Hyperdoctrine Semantics: An Invitation

SHAY LOGAN AND GRAHAM LEACH-KROUSE[1]

Abstract: Categorial logic, as its name suggests, applies the techniques and machinery of category theory to topics traditionally classified as part of logic. We claim that these tools deserve attention from a greater range of philosophers than just the mathematical logicians. We support this claim with an example. In this paper we show how one particular tool from categorial logic—hyperdoctrines—suggests interesting metaphysics. Hyperdoctrines can provide semantics for quantified languages, but this account of quantification suggests a metaphysical picture quite different from the one suggested by standard model-theoretic semantics.

Keywords: Hyperdoctrines, Categorical Logic, Metaphysics, Semantics

In this paper, we wish to suggest that a tool from category theory, and in particular categorial logic—the theory of hyperdoctrines—is of metaphysical interest. It presents an alternative to a viewpoint that has become entrenched (at least in some circles) to the point of invisibility. The first three sections of our paper are a crash-course in hyperdoctrine semantics for classical first-order logic. The final section argues that a focus on first-order model theory has distorted many philosopher's metaphysical theorizing, and uses the results of the first three sections to sketch an alternative.

1 Language and Logic

We call the language we work with throughout this paper $\mathcal{L}$. Each well-formed expression in $\mathcal{L}$ has the form '$\phi \mid X$' with ϕ a sequence of symbols called the *untyped part* of the expression and X a set of variables called the *typing part* of the expression. Philosophically, we understand the typing part of an $\mathcal{L}$-expression to specify something like the 'dimensions' along which the untyped part is taken to be incomplete.

[1]Many thanks to the audience at Logica, as well as to Andrew Tedder, Teresa Kouri, Eileen Nutting, Blane Worley, and an anonymous referee, who helped us sharpen and refine the ideas that follow.

The typing part can make a difference to the meaning of an expression even if not all of its variables appear in the untyped part. As an aid to understanding, consider a polynomial like $x - y$. The set of points at which this polynomial is zero can be viewed as a line in a two dimensional space if the only variables under consideration are x and y. But it can also be viewed as a plane in a three dimensional space when the variable z is in play. In $\mathcal{L}$, context matters: the zero locus of $x - y$ in the $\{x, y\}$-context is a line; the zero locus of $x - y$ in the $\{x, y, z\}$-context is a plane.

Formally, the terms of our language are constructed from a vocabulary consisting of variables $x_1, x_2, \ldots$, constants $c_1, c_2, \ldots$, the separator symbol '$|$', and the brackets '$\{$' and '$\}$', which we will tend to drop. We will regard a sequence τ of variables and constants with a shared context X as a single term $\tau \mid X$. If there are n terms in τ, we will say that τ is n-ary.

Definition 1 (Terms of $\mathcal{L}$)**.** If c is a constant, then $c \mid \varnothing$ is a unary term. If x is a variable, then $x \mid \{x\}$ is a unary term. If $\tau \mid X$ is an n-ary term and $\sigma \mid Y$ is an m-ary term, then $\tau\sigma \mid X \cup Y$ is an $n + m$-ary term. If $\tau \mid X$ is an n-ary term and y is a variable, then $\tau \mid X \cup \{y\}$ is an n-ary term.

For each n we recognize n-adic predicates $R_n^1, R_n^2, \ldots$. For concreteness, we will recognize three connectives ($\neg$, $\wedge$, and $\rightarrow$), one quantifier ($\forall$), and take '$\exists$' and '$\vee$' to be defined. We specify the set of formulas as follows:

Definition 2 (Formulas of $\mathcal{L}$)**.** If R is an n-adic predicate and $\tau \mid X$ is an n-ary term, then $R\tau \mid X$ is a formula. If $\phi \mid X$ and $\psi \mid X$ are formulas, then so are $\neg\phi \mid X$, $(\phi \wedge \psi) \mid X$, and $(\phi \rightarrow \psi) \mid X$. If $\phi \mid X$ is a formula and y is a variable, then $\phi \mid X \cup \{y\}$ is a formula. Last, if $\phi \mid X$ is a formula and $x \in X$, then $\forall x\phi \mid X - \{x\}$ is a formula.

