again that it follows from this that numeral multiplication has its expected properties.

Example 3.6. Another useful detail: suppose $\langle \sigma' \mid X\rangle$ and $\langle \tau' \mid Y\rangle$ are the same term. It follows that $\langle \sigma \mid X\rangle$ and $\langle \tau \mid Y\rangle$ are the same term.

This is easy to see: suppose that $\langle \sigma \mid X\rangle$ and $\langle \tau \mid Y\rangle$ are different terms. Without loss of generality we can take both σ and τ to be irreducible. Notice that any reduction that applies to σ' applies to σ and any reduction that applies to τ' similarly applies to τ. Thus σ' and τ' are also irreducible. And thus since $\langle \sigma \mid X\rangle$ and $\langle \tau \mid Y\rangle$ are different, $\langle \sigma' \mid X\rangle$ and $\langle \tau' \mid Y\rangle$ are also different.

This example has a motivationally important philosophical corollary: it says that the system we've given here has the feature we took Meyer to want. More to the point, it tells us that we do not have to *say* that things with identical successors are themselves identical. Or at least, we don't need to say that things whose successors are identical at the level of syntax are identical since it follows from σ and τ having successors that are identical at the level of syntax that σ and τ are themselves identical at the level of syntax.

3.2 Important Definitions and Results

Natural numbers objects, introduced by Lawvere in his seminal [15] are designed to be a 'completely categorial' analogue of the natural numbers that can be defined inside any topos. It wasn't too long after they were introduced before people noticed that with mild modifications, the same essential notion could be defined in an arbitrary category with finite products. These mildly-modified objects came to be called parameterized natural numbers objects (PNNOs).[4]

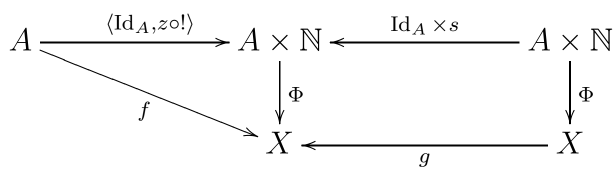
The fact that an object that (in a way that can be made concrete) 'looks like' the natural numbers can be defined in an arbitrary category with finite products suggests that such categories are a natural playground in which to find interesting models of arithmetic. When we combine this insight with the fact that hyperdoctrines—the core piece of categorial-logic machinery

[4]See e.g. [14] for a discussion.

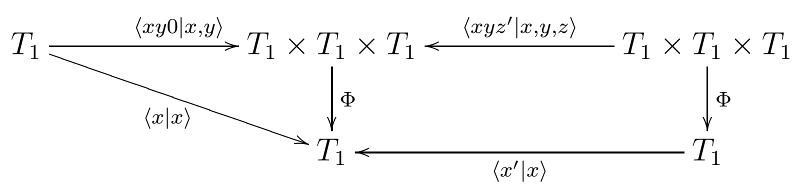
deployed in this paper—provide a natural playground in which to find interesting models of relevant logics, it becomes clear that we ought to be able to build *semantically* interesting relevant arithmetics by combining the PNNO-ish and hyperdoctrinal constructions. The main thrust of this paper is to do just that.

For the moment, we focus on the PNNO aspect. To that end, we adopt the following definition from [14]:

Definition 3.7. In a category with finite products, A parameterized natural numbers object is an object $\mathbb{N}$ together with maps $z : \mathbf{1} \to \mathbb{N}, s : \mathbb{N} \to \mathbb{N}$ such that given any objects A, X and maps $f : A \to X, g : X \to X$, there is a *unique* map $\Phi : A \times \mathbb{N} \to X$ making the following diagram commute:



We would love to discover that $\mathcal{B}$ is harboring a PNNO. If it were, the only natural candidate for the $\mathbb{N}$-role would be T_1, and for s and z, $\langle x' \mid x \rangle$ and $\langle 0 \mid \varnothing \rangle$ respectively. Sadly, things don't quite pan out. The difficulty is with the uniqueness clause. Consider the following diagram:



As it turns out, there is more than one way to solve for Φ here. The standard recursive definition of addition in our setting works out as $\mathsf{pr}[x \mid x, y; x' \mid x](xy)$. But consider how we would write $x + (y + z)$ and $(x + y) + z$, respectively:

$$\mathsf{pr}[v \mid v, w; v' \mid v](x\, \mathsf{pr}[v \mid v, w; v' \mid v](yz))$$
$$\mathsf{pr}[v \mid v, w; v' \mid v](\mathsf{pr}[v \mid v, w; v' \mid v](xy)z)$$

These are both in normal form; therefore neither rewrites to the other. But if we precompose these with the untyped term $xy0$, then in each case we *can* rewrite:

$$\mathsf{pr}[v \mid v, w; v' \mid v](x\,\mathsf{pr}[v \mid v, w; v' \mid v](y0)) \rightharpoonup \mathsf{pr}[v \mid v, w; v' \mid v](xy)$$

$$\mathsf{pr}[v \mid v, w; v' \mid v](\mathsf{pr}[v \mid v, w; v' \mid v](xy)0) \rightharpoonup \mathsf{pr}[v \mid v, w; v' \mid v](xy)$$

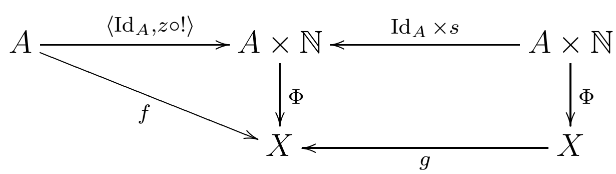
The first by rewriting the inner term, and the second by rewriting the outer. Similarly

$$\begin{aligned}\mathsf{pr}[v \mid v, w; v' \mid v](x\,\mathsf{pr}[v \mid v, w; v' \mid v](yz')) &\rightharpoonup \mathsf{pr}[v \mid v, w; v' \mid v](x\,\mathsf{pr}[v \mid v, w; v' \mid v](yz)') \\ &\rightharpoonup \mathsf{pr}[v \mid v, w; v' \mid v](x\,\mathsf{pr}[v \mid v, w; v' \mid v](yz))'\end{aligned}$$

$$\mathsf{pr}[v \mid v, w; v' \mid v](\mathsf{pr}[v \mid v, w; v' \mid v](xy)z') \rightharpoonup \mathsf{pr}[v \mid v, w; v' \mid v](\mathsf{pr}[v \mid v, w; v' \mid v](xy)z)'$$

So, the two distinct ways of associating a double addition each make the diagram commute.[5] It follows that there's no PNNO to be found in $\mathcal{B}$. What $\mathcal{B}$ *does* have, though, is a *weak* PNNO, which (following [22]) we define as follows:

Definition 3.8. In a category with finite products, a weak parameterized natural number object is an object $\mathbb{N}$ together with maps $z : \mathbf{1} \to \mathbb{N}, s : \mathbb{N} \to \mathbb{N}$ such that given any objects A, X and maps $f : A \to X, g : X \to X$, there is a map $\Phi : A \times \mathbb{N} \to X$ making the following diagram commute:



[5] There's a more theoretically satisfying explanation here, which we relegate to this footnote, as available space and time preclude a rigorous presentation. As it stands, equality of terms is, for us, decidable, since we can always compare normal forms. If our rewrite system *did* support interpreting T_1 as a PNNO, then the uniqueness clause would entail that our notion of term equality admitted an inference analogous to Goodstein's induction rule for equational primitive recursive arithmetic:

$$\frac{F(0) = G(0) \qquad H(x, F(x)) = F(S(x)) \qquad H(x, G(x)) = G(S(x))}{F(x) = G(x)}$$

and this would render the obvious theory of term equality over our terms (essentially) undecidable, contradicting the fact that equality of our terms is in fact decidable. See [9, p105] for background on Goodstein's rule.

Theorem 3.9. *The category $\mathcal{B}$ has finite products, a terminal object T_0, and a weak PNNO given by T_1 together with the arrows $T_0 \xrightarrow{\langle 0|\varnothing\rangle} T_1$ and $T_1 \xrightarrow{\langle z'|z\rangle} T_1$.*

Before we turn to proving this we state without proof the following easy-to-verify but extremely useful fact:

Fact 3.10. *Each term has a representative of the form $\tau \mid X_1^n$ for some n.*

Proof of Theorem 1. The first two parts are straightforward and left to the reader. To see that the given data define a parameterized natural numbers object, suppose we have the following:

$$T_m \xrightarrow[\langle\sigma|X\rangle]{} T_k \xleftarrow[\langle\tau|Y\rangle]{} T_k$$

Also suppose without loss of generality (since we can always change representative if needed) that $z \notin X \cup Y$ and $\overline{X,z} = \overline{X}z$. Our task is to show that there is a term $\langle \rho \mid X, z\rangle$ such that the following commutes:

$$\begin{array}{ccccc}
T_m & \xrightarrow{\left\langle \overline{X}0|X\right\rangle} & T_m \times T_1 & \xleftarrow{\left\langle \overline{X}z'|X,z\right\rangle} & T_m \times T_1 \\
 & \searrow^{\langle\sigma|X\rangle} & \downarrow{\scriptstyle\langle\rho|X,z\rangle} & & \downarrow{\scriptstyle\langle\rho|X,z\rangle} \\
 & & T_k & \xleftarrow[\langle\tau|Y\rangle]{} & T_k
\end{array}$$

I claim $\left\langle \mathsf{pr}[\sigma \mid X; \tau \mid Y](\overline{X,z}) \mid X, z\right\rangle$ will do the job.