We adopt the usual conventions regarding outermost parentheses and similar matters. To indicate substitutions, we declare that if τ, σ, and η are constants or variables, then $\tau(\sigma/\eta)$ is η if $\tau = \sigma$, and otherwise τ. We read this as "replace σ with η". If $\overline{\sigma}$ and $\overline{\eta}$ are sequences of constants or variables, then $(\overline{\sigma}/\overline{\eta})$ abbreviates a simultaneous replacement of each σ_i with η_i. We suppose that this is done in some way that avoids collision.

If ϕ is the untyped part of a formula, $\phi(\sigma/\eta)$ is the result of replacing each free occurrence of a constant or variable t in ϕ by an occurrence of $t(\sigma/\eta)$. If X is a set of variables, then $\overline{X}$ abbreviates the sequence of those variables taken in increasing order (by their subscript indices). If X and Y are sets of variables with $\text{card}(X) = \text{card}(Y)$ then we call a replacement of the form $\phi(\overline{Y}/\overline{X})$ a change of variables. Finally, we say that $\phi(\overline{\sigma}/\overline{\eta})$ is a *proper* substitution instance of ϕ if each η_i is freely substitutable for σ_i in ϕ.

We will write **K** for the subset of $\mathcal{L}$ that is, apart from typing, plain-old classical logic. More to the point, we say that if ϕ is a theorem of classical logic and $\phi \mid X$ is well formed, then $\phi \mid X \in \mathbf{K}$. But our main interest in what follows is not actually in **K** itself, but in the notion of **K**-provability. For the latter, we restrict to the special case of single-premise provability.[2] We will write $\phi \mid X \vdash_{\mathbf{K}} \psi \mid Y$ to mean that $\psi \mid Y$ is **K**-provable from $\phi \mid X$, and we define this relation as follows:

Definition 3. $\phi \mid X \vdash_{\mathbf{K}} \psi \mid Y$ iff there is a sequence of formulas $\psi_1 \mid X_1, \psi_2 \mid X_2, \ldots, \psi_n \mid X_n$ with $\psi_n \mid X_n = \psi \mid Y$ such that for all $1 \leq i \leq n$, either $\psi_i \mid X_i = \phi \mid X$, or for some $j < i$, $X_j = X_i$ and $\psi_j \rightarrow \psi_i \mid X_j \in \mathbf{K}$, or for some $j < i$ and $k < i$, $X_i = X_j = X_k$ and $\psi_i = \psi_j \wedge \psi_k$, or for some $j < i$, $\psi_i = \psi_j(\overline{X_j}/\overline{X_i})$ is a proper substitution instance of ψ_j.

Note that we allow a proper change of variables in the course of a proof. Why? Consider $x = a \mid x$ and $y = a \mid y$.[3] $x = a$, regarded as incomplete only along the x-dimension, defines the same property (intuitively the property of being identical to a) as $y = a$ does when regarded as incomplete only along the y-dimension. So we ought to adopt a mechanism that lets us count $x = a \mid x$ as expressing the same thing as $y = a \mid y$. To accomplish this, it's clear we ought to adopt *some sort* of variable-substitution policy.

Not just any policy will do, though. Changes of variables in our technical sense are always *monotone*. To see why this must be, consider the formulas $x \leq y \mid x, y$ and $y \leq x \mid x, y$. Each of these formulas defines the less-than-or-equal-to relation. Yet we wouldn't want to regard them as equivalent—if they were equivalent, then conjoining them wouldn't give us anything new. But of course it does: $x \leq y \wedge y \leq x \mid x, y$ defines the identity relation, which neither $x \leq y \mid x, y$ nor $y \leq x \mid x, y$, taken on its own, does.

What goes wrong is that $x \leq y \mid x, y$ and $y \leq x \mid x, y$ are true of different sets of tuples. If the first is true of $1, 2$ (in that order), then the second is true of $2, 1$ (in *that* order). Now, recall that our variables come equipped with an ordering: the one imposed on them by their subscripts. This imposes an ordering on the dimensions of incompleteness of a formula expressed in

[2]The extension to multipremise provability is straightforward, provided all the premises are required to share a context.

[3]Neither of these is in fact a formula in $\mathcal{L}$. Don't get hung up on this; just pretend for a moment that we have a language with identity (and, in a moment, with inequalities).

those variables. Formulas can be regarded as being "true of" a sequence of things if the result of supplying the members of the sequence to the dimensions of incompleteness *in order* is true. That is why our extended notion of provability only allows variable changes that are *monotone*: monotone changes preserve the feature of tracking which tuples a formula is 'true of'.