To verify this, we need to prove the following two identities:

$$\left\langle \mathsf{pr}[\sigma \mid X; \tau \mid Y](\overline{X,z}) \mid X, z\right\rangle \circ \left\langle \overline{X}0 \mid X\right\rangle = \langle \sigma \mid X\rangle \tag{A}$$

$$\left\langle \mathsf{pr}[\sigma \mid X; \tau \mid Y](\overline{X,z}) \mid X, z\right\rangle \circ \left\langle \overline{X}z' \mid X, z\right\rangle = \langle \tau \mid Y\rangle \circ \left\langle \mathsf{pr}[\sigma \mid X; \tau \mid Y](\overline{X,z}) \mid X, z\right\rangle \tag{B}$$

For (A), observe that

$$\begin{aligned}
\left\langle \mathsf{pr}[\sigma \mid X; \tau \mid Y](\overline{X,z}) \mid X, z\right\rangle \circ \left\langle \overline{X}0 \mid X\right\rangle &= \left\langle \mathsf{pr}[\sigma \mid X; \tau \mid Y](\overline{X}0) \mid X\right\rangle \\
&= \left\langle \sigma(\overline{X}/\overline{X}) \mid X\right\rangle \\
&= \langle \sigma \mid X\rangle
\end{aligned}$$

For the second, observe that

$$\begin{aligned}\left\langle \mathsf{pr}[\sigma \mid X;\tau \mid Y](\overline{X,z}) \mid X,z\right\rangle \circ \left\langle \overline{X}z' \mid X,z\right\rangle \\ &= \left\langle \mathsf{pr}[\sigma \mid X;\tau \mid Y](\overline{X}z') \mid X,z\right\rangle \\ &= \left\langle \tau\left(\overline{Y}/\,\mathsf{pr}[\sigma \mid X;\tau \mid Y](\overline{X}z)\right) \mid X,z\right\rangle \\ &= \langle \tau \mid Y\rangle \circ \left\langle \mathsf{pr}[\sigma \mid X;\tau \mid Y](\overline{X,z}) \mid X,z\right\rangle\end{aligned}$$

□

3.3 Places to Fiddle 1: NNOs

We mentioned in the introduction that there would be various ways to go about fiddling with the apparatus we presented. We've just seen a prime place for such fiddling: rather than embedding a *weak*, *parameterized* NNO into our type theory, we could instead have shoehorned in a full PNNO. We could also move to a Cartesian closed category of types (this would roughly correspond to adding higher-order vocabulary), and replace the PNNO with a regular unparameterized natural numbers object, since in the Cartesian closed context, the parameters are not needed. Further enrichments of the underlying category of types can be expected to yield more structure at the level of term rewriting, and deliver up more "logic-free" arithmetic and mathematics.

Such choices should not be made lightly. One philosophically important thing these decisions will turn on is whether we think syntax should be decidable. It is initially hard to see how to give this up. Given two arithmetic utterances, whether the first utterance expresses something *equivalent* to the second utterance is, of course, not a matter we generally expect to be able to settle in finite time. But whether the two utterances are in fact utterances of the very same expression is something that it sure seems we *ought* to be able to verify in finite time.

But that's just an intuition, and sometimes there are grounds for challenging even very strongly held intuitions. We argue that in this case, a closer look suggests that the intuition isn't as well-grounded as it initially seems. Here is one way to see this:

Begin with the observation we made at the very beginning of the paper: without a doubt, our standard logical and arithmetic practices build in at least *some* identities at the level of syntax. We acknowledge, of course, the obvious fact that whenever we *do* endorse the claim that utterance_1 and utterance_2 are utterances of the very same expression, we determine that this is so using only finite resources and finite time. But if *this* is the sole source of the intuition that our arithmetic syntax ought to be decidable (and we don't know what *else* might be grounding such an intuition), then the intuition strikes us as not one we ought to cling to all that tightly.

This is because we are, it seems, in the same boat in the case of arithmetic itself. Each member of the set of arithmetic computations we ever have performed has been performed in finite time, and we don't expect that to change.[6] Should we conclude from this that we ought to be formalizing our arithmetical practice in such a way that the resulting theory is decidable? Surely not—or at least, surely not *on these grounds alone*. The fact of the matter in the case of arithmetic seems to be that we muddle through using, in the finite time available to us, a system where not everything can be decided in finite time.

The evidence seems to be completely parallel in the two cases. In both cases, we have the obvious fact that we engage in the practice at hand using only finite time and finite resources. In the case of arithmetic we do not take this evidence to conclusively determine that the correct theory should be decidable in finite time. We thus see no reason to think the evidence *is* conclusive in the case of syntax either. And, at least to us, it's not clear what other evidence is going to be forthcoming.

If this argument (or some more cautious working-out of it) holds water, it would point to a philosophically reputable route to adopting much more robust theories of identity at the level of syntax than the one we've given here. Regardless, it seems clear that even going so far as including an NNO and exponential types in our type theory is a less suspect option than is the choice to add an omega rule to our logic—after all, one can work with an undecidable syntax without ever actually engaging in anything requiring infinite resources, but one cannot ever even once deploy an omega rule without expending infinite resources.

Overall, it seems to us that this is a good place to poke and prod the machinery in hopes of scaring up $R^{\#\frac{1}{2}}$-ish theory candidates.

[6]For an alternative perspective see [21].

4 Linguistic Matters 3: Languages and Logic

So far we've dealt only with the 'term'y part of the language. We'll now turn to constructing the rest of the language, and then augment it with a logic. Since we're having such fun with them already, we'll also add in another equivalence relation to keep track of.

4.1 The Languages

The elementary language is constructed as follows:

- If $\sigma \mid X$ and $\tau \mid X$ are unary elementary terms then $\sigma = \tau \mid X$ is an (atomic) elementary formula.
- If $\phi \mid X$ and $\psi \mid X$ are elementary formulas, then so are $\mathtt{t} \mid X$, $\sim\phi \mid X$, $\phi \wedge \psi \mid X$, and $\phi \rightarrow \psi \mid X$.
- If $y \in X$ and $\phi \mid X$ is an elementary formula, then $\forall y \phi \mid X \setminus \{y\}$ is an elementary formula.

We take $\vee$, $\leftrightarrow$ and $\exists$ to be defined in the usual ways. We define $\phi \otimes \psi \mid X$ to mean $\sim(\phi \rightarrow \sim\psi) \mid X$. As with elementary terms, we see elementary formulas as having untyped parts and typing parts, and we define the notion of an untyped elementary formula as well. If τ and σ are untyped n-ary elementary terms, then $\tau = \sigma$ abbreviates the untyped elementary formula $\tau_1 = \sigma_1 \otimes \cdots \otimes \tau_n = \sigma_n$. If Y is a finite set of variables, then $\overline{\forall Y}$ will mean $\forall y_1 \forall y_2 \ldots \forall y_n$, where $Y = \{y_1, \ldots, y_n\}$ and the y_i's are listed in increasing subscript-order.

We can lift the rewrite rules we applied to terms to the new expressions introduced here, and we will do so without comment. As with terms, we say that the elementary formulas $\phi \mid X$ and $\psi \mid Y$ are *directly equivalent* when either (a) $X = Y$ and $\phi \risingdotseq \psi$ or (b) $\phi \mid X = \psi(\overline{Y}/\overline{X}) \mid X$, and we take *equivalence* simpliciter to be the reflexive and transitive closure of direct equivalence. Finally, we lift the pointy-bracket notation to this new setting as well, and write $\langle \phi \mid X \rangle$ for the equivalence class of $\phi \mid X$. We will call $\langle \phi \mid X \rangle$ a *formula*.

4.2 The Logic

Below we give a Hilbert-style axiomatization of the logic we are interested in. Note that the logic we define is a logic on the *formulas*, not on the *elementary formulas*. Thus, the logic does not distinguish between different 'versions' of a formula that differ only at the level of the rewrite rules. That is, formulas that are transparently—read primitive recursively—the same are not to be distinguished. In addition to this strong theory of identity at the level of syntax, we've adopted very strong axioms governing the equality relation. In §7.1 we will discuss why we adopt such strong axioms. But it's worth noting already that the identity axioms look to be another part of the control panel where a bit of button-mashing could actually pay off.

Finally, we should note that we have adopted no arithmetical axioms. All the arithmetic that the system is capable of 'doing' is done in the syntax itself. This is another place where change seems possible.[7]

At any rate, our axioms are all the well-formed instances of each of the following:

A1 $\langle \mathsf{t} \mid X \rangle$

A2 $\langle \mathsf{t} \to (\alpha \to \alpha) \mid X \rangle$

A3 $\langle \alpha \to \alpha \mid X \rangle$

A4 $\langle (\alpha \wedge \beta) \to \alpha \mid X \rangle$

A5 $\langle (\alpha \wedge \beta) \to \beta \mid X \rangle$

A6 $\langle ((\alpha \to \beta) \wedge (\alpha \to \gamma)) \to (\alpha \to (\beta \wedge \gamma)) \mid X \rangle$

A7 $\langle (\alpha \wedge (\beta \vee \gamma)) \to ((\alpha \wedge \beta) \vee (\alpha \wedge \gamma)) \mid X \rangle$

A8 $\langle {\sim}{\sim}\alpha \to \alpha \mid X \rangle$

A9 $\langle (\alpha \to {\sim}\beta) \to (\beta \to {\sim}\alpha) \mid X \rangle$

A10 $\langle (\alpha \to \beta) \to ((\beta \to \gamma) \to (\alpha \to \gamma)) \mid X \rangle$

[7]It's important to note that changes at this level are a bit more fraught than some of the other changes we've highlighted. The reason, which will become clear in light of the discussion in the remainder, is that changes at the level of the axioms will result in changes in the sorts of algebras we have to use in the semantics, whereas changes in the rewrite rules only result in changes in the category $\mathcal{B}$ that we defined above.

A11 $\langle \alpha \to ((\alpha \to \beta) \to \beta) \mid X \rangle$

A12 $\langle (\alpha \to (\alpha \to \beta)) \to (\alpha \to \beta) \mid X \rangle$

A13 $\langle \forall x \phi \to \phi(x/y) \mid X \rangle$ where y is free for x in ϕ.

A14 $\langle \forall x(\phi \to \psi) \to (\phi \to \forall x \psi) \mid X \rangle$ where x is not free in ϕ.

A15 $\langle \forall x(\phi \vee \psi) \to (\phi \vee \forall x \psi) \mid X \rangle$ where x is not free in ϕ.