Now that you know why we've defined **K**-provability the way we have, we end the section by noting a few important facts about this relation:

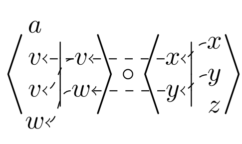
Lemma 1. If $\phi \mid X \vdash_{\mathbf{K}} \psi \mid Y$, then $\phi(\overline{X}/\overline{Y}) \to \psi \mid Y \in \mathbf{K}$.

Lemma 2. If $\phi \mid X \vdash_{\mathbf{K}} \psi \mid Y$, then $\phi \mid X \cup \{z\} \vdash_{\mathbf{K}} \psi \mid Y \cup \{z\}$.

It turns out that there is a natural way to view (equivalence classes of) terms of $\mathcal{L}$ as arrows in a category. Our goal at the moment is to very concretely describe this category.

To help keep concepts clearly delimited, we will write $\langle \tau \mid X \rangle$ for the arrow associated with the term $\tau \mid X$. As a preview of what's to come, we offer the following summary: the category we are constructing has for its objects the 'types' T_n, where n is a natural number. Each arrow $\langle \tau \mid X \rangle$ will have domain $T_{\text{card}(X)}$ and codomain $T_{\text{len}(\tau)}$. It follows that $\langle \tau_2 \mid X_2 \rangle \circ \langle \tau_1 \mid X_1 \rangle$ is defined just when $\text{len}(\tau_1) = \text{card}(X_2)$. Composition in $\mathcal{B}$ is just careful substitution. That is, the composition $\langle \tau_2 \mid X_2 \rangle \circ \langle \tau_1 \mid X_1 \rangle$ is formed by substituting the symbols constituting τ_1 for the variables occurring in X_2.

To say a bit more, it helps to first look at an example: consider $\langle xy \mid x, y, z \rangle : T_3 \longrightarrow T_2$ and $\langle avvw \mid v, w \rangle : T_2 \longrightarrow T_4$. The composition $\langle avvw \mid v, w \rangle \circ \langle xy \mid x, y, z \rangle$ should be an arrow $T_3 \longrightarrow T_4$. Here's how to make this happen: pair xy in the untyped part of $\langle xy \mid x, y, z \rangle$ with v, w in the typing part of $\langle avvw \mid v, w \rangle$. Use that pairing to replace the symbols of $avvw$, creating a new untyped part $axxy$ compatible with the typing part x, y, z. Or, in a picture, composition works like this in the case at hand:



Either way we describe it, the result is the same:

$$\langle avvw \mid v, w \rangle \circ \langle xy \mid x, y, z \rangle = \langle axxy \mid x, y, z \rangle$$

More generally:

Definition 4. if $\langle \tau_1 \mid X_1 \rangle$ and $\langle \tau_2 \mid X_2 \rangle$ are composable, then

$$\langle \tau_2 \mid X_2 \rangle \circ \langle \tau_1 \mid X_1 \rangle := \langle \tau_2(\overline{X_2}/\tau_1) \mid X_1 \rangle$$

That is to say, replace the variables X_2 (taken in order of subscripts) occurring in τ_2 with the symbols constituting τ_1, and view the result as a term in the context X_1.

There's a bit of a problem here, however: if we associate each term with a unique arrow, the composition above doesn't give us a category. Recall that to be a category, each object must have a unique identity arrow. The natural candidate for the identity arrow $T_2 \longrightarrow T_2$, for example, is something of the form $\langle x_1 x_2 \mid x_1, x_2 \rangle$. But it's equally natural to consider $\langle x_3 x_4 \mid x_3, x_4 \rangle$. More generally, the only plausible candidates for identity arrows at T_n are terms of the form $\langle \overline{X} \mid X \rangle$. All such arrows are in fact *left* identities. That is, we have for example that

$$\langle x_1 x_2 \mid x_1, x_2 \rangle \circ \langle a x_3 \mid x_3 \rangle = \langle a x_3 \mid x_3 \rangle$$

But the fact that *all* of these arrows are left identities immediately entails that none of them are right identities. Again this is easy to see by examining the following simple example:

$$\langle x_3 x_4 \mid x_3, x_4 \rangle \circ \langle x_1 x_2 \mid x_1, x_2 \rangle = \langle x_1 x_2 \mid x_1, x_2 \rangle$$

Luckily, this example suggests how to correct the problem. Given the definition of composition, it's clear that applying any of the candidate identity arrows on the right of a term is exactly the same thing as applying a change of variables. Thus, for example, $\langle \tau \mid X \rangle \circ \langle \overline{Y} \mid Y \rangle = \langle \tau(\overline{X}/\overline{Y}) \mid Y \rangle$. So, rather than taking arrows to be terms *simpliciter*, we take arrows to be equivalence classes of terms with equivalence being given by monotone change of variables. More explicitly, we say that $\langle \tau \mid X \rangle$ and $\langle \sigma \mid Y \rangle$ are equivalent when $\langle \sigma \mid Y \rangle = \langle \tau(\overline{X}/\overline{Y}) \mid Y \rangle$.