A16 $\langle \tau = \tau \mid X \rangle$

A17 $\langle (\sigma = \tau \otimes \phi(v/\sigma)) \to \phi(v/\tau) \mid X \rangle$ provided each τ_i, σ_i is free for v in ϕ

We close these axioms under the well-formed instances of the following rules:

R1 $$\frac{\langle \alpha \mid X \rangle \qquad \langle \alpha \to \beta \mid X \cup Y \rangle}{\langle \beta \mid Y \rangle}$$

R2 $$\frac{\langle \alpha \mid X \rangle \qquad \langle \beta \mid Y \rangle}{\langle \alpha \wedge \beta \mid X \cup Y \rangle}$$

R3 $$\frac{\langle \phi \mid X \rangle}{\langle \forall x \phi \mid X - \{x\} \rangle}$$

When $\langle \phi \mid X \rangle$ is a theorem, we write $\vdash \langle \phi \mid X \rangle$.[8]

A few useful lemmas are worth stating immediately:

Lemma 4.1. *If v does not occur in X or in ϕ and $\vdash \langle \phi \mid X, v \rangle$ then $\vdash \langle \phi \mid X \rangle$.*

Lemma 4.2. *Antilogism $\vdash \langle (\phi_1 \otimes \phi_2) \to \phi_3 \mid X \rangle$ iff $\vdash \langle (\phi_1 \otimes \sim \phi_3) \to \sim \phi_2 \mid X \rangle$.*

Lemma 4.3. *$\vdash \langle (\phi_1 \otimes \phi_2) \to \psi \mid X \rangle$ iff $\vdash \langle \phi_1 \to (\phi_2 \to \psi) \mid X \rangle$.*

Definition 4.4. We say that $\langle \phi \mid X \rangle$ and $\langle \psi \mid X \rangle$ are equivalent when they are both well-formed and $\langle \phi \leftrightarrow \psi \mid X \rangle$ is a theorem, and we write $[\phi \mid X]$ for the class of formulas equivalent to $\langle \phi \mid X \rangle$. Finally, we say $[\phi \mid X] \leq [\psi \mid X]$ when for some (and hence every) pair of members $\langle \phi_0 \mid Y \rangle$ of $[\phi \mid X]$ and $\langle \psi_0 \mid Y \rangle$ of $[\psi \mid X]$, $\phi_0 \to \psi_0 \mid Y$ is a theorem.

[8] Aside from the addition of identity and $\mathtt{t}$, this is exactly the axiomatization of the logic R given in [2].

We leave to the reader the task of checking that all this is well-defined. We extend other notions from pointy-bracket land to square-bracket land in the expected ways; e.g. we write $\vdash [\phi \mid X]$ to mean that $\vdash \langle\psi \mid Y\rangle$ for some (and hence every) member $\langle\psi \mid Y\rangle$ of $[\phi \mid X]$.

5 Preview

In the remainder of the paper we build a semantic theory that validates (as we prove) all and only the theorems of the logic we just defined. Before we do that, it's worthwhile to pause to give an overview of how the construction will proceed.

We will give an essentially algebraic semantics. Recall that at the heart of many completeness proofs is the construction of an object often called the *canonical model*. In algebraic settings, the canonical model is typically constructed from equivalence classes of formulas, on which one defines operations corresponding to the connectives and operators in the logic. As an example, to show classical logic is complete with respect to the class of boolean algebras, one standardly constructs what has come to be known as the Lindenbaum algebra.

But regardless of whether they proceed via the method of canonical models or by any other means, completeness proofs *typically* come *after* soundness proofs. It is also *typical* for both soundness proofs and completeness proof to come *after* the semantics is introduced.

Apart from the construction of the canonical model, we will be atypical in all respects. More to the point, our *first* step is to arrange the equivalence classes $[\phi \mid X]$ into an appropriate canonical model. Once we have the canonical model on hand, we will *then* build the rest of our semantic theory. Intuitively, one can imagine proceeding as follows: first, we 'abstract away' from the particular features of the canonical model in order to arrive at broader a family of objects we will call models. With new models in hand, we define a notion validity. We are constrained in these tasks by the requirement that the canonical model remain a model and that only theorems be validated by the canonical model. Provided we meet these constraints, proving completeness—that is, that only validated formulas are theorems—is entirely trivial—any non-theorem will counterexampled by the canonical model. But the logic being *sound* will be quite the surprising result. After all, we'll be evaluating formulas in a range of structures quite different from the one we started at—and which we arrived at not by reflection on truth or meaning or anything like that, but rather by a process of abstraction guided by heuristics and mathematical intuition. That following these guides *doesn't* lead us to models that falsify any of our theorems is downright shocking.

Philosophers of a certain bent will be crabby about this order of going about things. They will claim that completeness proofs, when done correctly, ought to be surprising. The view of logic espoused by such philosophers is something like this:

> We have a fairly good sense for what the world is like. Using this sense of what the world is like, we build models. We then define the set of formulas we're interested in to be the *valid* formulas—that is, the formulas that are true in all the models. After doing this, we then set about trying to figure out what we've done—that is, we set ourselves the task of actually gathering up all the valid formulas. Since these are the *only* sentences we're interested in, the fact that we have a sound semantic theory is completely unsurprising. The goal, however, isn't to gather up just *some* of the valid formulas but to in fact gather up *all* of them. Intuitively, this is a pretty hard thing to do. So we should find it surprising that we're actually able to do the job. Thus completeness proofs—which say not only that we *can do* but that we in fact *have done* the job—ought to be surprising.

This way of thinking about logic is familiar and friendly and, to our minds, entirely backwards. What we have a good understanding of is what *language* is like, not what *the world* is like. This understanding of language tells us something about how the world-related concepts we deploy—e.g. terms, predicates, and the like—interact. We can capture all of that in our canonical model, which organizes all of this information in a mathematically perspicuous framework. What we *don't* (and, perhaps, *can't*?) know is how well our language maps onto whatever extra-linguistic reality there is. And this for the simple reason that we don't know much about said extra-linguistic reality at all! Supposing we *do* know about such matters is (puzzlingly) where the grumpy-at-us philosophers seem to start.

Be that as it may, since we don't know how well word and world match up, we must admit that things we've distinguished may turn out to be, on interpretation into the extra-linguistic realm, not distinct. Or there might be more structure to the extra-linguistic world than there is in our language. We capture these possibilities by building natural generalizations of our canonical model.[9] When it turns out that the logic we began with is sound *across all the generalizations*, we should be surprised—after all, we used hints from language, and now we're ending up at something that tells us how the world might be organized. In fact, based on other experiences in which implausibly-good explanations show up, we might think this surprisingness is itself evidence that we've hit on something deep.[10]

At any rate, we've said all this as part of a defense of why it should be the case that, on the view we take, it's soundness and *not* completeness that is the surprising metatheoretic result. And, regardless of whether you buy the story we're telling, it tells you both how the remainder of the paper will be organized and why we've chosen that way of organizing it. In more detail, the next section will describe how we arrange the classes $[\phi \mid X]$ into canonical-model-shape. We will then take another section to recognize how the first-order machinery of our logic is realized in the canonical model. After that, we'll generalize and provide the full semantics before we turn to the metatheory. We will end the paper by examining how much arithmetic we get.

6 First Steps

The basic plan for the construction of the canonical model—which we will sometimes call *the syntactic hyperdoctrine*—is this: to each type T_n, we associate an algebra $S(T_n)$ of propositional functions in n variables. To each class of terms $\langle \tau \mid X \rangle$, we associate a morphism of algebras $S(T_{\mathrm{len}(\tau)}) \longrightarrow S(T_{\mathrm{card}(X)})$. Intuitively, these morphisms correspond to substitutions—they map each propositional function $[\phi \mid Y]$ in $S(T_{\mathrm{len}(\tau)})$ to the propositional function $[\phi(\overline{Y}/\tau) \mid X]$ in $S(T_{\mathrm{card}(X)})$. The structure of the resulting system

[9]This ignores the possibility that the extra-linguistic reality is organized in some way that is *utterly* different than what our language suggests. Being frank, we're not sure what to make of this worry—on the one hand, it feels like it's certainly *possible* that we're utterly wrong about everything. On the other hand, what it would mean to say that we are is not entirely clear. Regardless, it seems that language gives us not just our best but also our only guess about the structure of the world, so making the assumption that it's at least right in regard to the gross details seems a not too-irresponsible thing to do.

[10]It strikes us that, pursued to its natural conclusion, this line of thought leads to a distinctly Popperian sort of antiexceptionalism about logic.

of algebras and morphisms then determines how we interpret all the other elements. In particular, both quantifiers and identities 'show up' as adjoints in a way to be explained below. In order to easily state the results and definitions we require, the following analogue of Fact 3.10 is crucial:

Fact 6.1. *If card$(X) = n$, then for some untyped formula ψ, $[\phi \mid X] = [\psi \mid X_1^n]$.*

Definition 6.2. We define the algebras $S(T_n)$ as follows:

- As a set, $S(T_n)$ is the collection of classes of formulas of the form $[\phi \mid X_1^n]$.
- As a lattice $S(T_n)$ is ordered by $\leq$ with meet $\sqcap$ and join $\sqcup$ defined by $[\phi \mid X_1^n] \sqcap [\psi \mid X_1^n] := [\phi \wedge \psi \mid X_1^n]$ and $[\phi \mid X_1^n] \sqcup [\psi \mid X_1^n] := [\phi \vee \psi \mid X_1^n]$, respectively.
- The lattice $S(T_n)$ is equipped with a De Morgan involution we denote by $\neg$, and define by $\neg[\phi \mid X_1^n] := [\sim \phi \mid X_1^n]$.
- Finally we define a monoidal operation $\boxtimes$ on the De Morgan lattice $S(T_n)$ by saying $[\phi \mid X_1^n] \boxtimes [\psi \mid X_1^n] := [\phi \otimes \psi \mid X_1^n]$.

Fact 6.3. *$S(T_n)$, with the above operations, is a De Morgan Monoid.*[11]

Now a confession: we said above that each class of terms $\langle \tau \mid X \rangle$ would be associated to the substitution morphism given by $[\phi \mid Y] \longmapsto [\phi(\overline{Y}/\tau) \mid X]$. This isn't *exactly* right—which is to say that we lied to you when we said this. We promise it was for your own good, though, as the lie is a lie is for reasons that are both uninteresting and annoying.

In short, the problem is that we need to take care of a bit of 'free-for' fiddliness. More to the point, we are still using elementary formulas (e.g. $\phi \mid Y$) to stand for the formulas $\langle \phi \mid Y \rangle$ that represent the equivalence classes $[\phi \mid Y]$. And, when it comes to elementary formulas, there is a worry that we might substitute in variables that are not free for the variable they're replacing.

But fixing this isn't hard, and there are many ways to do it. We choose a particularly concrete solution, which looks a little uglier than some of the

[11] De Morgan monoids were introduced in [3]. Further examinations or applications of De Morgan monoids can be found in [24], [4], [25], [19], and [20].

alternatives. The idea is to choose the representative elementary term in $\langle \tau \mid X \rangle$ in such a way that each τ_i is free for the variables that are free in $\phi \mid Y$, which we are using as our representative of our representative of $[\phi \mid Y]$. An easy and effective way to accomplish this is to use for our representative an elementary term $\sigma \mid Z \in \langle \tau \mid X \rangle$ with no member of Z occurring anywhere in the elementary formula $\phi \mid X$. working through the details we're led to the following definition for our substitution function:

Definition 6.4. For each term ($\mathcal{B}$-arrow) $\langle \tau \mid X \rangle : T_{\mathrm{card}(X)} \longrightarrow T_{\mathrm{len}(\tau)}$, the morphism $S_{\langle \tau \mid X \rangle} : S(T_{\mathrm{len}(\tau)}) \longrightarrow S(T_{\mathrm{card}(X)})$ is defined by $[\phi \mid Y] \longmapsto [\phi(\overline{Y}/\tau(\overline{X}/\overline{Z})) \mid Z]$ where Z is some set of variables no member of which occurs anywhere in $\phi \mid Y$.

Fact 6.5. *$S_{\langle \tau \mid X \rangle}$ is a well-defined De Morgan monoid homomorphism*

6.1 Some Special Functors

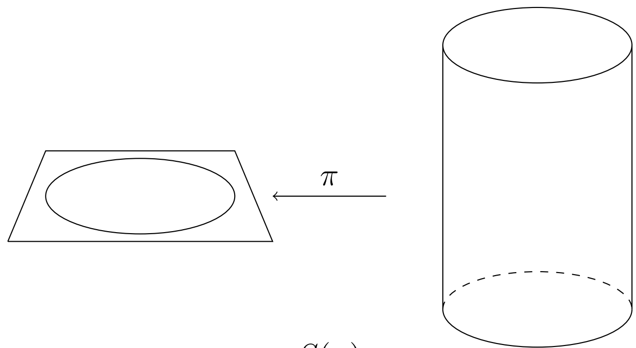
We can summarize what we've seen so far as follows:

Lemma 6.6. *S is a contravariant functor that maps each $\mathcal{B}$-object (type) T_n to the De Morgan Monoid $S(T_n)$ of n-adic propositional functions and maps each $\mathcal{B}$-morphism (term) $\langle \tau \mid X \rangle$ to the algebra homomorphism $S_{\langle \tau \mid X \rangle}$.*

For reasons that will become clear later, we will think of S as a functor from $\mathcal{B}$ to the category of posets, rather than as a functor from $\mathcal{B}$ to the category of De Morgan monoids. Be that as it may, certain morphisms in $\mathcal{B}$ correspond to algebra homomorphisms that play an important role in the story we're telling. So we'll pause to have a look at them.

The first class of special $\mathcal{B}$-morphisms we will look at are the projections $T_{n+m} \longrightarrow T_n$. A short bookkeeping exercise reveals that these correspond to the terms $\langle \overline{X_1^n} \mid X_1^{n+m} \rangle$. S promotes these morphisms to what might be thought of as *inclusion* morphisms, including the algebra $S(T_n)$ of propositional functions with n free variables into the algebra $S(T^{n+m})$ of propositional functions with $n+m$ free variables. For example, it's easy to verify by inspection that $S_{\langle \overline{X_1^n} \mid X_1^{n+m} \rangle}$ transports $[\phi \mid X_1^n]$ to $[\phi \mid X_1^{n+m}]$.

Strictly for intuition, we offer some analogies. If the reader is inclined to geometric thinking, this is quite like taking the preimage under projection of a set defined by ϕ:

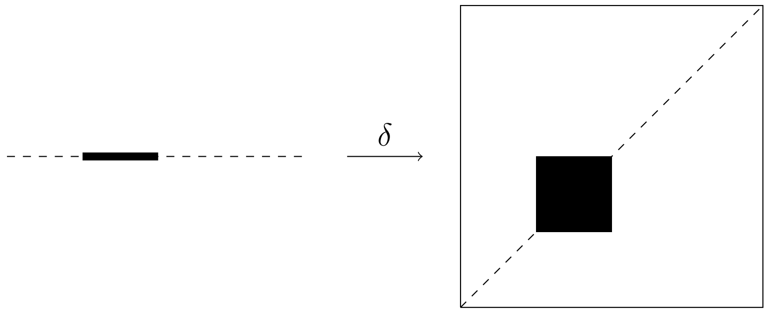


$$\langle x^2 + y^2 = 0 \mid x, y\rangle \xrightarrow{S(\pi)} \langle x^2 + y^2 = 0 \mid x, y, z\rangle$$

Or, if the reader prefers to think of the types of variables in a context using an analogy with propositions (the so-called propositions-as-types perspective), then this is rather like applying a round of weakening to the left hand side of a sequent.[12]

The next class of $\mathcal{B}$-morphisms we examine are the diagonal morphisms, which we can identify with the terms $\langle\overline{X_1^n X_1^n} \mid X_1^n\rangle$. Note that such morphisms *identify* (hint hint) distinct variables in a formula—for example, $S_{\langle xx|x\rangle}$ transports $[\phi(u,v) \mid u,v]$ to $[\phi(x,x) \mid x]$. Again, for intuition, we note that geometrically, $S_{\langle\overline{X_1^n X_1^n}|X_1^n\rangle}$ is akin to taking a preimage along the map δ that embeds a lower-dimensional space as a diagonal in a higher dimensional space.

[12]This is a bit clearer if one adopts the suggestive and more standard notation $x :: X \vdash \phi$ for what we would write as $\phi \mid x$. In this case, an application of $S(\pi)$, where $\pi : X \times Y \to X$ moves $x :: X \vdash \phi$ to $x :: X, y :: Y \vdash \phi$. We should note that just because adding variables to a context is interestingly *analogous* to the structural rule of weakening, that does not mean that this framework automatically validates the actual structural rule of weakening. If it did automatically validate that rule, that would of course be problematic from the perspective of relevance.



$$\langle 1 < x < 2 \wedge 1 < x < 2 \mid x\rangle \xleftarrow{S(\delta)} \langle 1 < x < 2 \wedge 1 < y < 2 \mid x, y\rangle$$

Or, again, if the reader prefers to employ the propositions-as-types perspective, then this is rather like applying a round of contraction to the left-hand side of a sequent, in that one replaces a requirement of two variables of the same type (two proofs of a certain proposition) with the requirement of one variable of that type.

7 First-Order Structure

Now we turn to the *first*-order structure of the canonical model. This structure arises via adjoints to the special functors identified in §6.1. The induced structure has two main parts: quantifiers and identities. They are intimately connected, as we will see.

Lemma 7.1. *Our implications support a version of Lawvere's rule for equality. If $\langle \tau \mid X\rangle$ is a unary term and v is a variable not in $\tau \mid X$ or $\phi \mid X$ we have*

$$\vdash [(\tau = v \otimes \phi) \to \psi \mid X, v] \qquad \textit{iff} \qquad \vdash [\phi \to \psi(v/\tau) \mid X]$$

Proof. Without loss of generality, assume nothing in X is bound in ϕ or in ψ.

For left to right, notice that if $\vdash [(\tau = v \otimes \phi) \to \psi \mid X, v]$, then $\vdash [\forall v((\tau = v \otimes \phi) \to \psi) \mid X]$. Thus $\vdash [(\tau = \tau \otimes \phi(v/\tau)) \to \psi(v/\tau) \mid X]$. But v does not occur in ϕ, so $\vdash [(\tau = \tau \otimes \phi) \to \psi(v/\tau) \mid X]$. So by Lemma 4.3, $\vdash [\tau = \tau \to (\phi \to \psi(v/\tau)) \mid X]$. So $\vdash [\phi \to \psi(v/\tau) \mid X]$.

For right to left, suppose $\vdash [\phi \to \psi(v/\tau) \mid X]$. Then $\vdash [{\sim}\psi(v/\tau) \to {\sim}\phi \mid X]$. Thus we also have that $\vdash [{\sim}\psi(v/\tau) \to {\sim}\phi \mid X, v]$. Also by A17,

$\vdash [(\tau = v \otimes \sim\psi) \to \sim\psi(v/\tau) \mid X, v]$. So $\vdash [(\tau = v \otimes \sim\psi) \to \sim\phi \mid X, v]$, and thus by Lemma 4.2, $\vdash [(\tau = v \otimes \phi) \to \psi \mid X, v]$. $\square$

We implicitly generalize this lemma to the case where $\langle \tau \mid X \rangle$ is an n-ary term for $n > 1$ in the obvious way.

Definition 7.2. Let $\langle \tau \mid X_1^n \rangle$ be an m-ary term, $\langle \phi \mid X_1^n \rangle$ be a formula, and Y be a set of n variables that do not occur in $\phi \mid X_1^n$. Then

- $\Pi_{\langle \tau \mid X_1^n \rangle}$ names the order-preserving function given by
$$[\phi \mid X_1^n] \longmapsto [\overline{\forall Y}(\tau(\overline{X_1^n}/\overline{Y}) = \overline{X_1^m} \to \phi(\overline{X_1^n}/\overline{Y})) \mid X_1^m];$$
- $\Sigma_{\langle \tau \mid X_1^n \rangle}$ names the order-preserving function given by
$$[\phi \mid X_1^n] \longmapsto [\overline{\exists Y}(\tau(\overline{X_1^n}/\overline{Y}) = \overline{X_1^m} \otimes \phi(\overline{X_1^n}/\overline{Y})) \mid X_1^m].$$

Note that neither $\Pi_{\langle \tau \mid X_1^n \rangle}$ nor $\Sigma_{\langle \tau \mid X_1^n \rangle}$ is an algebra homomorphism. Thus, in order to admit them into the structure of the canonical model in any way, we must (as mentioned above) view the functor S as mapping from $\mathcal{B}$ to the category of posets. This seems like an inconvenience, until we prove the next two theorems

Theorem 7.3. *$\Pi_{\langle \tau \mid X_1^n \rangle}$ is right adjoint to $S_{\langle \tau \mid X_1^n \rangle}$.*[13]

Proof. Let $\mathrm{len}(\tau) = m$. It suffices to show that $S_{\langle \tau \mid X_1^n \rangle}[\phi \mid X_1^m] \leq [\psi \mid X_1^n]$ iff $[\phi \mid X_1^m] \leq \Pi_{\langle \tau \mid X_1^n \rangle}[\psi \mid X_1^n]$. Let $\mathrm{card}(Y) = n$ and suppose no member of Y occurs anywhere in sight.

We reason as follows:

$$\begin{aligned}
S_{\langle \tau \mid X_1^n \rangle}[\phi \mid X_1^m] \leq [\psi \mid X_1^n] &\text{ iff } [\phi(\overline{X_1^m}/\tau(\overline{X_1^n}/\overline{Y})) \mid Y] \leq [\psi(\overline{X_1^n}/\overline{Y}) \mid Y] \\
&\text{ iff } [\sim\psi(\overline{X_1^n}/\overline{Y}) \mid Y] \leq [\sim\phi(\overline{X_1^m}/\tau(\overline{X_1^n}/\overline{Y})) \mid Y] \\
&\text{ iff } [\tau(\overline{X_1^n}/\overline{Y}) = \overline{X_1^m} \otimes \sim\psi(\overline{X_1^n}/\overline{Y}) \mid X_1^m, Y] \leq [\sim\phi \mid X_1^m, Y] \\
&\text{ iff } [\tau(\overline{X_1^n}/\overline{Y}) = \overline{X_1^m} \otimes \phi \mid X_1^m, Y] \leq [\psi(\overline{X_1^n}/\overline{Y}) \mid X_1^m, Y] \\
&\text{ iff } [\phi \mid X_1^m, Y] \leq [\tau(\overline{X_1^n}/\overline{Y}) = \overline{X_1^m} \to \psi(\overline{X_1^n}/\overline{Y}) \mid X_1^m, Y] \\
&\text{ iff } [\phi \mid X_1^m] \leq [\overline{\forall Y}(\tau(\overline{X_1^n}/\overline{Y}) = \overline{X_1^m} \to \psi(\overline{X_1^n}/\overline{Y})) \mid X_1^m] \\
&\text{ iff } [\phi \mid X_1^m] \leq [\Pi_{\langle \tau \mid X_1^n \rangle}(\psi \mid X_1^n)]
\end{aligned}$$

[13] It's worth pointing out that there's an important difference between the behavior of Π and the behavior of S that our notation obscures: Π is 'covariant' in its subscript where S is 'contravariant' in its subscript. What we mean by this is that where $\langle \tau \mid X \rangle$ maps from $T_{\mathrm{card}(X)}$ to $T_{\mathrm{len}(\tau)}$, $\Pi_{\langle \tau \mid X_1^n \rangle}$ maps from $S(T_{\mathrm{card}(X)})$ to $S(T_{\mathrm{len}(\tau)})$, but $S_{\langle \tau \mid X \rangle}$ maps from $S(T_{\mathrm{len}(\tau)})$ to $S(T_{\mathrm{card}(X)})$.

□

Theorem 7.4. $\Sigma_{\langle\tau|X_1^n\rangle}$ *is left adjoint to* $S_{\langle\tau|X_1^n\rangle}$

Proof. Let $\mathrm{len}(\tau) = m$. It suffices to show that $\Sigma_{\langle\tau|X_1^n\rangle}[\phi \mid X_1^n] \leq [\psi \mid X_1^m]$ iff $[\phi \mid X_1^n] \leq S_{\langle\tau|X_1^n\rangle}[\psi \mid X_1^m]$. Let $\mathrm{card}(Y) = n$ and suppose no member of Y occurs anywhere in sight. We use the previous result, the definitions of the existential and of $\otimes$, and the fact that $S_{\langle\tau|X_1^n\rangle}$ is a homomorphism. The reasoning is as follows:

$$
\begin{aligned}
\Sigma_{\langle\tau|X_1^n\rangle}[\phi \mid X_1^n] \leq [\psi \mid X_1^m] &\text{ iff } [\exists\overline{Y}(\tau(\overline{X_1^n}/\overline{Y}) = \overline{X_1^m} \otimes \phi(\overline{X_1^n}/\overline{Y})) \mid X_1^m] \leq [\psi \mid X_1^m] \\
&\text{ iff } [\sim\forall\overline{Y}(\tau(\overline{X_1^n}/\overline{Y}) = \overline{X_1^m} \rightarrow \sim\phi(\overline{X_1^n}/\overline{Y})) \mid X_1^m] \leq [\psi \mid X_1^m] \\
&\text{ iff } [\sim\psi \mid X_1^m] \leq [\forall\overline{Y}(\tau(\overline{X_1^n}/\overline{Y}) = \overline{X_1^m} \rightarrow \sim\phi(\overline{X_1^n}/\overline{Y})) \mid X_1^m] \\
&\text{ iff } [\sim\psi \mid X_1^m] \leq [\Pi_{\langle\tau|X_1^n\rangle}(\sim\phi \mid X_1^m)] \\
&\text{ iff } S_{\langle\tau|X_1^n\rangle}[\sim\psi \mid X_1^m] \leq [\sim\phi \mid X_1^m] \\
&\text{ iff } \neg S_{\langle\tau|X_1^n\rangle}[\psi \mid X_1^m] \leq \neg[\phi \mid X_1^m] \\
&\text{ iff } [\phi \mid X_1^m] \leq S_{\langle\tau|X_1^n\rangle}[\psi \mid X_1^m]
\end{aligned}
$$

□

As an important special case consider a projection $T_2 \to T_1$; e.g. $\langle x \mid x, y\rangle$. We then observe that $\Pi_{\langle x|x,y\rangle}$ is the map given by

$$[\phi(x,y) \mid x,y] \longmapsto [\forall v\forall w(v = x \rightarrow \phi(v,w)) \mid x] = [\forall y\phi(x,y) \mid x]$$

and $\Sigma_{\langle x|x,y\rangle}$ is the map given by

$$[\phi(x,y) \mid x,y] \longmapsto [\exists v\exists w(v = x \otimes \phi(v,w)) \mid x] = [\exists y\phi(x,y) \mid x]$$

In words, (universal/existential) quantification along the jth variable is (right/left) adjoint to projection onto all but the jth component.

As a second example of an important class of special cases, consider a diagonal $T_1 \to T_2$; e.g. $\langle xx \mid x\rangle$. Note that $\Sigma_{\langle xx|x\rangle}$ is the map

$$[\phi(x) \mid x] \longmapsto [\exists v(x = v \otimes y = v \otimes \phi(v)) \mid x, y].$$

If we examine in particular the action of $\Sigma_{\langle xx|x\rangle}$ on the class $[\mathsf{t} \mid x]$, we see that

$$[\mathsf{t} \mid x] \longmapsto [\exists v(x = v \otimes y = v \otimes \mathsf{t}) \mid x, y]$$

It's not immediately obvious this is at all interesting. To see that it is, first notice that it's not hard to show that $[\exists v(x = v \otimes y = v \otimes \mathtt{t}) \mid x, y] \leq [x = y \mid x, y]$. This is neat, but is made much more neat by the fact that the converse inequality also holds!

To see this, first do a bit of fiddling on a napkin and verify that (i) $\vdash \langle (x = x \otimes x = y) \to (x = x \otimes y = x) \mid x, y \rangle$. Then, by the following path, observe that (ii) $\vdash \langle x = y \to (x = x \otimes x = y) \mid x, y \rangle$:

- First, by A12, $\vdash \langle x = x \to ((x = x \to \neg x = y) \to \neg x = y) \mid x, y \rangle$.
- Next, by A16, $\vdash \langle x = x \mid x, y \rangle$.
- So $\vdash \langle (x = x \to \neg x = y) \to \neg x = y \mid x, y \rangle$, and thus, via fiddling,
- $\vdash \langle x = y \to (x = x \otimes x = y) \mid x, y \rangle$, which is what we were after.

With (i) and (ii) on hand, it's smooth sailing: just combine them to get $\vdash \langle x = y \to (x = x \otimes y = x) \mid x, y \rangle$, then return to the trusty old napkin to verify that $\vdash \langle (x = x \otimes y = x) \to \exists v(x = v \otimes y = v) \mid x, y \rangle$ as well, from which a bit of elbow grease finishes the job.

Thus identity claims are exactly what we get by evaluating the adjoint to the diagonal at $\mathtt{t}$. So *both* of the distinguished functors we identified above have adjoints that correspond to distinguished logical concepts: quantification in the case of projections and identity in the case of the diagonal.

7.1 A Brief Discussion of Identity

As we mentioned, the theory of identity we adopted is quite strong. Following the discussion in [5], we can see reasons to worry the theory is *too* strong by making the following observations. First, recall that formulas of the form $P \to (Q \to P)$ are, in Dunn's words 'dread relevance destroyers'. Now observe that we can quite easily prove something of roughly this very form in a few simple steps from our A17:

- Begin with the axiom $\langle (x = y \otimes u = v) \to u = v \mid x, y, u, v \rangle$.
- Notice that by the definition of $\otimes$ and axiom A9 (contraposition), $\otimes$ is commutative.
- So $\vdash \langle (u = v \otimes x = y) \to u = v \mid x, y, u, v \rangle$.

- But then by Lemma 4, $\vdash \langle u = v \to (x = y \to u = v) \mid x, y, u, v\rangle$ as we feared.

In addition to Dunn's work, the two most detailed philosophical examinations of identity in quantified relevance logics that we are aware of are undertaken by Ed Mares (see [17]) and Philip Kremer (see [10], [11], [12], and [13].) Both Mares and Dunn aggressively restrict the identity axioms we've given. Whatever its merits, we won't pursue this approach here as it would destroy the adjointness features we've just identified.

Since we *don't* pursue this strategy, we owe a defense of our theory of identity. The defense comes in three parts. The first part of our defense draws on Philip Kremer's examination of Dunn's theory in [12]. It's worth our while to quote from this paper at length:

> [Dunn's central move] is the rejection of $p \to (x = c \to p)$ based on its similarity to $p \to (q \to p)$. The appropriateness of this rejection depends on the intended interpretation of identity, though Dunn's motivation does not explicitly rely on any particular interpretation. Rather, it relies on our typographical intuitions.
>
> Typographical intuitions are useful and underlie much of the motivation for relevance logic. But these intuitions are defeasible. Consider the atomic propositional constant, t, which is often added to relevance logics. t is standardly interpreted as the conjunction of all theorems. Despite the typographical similarity of $p \to (\mathsf{t} \to p)$, to $p \to (q \to p)$ the former is accepted as a theorem of the relevance logic R: we rely in the end on the interpretation of the new logical vocabulary.
>
> Like 't', '=' is a piece of logical vocabulary, open to interpretation... The moral to be drawn is that Dunn's notion of relevant predication implicitly relies on some weaker interpretation of identity.[12, pp.40ff]

There are two points being made here that are worth sussing apart. The first is roughly the following: Dunn has observed that our choice of identity axiom commits us to something that is typographically quite like something we don't like. In particular, $u = v \to (x = y \to u = v) \mid x, y, u, v$ is quite like $P \to (Q \to P)$. Since we fear the latter we should fear the former. But notice that reason to fear and reason to back down are different things. So we might take Dunn's observation as gesturing in the direction of a need for caution without taking it to be a reason to stop altogether.

The second point we take Kremer to be making is a complaint about the theory of relevant predication itself. What Kremer seems to be pointing to is something roughly circular in Dunn's reasoning. In particular, Dunn is attempting to work out a theory of relevant predication. Relevant predication, for Dunn, seems to have something to do with, as Kremer puts it 'what follows relevantly from claims of the form $x = y$'.[14] But in order to keep relevant and ordinary predication from collapsing into one another, Dunn has to interpret '=' in a nonstandard way. And the only motivation available for such a move is that it's what relevant predication requires. Thus relevant predication is grounded in some way on features of relevant identity which is itself grounded in some way on relevant predication again. Working the details out here will reveal whether there is something genuinely circular or not. We take it, however, that this leaves us with reason to be concerned about relevant predication.

However, at the end of the day, Kremer comes down on the side of Dunn's theory of relevant predication. So we'd best not hang our whole defense on Kremer's worries, since Kremer, at least, thinks Kremer's worries can in fact be addressed. So we move on to other proposals.

The most straightforward of these is this: there is more than just identity at play in our derivation of a problematic formula using our strong theory of identity. In particular, the derivation also relied on $\otimes$ being commutative. Thus, were we to reject this commutativity, we could perhaps be safe.

We had to qualify the final sentence here with 'perhaps' because there is a worry to be had: the theory of identity we *want* is one in which identity arises, as it does here, from the evaluation at $\mathtt{t}$ of an adjoint to the diagonal. Whether we can accomplish this *at all* or *with a theory such as ours* once we reject commutativity for $\otimes$ is something we have not investigated.[15] But it seems to us that if evaluation at $\mathtt{t}$ of an adjoint to the diagonal gives us, in R, an unacceptably strong theory of identity, perhaps we should lay the blame not on using the thing we get by evaluation at $\mathtt{t}$ of an adjoint to the diagonal, but on R.

All of this, of course, depends for its plausibility on some reason for taking

[14] [12, p.38]

[15] The results in [18] seem to suggest that things can be worked out. But this at best takes us from the realm of complete guesses to the realm of educated guesses.

seriously evaluation at $\mathtt{t}$ of an adjoint to the diagonal as giving us the right account of identity. So our final defense of our theory of identity is a defense of this account itself. We offer two considerations in favor.

First: regarding identity as arising from an adjoint functor in this way gives a simple unification of the standard properties of identity. Because $\Sigma_{\langle xx|x\rangle}$ is a left adjoint to $S_{\langle xx|x\rangle}$, reflexivity arises as the unit natural transformation $e_1 \leq S_{\langle xx|x\rangle}\Sigma_{\langle xx|x\rangle}(e_1)$, corresponding to the implication

$$\langle \mathtt{t} \to x = x \mid x\rangle$$

Similarly, Leibniz's law, in the formulation "If x is equal to something with the property ϕ, then x has ϕ" arises as the counit natural transformation $\Sigma_{\langle x|x\rangle}S_{\langle x|x\rangle}([\phi \mid x]) \leq [\phi \mid x]$, corresponding to the implication

$$\langle \exists v(x = v \otimes \phi(x/v)) \to \phi \mid x\rangle$$

So, both the standard introduction and elimination properties of identity are encapsulated by the adjoint functor presentation. This is certainly striking. More concretely, we can point out that it gives a simple and unified explanation of the behavior of the notion of identity, and that that explanation dovetails perfectly with a larger body of theory (namely category theory). So, there are strong abductive considerations in favor of this approach to identity.

Second: regarding identity as arising from an adjoint functor in this way clarifies the status of identity as a purely logical notion. It does so in a number of ways. But we will mention only one: unit/counit symmetry. The point here is that the adjoint functor presentation exhibits a profound symmetry between identification and quantification that would be quite difficult to see in any other way. Identity emerges from the left adjoint to the substitution induced by our diagonal map, $\langle xx \mid x\rangle$. Existential quantification emerges from the left adjoint to substitution induced by projection, for example $\langle x \mid x, y\rangle$. However, projection and the diagonal are, from the right perspective, symmetrical operations.

Recall that the product on a category $\mathcal{C}$ with finite products, regarded as a bifunctor $\mathbb{P} : \mathcal{C} \times \mathcal{C} \to \mathcal{C}$, is the right adjoint to the diagonal functor $\mathbb{D} : \mathcal{C} \to \mathcal{C} \times \mathcal{C}$, mapping each object X in $\mathcal{C}$ to a pair $\langle X, X\rangle$ in $\mathcal{C} \times \mathcal{C}$. So we have a unit natural transformation $\eta_\bullet : \mathbb{I} \to \mathbb{PD}$, where $\mathbb{I}$ is the identity functor, and a counit $\varepsilon_\bullet : \mathbb{DP} \to \mathbb{I}$. Concretely, these work out to $\eta_X : X \to X \times X$ and $\varepsilon_{\langle X,Y\rangle} : \langle X \times Y, X \times Y\rangle \to \langle X, Y\rangle$. Inspecting carefully, you will find that the components of the counit $\varepsilon_{\langle X,Y\rangle}$ are the projections π_1, π_2 associated with the product, and the unit η_X is the diagonal map δ.[16]

So, the argument would be this: existential quantification is a purely logical notion. But existential quantification is nothing but the left adjoint to substitution along a component of the counit of the adjunction $\mathbb{D} \dashv \mathbb{P}$. Similarly, identification is nothing but the left adjoint to substitution along the *unit* of the adjunction $\mathbb{D} \dashv \mathbb{P}$. So identification should also be admitted as purely logical—to do otherwise would be to draw the logical/non-logical boundary capriciously.

8 Semantics, Generally

We have our canonical model—the syntactic hyperdoctrine S—in hand. We now turn to building the rest of our semantic theory. Recall that our plan was to do this by 'abstracting away' from the particular features of the canonical model. We begin the process as follows:

Definition 8.1. A *De Morgan monoid pseudohyperdoctrine* is a contravariant functor $H : \mathcal{B} \longrightarrow \mathbf{Pos}$ such that[17]

1. The image of H is in the category of De Morgan monoids, in the sense that for each type T_n, $H(T_n)$ is a De Morgan monoid and for each term $\langle \tau \mid X\rangle$, $H_{\langle \tau|X\rangle}$ is a De Morgan monoid homomorphism;

2. Each $H_{\langle \tau|X\rangle}$ has a right adjoint $\Pi_{\langle \tau|X\rangle}$.

We leave it to the reader to verify the following fact:

Fact 8.2. *In any De Morgan monoid pseudohyperdoctrine, the map* $\Sigma_{\langle \tau|X_1^n\rangle}$ *defined by* $H(T_n) \ni h_n \longmapsto \neg\Pi_{\langle \tau|X\rangle}(\neg h_n)$ *is left adjoint to* $H_{\langle \tau|X\rangle}$.

For full hyperdoctrines, we will require two more features that seem important to the proper functioning of the syntactic hyperdoctrine, but which

[16] This has the nice consequence that properties like $\pi_1 \circ \delta = \mathrm{Id}$ emerge simply from standard results on adjunctions (in this case, by the unit-counit equations).

[17] We take a fairly liberal view about who, exactly, **Pos** is. In particular we suppose that *different* De Morgan monoids—even those with the same underlying posets—occur as *different* objects in **Pos**. If you are scandalized by our libertinism, please feel free to view each pseudohyperdoctrine as coming equipped with a section of the appropriate forgetful functor, and to make the appropriate adjustments to what follows.

are easy to overlook. The first two are analogues of our axioms A14 and A15. Let $h_n \in H(T_n)$ and $h_{\text{len}(\tau)} \in H_{\text{len}(\tau)}$. Then

3. $\Pi_{\langle\tau|X_1^n\rangle}\left(H_{\langle\tau|X_1^n\rangle}(h_{\text{len}(\tau)}) \sqcup (h_n)\right) \leq (h_{\text{len}(\tau)}) \sqcup \Pi_{\langle\tau|X\rangle}(h_n)$
4. $\Pi_{\langle\tau|X_1^n\rangle}\left(H_{\langle\tau|X_1^n\rangle}(h_{\text{len}(\tau)}) \boxtimes (h_n)\right) \leq (h_{\text{len}(\tau)}) \boxtimes \Pi_{\langle\tau|X\rangle}(h_n)$

Next, (5), we suppose that whenever all of it makes sense, the following diagram commutes:[18]

$$\begin{array}{ccc}
H(T_{m+1}) & \xrightarrow{\Pi_{\left\langle\overline{X_1^m}|X_1^{m+1}\right\rangle}} & H(T_m) \\
{\scriptstyle H_{\left\langle\tau x_{n+1}|X_1^{n+1}\right\rangle}}\downarrow & & \downarrow{\scriptstyle H_{\left\langle\tau|X_1^n\right\rangle}} \\
H(T_{n+1}) & \xrightarrow[\Pi_{\left\langle\overline{X_1^n}|X_1^{n+1}\right\rangle}]{} & H(T_n)
\end{array}$$

Roughly, this says that substitutions commute with quantification. Concretely, note that if $[\phi \mid X_1^{m+1}] \in H(T_{m+1})$, then $\Pi_{\left\langle\overline{X_1^m}|X_1^{m+1}\right\rangle}[\phi \mid X_1^{m+1}] = [\forall x_{m+1}\phi \mid X_1^m]$, and thus by the top path, we end up at $[\forall x_{m+1}\phi(\overline{X_1^m}/\tau) \mid X_1^n]$. When we follow the other path, we first compute $H_{\left\langle\tau x_{n+1}|X_1^{n+1}\right\rangle}[\phi \mid X_1^{m+1}] = [\phi(\overline{X_1^m}/\tau)(x_{m+1}/x_{n+1}) \mid X_1^{n+1}]$. Composing this with $\Pi_{\overline{X_1^n}|X_1^{n+1}}$ gives $[\forall x_{n+1}\phi(\overline{X_1^m}/\tau)(x_{m+1}/x_{n+1}) \mid X_1^n]$. Comparing these two results (with a few added parentheses as below) makes clear that they are in fact the same:

$$[(\forall x_{m+1}\phi)\,(\overline{X_1^m}/\tau) \mid X_1^n] = [\forall x_{n+1}\left(\phi(\overline{X_1^m}/\tau)(x_{m+1}/x_{n+1})\right) \mid X_1^n]$$

Finally, (6) we need a 'Frobenius Reciprocity' condition:

$$\Sigma_{\langle\tau|X_1^n\rangle}\left(H_{\langle\tau|X_1^n\rangle}(h_{\text{len}(\tau)}) \boxtimes h_n\right) = h_{\text{len}(\tau)} \boxtimes \Sigma_{\langle\tau|X_1^n\rangle}(h_n)$$

Note that this is essentially nothing more than a strengthened, existential version of the confinement axiom A14.

Definition 8.3. A De Morgan monoid hyperdoctrine is a De Morgan pseudohyperdoctrine that satisfies (3)-(6)

[18]There is a more general version of this that is typically included in this sort of semantic theory under the name of the Beck-Chevalley condition. We include only the following limited form, for the simple reason that it's all we need and it's easier to grasp than is the more general option.

Every De Morgan monoid hyperdoctrine 'comes equipped' with a full interpretation of the vocabulary of our language. More to the point, for each De Morgan monoid hyperdoctrine, there is a corresponding function that maps formulas $\langle \phi \mid X \rangle$ to semantic values $[\![\langle \phi \mid X \rangle]\!]$. Since pairing double brackets with pointy brackets is terribly ugly we will tend to drop the pointy brackets in the remainder. But the reader is encouraged to periodically meditate on the fact that they're implicitly present. At any rate, if H is a De Morgan monoid hyperdoctrine and e_n is the identity element in $H(T_n)$, then $[\![-]\!]$ is determined as follows:

- $[\![\tau = \sigma \mid X]\!] = H_{\langle \tau\sigma \mid X \rangle}(\Sigma_{\langle xx \mid x \rangle}(e_1))$.
- $[\![\mathsf{t} \mid X]\!] = e_{\mathrm{card}(X)}$.
- $[\![\sim \phi \mid X]\!] = \neg [\![\phi \mid X]\!]$
- $[\![\phi \wedge \psi \mid X]\!] = [\![\phi \mid X]\!] \sqcap [\![\psi \mid X]\!]$
- $[\![\phi \vee \psi \mid X]\!] = [\![\phi \mid X]\!] \sqcup [\![\psi \mid X]\!]$
- $[\![\phi \to \psi \mid X]\!] = \neg\left([\![\phi \mid X]\!] \boxtimes \neg [\![\psi \mid X]\!]\right)$
- $[\![\forall x_{n+1} \phi \mid X_1^n]\!] = \Pi_{\langle \overline{X_1^n} \mid X_1^{n+1} \rangle} [\![\phi \mid X_1^{n+1}]\!]$

Definition 8.4. We say that $\langle \phi \mid X \rangle$ is valid in H when $e_{\mathrm{card}(X)} \leq [\![\phi \mid X]\!]$ and is valid simpliciter when it is valid in H for every H.

Theorem 8.5. *If $\langle \phi \mid X \rangle$ is valid, then $\vdash \langle \phi \mid X \rangle$.*

Proof. Notice that $\vdash \langle \phi \mid X \rangle$ iff $[\mathsf{t} \mid X] \leq [\phi \mid X]$. Thus if $\nvdash \langle \phi \mid X \rangle$, then $\langle \phi \mid X \rangle$ is not valid in the canonical model, and thus not valid. Contraposing gives the result. □

Finally, before proving soundness we state the following lemma:

Lemma 8.6. *For any De Morgan monoid hyperdoctrine H, whenever it all makes sense, $H_{\langle \tau \mid X \rangle} [\![\phi \mid Y]\!] = [\![S_{\langle \tau \mid X \rangle} \langle \phi \mid Y \rangle]\!]$.*

Theorem 8.7. *If $\vdash \langle \phi \mid X \rangle$, then $\langle \phi \mid X \rangle$ is valid.*

Proof. By induction on the length of the proof witnessing $\vdash \langle \phi \mid X \rangle$. We examine only A13, A16, and A17. The reader interested in working out the details for the other axioms and rules may want to consult [18] or [23] for hints.

[A13] To start, do a bit of fiddling to see that $e_{\text{card}(X)} \leq [\![\forall x_{n+1}\phi \to \phi(x_{n+1}/y) \mid X_1^n]\!]$ iff $H_{\langle \overline{X_1^n, y} \mid X_1^n \rangle}[\![\forall x_{n+1}\phi \mid X_1^{n+1}]\!] \leq H_{\langle \overline{X_1^n, y} \mid X_1^n \rangle}[\![\phi \mid X_1^{n+1}]\!]$. Next, notice that by adjointness, $H_{\langle \overline{X_1^n} \mid X_1^{n+1} \rangle}\Pi_{\langle \overline{X_1^n} \mid X_1^{n+1} \rangle}[\![\phi \mid X_1^{n+1}]\!] \leq [\![\phi \mid X_1^{n+1}]\!]$. On the other hand we also have that

$$\begin{aligned} H_{\langle \overline{X_1^n} \mid X_1^{n+1} \rangle}\Pi_{\langle \overline{X_1^n} \mid X_1^{n+1} \rangle}[\![\phi \mid X_1^{n+1}]\!] &= H_{\langle \overline{X_1^n} \mid X_1^{n+1} \rangle}[\![\forall x_{n+1}\phi \mid X_1^n]\!] \\ &= [\![\forall x_{n+1}\phi \mid X_1^{n+1}]\!] \end{aligned}$$

And from here we simply have to apply $H_{\langle \overline{X_1^n, y} \mid X_1^n \rangle}$ to both sides to get what we want.

[A16] Notice that $\langle \tau\tau \mid X_1^n \rangle = \langle x_1 x_1 \mid x_1 \rangle \circ \langle \tau \mid X_1^n \rangle$. Thus we can compute as follows:

$$\begin{aligned} [\![\tau = \tau \mid X_1^n]\!] &= H_{\langle \tau\tau \mid X_1^n \rangle}(\Sigma_{\langle x_1 x_1 \mid x_1 \rangle}(e_1)) \\ &= H_{\langle \tau \mid X_1^n \rangle}(H_{\langle x_1 x_1 \mid x_1 \rangle}(\Sigma_{\langle x_1 x_1 \mid x_1 \rangle}(e_1))) \end{aligned}$$

Thus $e_n \leq [\![\tau = \tau \mid X_1^n]\!]$ since (a) $\Sigma_{\langle x_1 x_1 \mid x_1 \rangle}$ is left adjoint to $H_{\langle x_1 x_1 \mid x_1 \rangle}$ and (b) $H_{\langle \tau \mid X_1^n \rangle}$ is a De Morgan monoid homomorphism.

While we're here, it's worth noting that the converse inequality holds as well. To see this, note that since $e_1 \leq H_{\langle x_1 x_1 \mid x_1 \rangle}(e_2)$, adjointness gives $\Sigma_{\langle x_1 x_1 \mid x_1 \rangle}(e_1) \leq e_2$. Adjointness also gives that

$$\Sigma_{\langle \tau\tau \mid X_1^n \rangle}(H_{\langle \tau\tau \mid X_1^n \rangle}(\Sigma_{\langle x_1 x_1 \mid x_1 \rangle}(e_1))) \leq \Sigma_{\langle x_1 x_1 \mid x_1 \rangle}(e_1).$$

Thus we have $\Sigma_{\langle \tau\tau \mid X_1^n \rangle}(H_{\langle \tau\tau \mid X_1^n \rangle}(\Sigma_{\langle x_1 x_1 \mid x_1 \rangle}(e_1))) \leq e_2$, which is to say that $\Sigma_{\langle \tau\tau \mid X_1^n \rangle}[\![\tau = \tau \mid X_1^n]\!] \leq e_2$. One last application of adjointness then gives $[\![\tau = \tau \mid X_1^n]\!] \leq H_{\langle \tau\tau \mid X_1^n \rangle}(e_2) = e_n$. Together with the previous result we thus have that $[\![\tau = \tau \mid X_1^n]\!] = e_n$, which we will need in the sequel.

[A17] To begin, note that for any $h_n \in H(T_n)$, we have

$$\begin{aligned} H_{\langle \overline{X_{n+1}^{2n}} \mid X_1^{2n} \rangle}(h_n) \boxtimes \Sigma_{\langle \overline{X_1^n X_1^n} \mid X_1^n \rangle}(e_n) &= \Sigma_{\langle \overline{X_1^n X_1^n} \mid X_1^n \rangle}(H_{\langle \overline{X_1^n X_1^n} \mid X_1^n \rangle}(H_{\langle \overline{X_{n+1}^{2n}} \mid X_1^{2n} \rangle}(h_n)) \boxtimes e_n) \\ &= \Sigma_{\langle \overline{X_1^n X_1^n} \mid X_1^n \rangle}(h_n) \\ &= \Sigma_{\langle \overline{X_1^n X_1^n} \mid X_1^n \rangle}(H_{\langle \overline{X_1^n X_1^n} \mid X_1^n \rangle}(H_{\langle \overline{X_1^n} \mid X_1^{2n} \rangle}(h_n))) \\ &\leq H_{\langle \overline{X_1^n} \mid X_1^{2n} \rangle}(h_n) \end{aligned}$$

Where the first equality follows from Frobenius reciprocity and the inequality comes from adjointness. After commuting the first term in the resulting inequality, we continue the computation as follows:

$$\Sigma_{\langle\overline{X_1^n X_1^n}|X_1^n\rangle}(e_n) \boxtimes H_{\langle\overline{X_{n+1}^{2n}}|X_1^{2n}\rangle}(h_n) \leq H_{\langle\overline{X_1^n}|X_1^{2n}\rangle}(h_n)$$

so

$$H_{\langle\sigma\overline{X_2^n}\tau\overline{X_2^n}|X_1^n\rangle}(\Sigma_{\langle\overline{X_1^n X_1^n}|X_1^n\rangle}(e_n)) \boxtimes H_{\langle\sigma\overline{X_2^n}\tau\overline{X_2^n}|X_1^n\rangle}(H_{\langle\overline{X_{n+1}^{2n}}|X_1^{2n}\rangle}(h_n)) \leq H_{\langle\sigma\overline{X_2^n}\tau\overline{X_2^n}|X_1^n\rangle}(H_{\langle\overline{X_1^n}|X_1^{2n}\rangle}(h_n))$$

so

$$H_{\langle\sigma\overline{X_2^n}\tau\overline{X_2^n}|X_1^n\rangle}(\Sigma_{\langle\overline{X_1^n X_1^n}|X_1^n\rangle}(e_n)) \boxtimes H_{\langle\tau\overline{X_2^n}|X_1^n\rangle}(h_n) \leq H_{\langle\sigma\overline{X_2^n}|X_1^n\rangle}(h_n)$$

From here, observe first that

$$\begin{aligned}\Sigma_{\langle\overline{X_1^n X_1^n}|X_1^n\rangle}(e_n) &= \Sigma_{\langle\overline{X_1^n X_1^n}|X_1^n\rangle}[\![\mathsf{t} \mid X_1^n]\!] \\ &= [\![\overline{X_1^n} = \overline{X_{n+1}^{2n}} \mid X_1^{2n}]\!]\end{aligned}$$

With the second equality following by an argument similar to the one immediately before §7.1. Thus we have that

$$\begin{aligned}H_{\langle\sigma\overline{X_2^n}\tau\overline{X_2^n}|X_1^n\rangle}(\Sigma_{\langle\overline{X_1^n X_1^n}|X_1^n\rangle}(e_n)) &= H_{\langle\sigma\overline{X_2^n}\tau\overline{X_2^n}|X_1^n\rangle}([\![\overline{X_1^n} = \overline{X_{n+1}^{2n}} \mid X_1^{2n}]\!]) \\ &= [\![\sigma\overline{X_2^n} = \tau\overline{X_2^n} \mid X_1^n]\!] \\ &= [\![\sigma = \tau \otimes x_2 = x_2 \otimes \cdots \otimes x_n = x_n \mid X_1^n]\!] \\ &= [\![\sigma = \tau \mid X_1^n]\!]\end{aligned}$$

With the second equality following from Lemma 8.6 and the fourth from the just-verified fact that $[\![x_i = x_i \mid X_1^n]\!] = e_n$. So for any $h_n \in H(T_n)$ we have that

$$[\![\sigma = \tau \mid X_1^n]\!] \boxtimes H_{\langle\tau\overline{X_2^n}|X_1^n\rangle}(h_n) \leq H_{\langle\sigma\overline{X_2^n}|X_1^n\rangle}(h_n)$$

Thus, in particular, if $\phi \mid X_1^n$ is well formed, then we have

$$[\![\sigma = \tau \mid X_1^n]\!] \boxtimes H_{\langle\tau\overline{X_2^n}|X_1^n\rangle}[\![\phi \mid X_1^n]\!] \leq H_{\langle\sigma\overline{X_2^n}|X_1^n\rangle}[\![\phi \mid X_1^n]\!]$$

From which it follows that $e_n \leq [\![(\sigma = \tau \otimes \phi(x_1/\tau)) \to \phi(x_1/\sigma) \mid X_1^n]\!]$.[19] □

[19] If you squint at this proof, you will see that the system we've defined 'wants', in some sense, for the induced notion of identity at the type T_n to be a notion of identity for n-ary terms. We think this is an interesting-enough-for-a-footnote observation, hence this footnote.

9 Arithmetic

What we have, at the end of the day, is a minimal theory of arithmetic for relevance logic, with a semantics smoothly generalizing the algebraic semantics, via De Morgan monoids, for propositional relevance logic. How minimal is minimal? Let's take a moment to consider this question. We won't attempt to be maximally precise here, but only to give a general sense of how things stand.

Our theory makes no effort to capture standard features of the successor, such as $0 \neq 1$. Indeed, one can construct a relevant hyperdoctrine using the Boolean algebras induced by a classical one-element model of our vocabulary to witness that $\langle 0 \neq 1 \mid \varnothing \rangle$ is not valid. So, we don't build in even some of the basic axioms of Robinson arithmetic.

On the other hand, given a unary term $\sigma \mid X$, we can have the formula $\langle (\sigma = y \otimes \sigma = z) \to y = z \mid X, y, z \rangle$ as an instance of A17 (setting ϕ to $v = z$). It follows fairly directly that equations where a pr term flanks a free variable define provably total functions; since for each primitive recursive function f we have a pr term such that

$$\vdash \langle \mathsf{pr}_1[\sigma \mid X; \tau \mid Y](\underline{n_0}, \ldots, \underline{n_i}) = \underline{m} \rangle \qquad \text{if and only if} \qquad f(n_0, \ldots, n_i) = m$$

there is a sense in which our theory has more provably total functions than Robinson arithmetic—it has at least as many as $\mathrm{I}\Sigma^0_1$. However, we clearly fall short of $\mathrm{I}\Sigma^0_1$ in other ways. In particular, we do not have any meaningful induction. For example,

$$\langle \forall x \forall y \forall z ((x + y) + z = x + (y + z)) \mid x, y, z \rangle$$

(with the $+$ given by the appropriate pr term) is not valid in the syntactic hyperdoctrine, although the analogous statement is provable even in $\mathrm{I}\Delta^0_0$.

So, "minimal" is apparently, morally speaking, subtly incommensurable with Robinson's arithmetic but below $\mathrm{I}\Sigma^0_1$. Our theory's nearest relative might be something like the system PRE ("primitive recursive equations") described in Friedman's [6]. PRE represents primitive recursion via terms, as we do, although the background logic is classical and the representation takes recursion equations as axioms rather than using a rewrite system.

This analogy suggests a conjecture. Friedman proves that $\mathsf{PRE} + 0 \neq 1 + x \neq 0 \to \exists y(S(y) = x)$, when closed under an induction rule of the form

$$\frac{F(0) = G(0) \qquad F(x) = G(x) \to F(S(x)) = G(S(x))}{F(x) = G(x)}$$

has the strength of full PRA (that is, one recovers induction for open quantifier-free formulas). A very slight weakening of Friedman's induction rule is Goodstein's induction rule:

$$\frac{F(0) = G(0) \qquad H(x, F(x)) = F(S(x)) \qquad H(x, G(x)) = G(S(x))}{F(x) = G(x)}$$

and, Goodstein's rule figures prominently in the axiomatization of equational PRA (see [9] for details). But as noted above (see footnote 5), adjusting our type theory to make T_1 a full PNNO, rather than a weak PNNO (with corresponding modifications to our rewrite rules) would underwrite a version of Goodstein's rule. So the conjecture is this: adjusting our theory to incorporate a full PNNO at the level of types will result in the validity of an axiom of (quantifier free) induction.

A more ambitious conjecture might even go as follows. Tweaking our theory so that the category of types and terms is Cartesian closed and contains a natural numbers object would yield a term rewriting system closely analogous to Gödel's system T.[20] Given the close relationship between system T and arithmetic at the level of PA and HA, we might hope in this context to recover full arithmetical induction. Continued exploration in this direction, with richer and richer categories of types and terms in the background, is the program that we sketched in the introduction of this essay, leading us from the present $R^{\frac{1}{2}\#}$ to something that could play the role of a genuine $R^{\#\frac{1}{2}}$.

The extremely broad picture here is something like this. One can imagine a whole program of non-classical arithmetic, built on two pillars. The first pillar would be a conception of consequence—no matter how non-classical—as a matter of the structure of fibers in a hyperdoctrine. The second pillar would be a conception of arithmetical computation as properly represented by rewrite rules at the level of syntax, with various levels of arithmetical strength (including perhaps various forms of induction) given by the various possible levels of structure in an underlying syntactic category of types and terms. Such a program would cleanly factorize the contributions to our overall arithmetic of our logical principles, on the one hand, and of our computational resources, on the other.

In the above we've found one illustration (inspired by Meyer's remarks) of how such a program might be carried out. We hope more will be forthcoming.

[20] For details, see [1]. For the original presentation, see [8].

References

[1] J. Avigad and S. Feferman. Gödel's functional ("Dialectica") interpretation. In S. Buss, editor, *Handbook of Proof Theory*, pages 337–405. Elsevier, 1999.

[2] R. T. Brady. Natural deduction systems for some quantified relevant logics. *Logique et Analyse*, 27(108):355–378, 1984.

[3] J. M. Dunn. *The Algebra of Intensional Logics*. PhD thesis, University of Pittsburgh, 1966.

[4] J. M. Dunn. Relevance logic and entailment. In *Handbook of Philosophical Logic*, pages 117–224. Springer, 1986.

[5] J. M. Dunn. Relevant predication 1: The formal theory. *Journal of Philosophical Logic*, 16(4):347–381, 1987.

[6] H. Friedman. Provable equality in primitive recursive arithmetic with and without induction. *Pacific Journal of Mathematics*, 57(2), 1975.

[7] H. Friedman and R. K. Meyer. Whither relevant arithmetic? *Journal of Symbolic Logic*, pages 824–831, 1992.

[8] K. Gödel. *Über eine bisher noch nicht benützte Erweiterung des finiten Standpunktes. Dialectica*, pages 280–287, 1958.

[9] R. Goodstein. *Recursive Number Theory.* North Holland, 1957.

[10] P. Kremer. Relevant predication: Grammatical characterisations. *Journal of Philosophical Logic*, 18(4):349–382, 1989.

[11] P. Kremer. *Real Properties, Relevance Logic, and Identity.* PhD thesis, University of Pittsburgh, 1994.

[12] P. Kremer. Dunn's relevant predication, real properties and identity. *Erkenntnis*, 47(1):37–65, 1997.

[13] P. Kremer. Relevant identity. *Journal of Philosophical Logic*, 28(2):199–222, 1999.

[14] J. Lambek and P. J. Scott. *Introduction to Higher-Order Categorical Logic*. Cambridge University Press, 1988.

[15] F. W. Lawvere. An elementary theory of the category of sets. *Proceedings of the National Academy of Sciences of the United States of America*, 52(6):1506, 1964.

[16] K. Manders. The Euclidean diagram. In P. Mancosu, editor, *The Philosophy of Mathematical Practice*, pages 80–133. Oxford University Press, 2008.

[17] E. D. Mares. Semantics for relevance logic with identity. *Studia Logica*, 51(1):1–20, 1992.

[18] Y. Maruyama. Full Lambek hyperdoctrine: Categorical semantics for first-order substructural logics. In *International Workshop on Logic, Language, Information, and Computation*, pages 211–225. Springer, 2013.

[19] T. Moraschini, J. G. Raftery, and J. Wannenburg. Varieties of De Morgan monoids: Covers of atoms. *The Review of Symbolic Logic*, pages 1–37, 2018.

[20] T. Moraschini, J. G. Raftery, and J. Wannenburg. Varieties of De Morgan monoids: Minimality and irreducible algebras. *Journal of Pure and Applied Algebra*, 223(7):2780–2803, 2019.

[21] I. Németi and H. Andréka. Can general relativistic computers break the Turing barrier? In *Conference on Computability in Europe*, pages 398–412. Springer, 2006.

[22] L. Román. Cartesian categories with natural numbers object. *Journal of Pure and Applied Algebra*, 58(3):267–278, 1989.

[23] R. A. Seely. Hyperdoctrines, natural deduction and the Beck condition. *Mathematical Logic Quarterly*, 29(10):505–542, 1983.

[24] A. Urquhart. The undecidability of entailment and relevant implication. *Journal of Symbolic Logic*, 49(4):1059–1073, 1984.

[25] A. Urquhart. Duality for algebras of relevant logics. *Studia Logica*, 56(1-2):263–276, 1996.