Lemma 3. Composition as defined in Definition 4 is well-defined on equivalence classes of terms; furthermore, the types, maps, and identities above, taken with this composition, constitute a category.

Lemma 4. If $\langle \tau \mid Y \rangle : T_m \longrightarrow T_n$, then for some σ, $\langle \tau \mid Y \rangle = \langle \sigma \mid x_1, \ldots, x_n \rangle$.

The set $\{x_1, \ldots, x_n\}$ will play a large role in the remainder. Accordingly, we abbreviate $\{x_1, \ldots, x_n\}$ as X_1^n.

Definition 5 (The Base Category)**.** $\mathcal{B}$ is the category that has the types T_n for objects and that has equivalence classes of terms as arrows.

Definition 6. For each n,

1. $\mathcal{L}(T_n)$ is the set of formulas $\phi \mid X$ with $\text{card}(X) = n$. We say that such formulas have type T_n.

2. If $\phi \mid X$ and $\psi \mid Y$ are both in $\mathcal{L}(T_n)$, then we say $\phi \mid X$ and $\psi \mid Y$ are *equivalent in **K*** (written: $\phi \mid X \cong_{\mathbf{K}} \psi \mid Y$) when both $\phi \mid X \vdash_{\mathbf{K}} \psi \mid Y$ and $\psi \mid Y \vdash_{\mathbf{K}} \phi \mid X$.

3. For $\phi \mid X \in \mathcal{L}(T_n)$, we write $[\phi \mid X]_{\mathbf{K}}$ for the $\cong_{\mathbf{K}}$-equivalence class of $\phi \mid X$.

4. We write $\mathbf{S}(T_n)$ for the poset with underlying set containing the classes $[\phi \mid X]$ with $\text{card}(X) = n$ and with $[\phi \mid X] \leq^n_{\mathbf{K}} [\psi \mid Y]$ iff $\phi \mid X \vdash_{\mathbf{K}} \psi \mid Y$.

5. For each n, we define the unary operation $'^n$ on $\mathbf{S}(T_n)$ by setting $[\phi \mid X]'^n = [\neg\phi \mid X]$.

We leave it to the reader to check that $\cong_{\mathbf{K}}$ is in fact an equivalence relation. When they can be inferred from context (and they essentially always can) we omit most of the superscripts and subscripts.

Lemma 5. $'$ and $\leq$ are well-defined

Lemma 6. Every member of $\mathbf{S}(T_n)$ can be written in the form $[\phi \mid x_1 \ldots x_n]$.

Lemma 7. For all n, the structure $\langle \mathbf{S}(T_n), \leq, ' \rangle$ is a Boolean algebra.

Lemma 8. $[\phi \mid X] = \top_{\text{card}(X)}$ iff $\phi \mid X \in \mathbf{K}$, where $\top_n$ is the supremum of the Boolean algebra $\langle \mathbf{S}(T_n), \leq, ' \rangle$.

2 Limning the Remaining Structure

We've now seen that $\mathcal{B}$'s objects naturally correspond to certain algebras of formulas. As you might expect, it turns out that $\mathcal{B}$'s arrows naturally correspond to homomorphisms of these algebras. To get an intuition for what this is going to look like, consider a formula like $Rx_1x_2 \mid x_1, x_2, x_3$. Our typing conventions demand that we view this formula as incomplete along dimensions x_1, x_2 and x_3. Picture these as three 'slots' into which name-shaped things can be placed. Now notice that if $\tau \mid X$ is a term and $\mathrm{len}(\tau) = 3$, then the untyped part of the term, τ, just is a sequence of three name-shaped things—which is exactly what $Rx_1x_2 \mid x_1, x_2, x_3$ was looking for! Given such a term, we should be able to apply it to $Rx_1x_2 \mid x_1, x_2, x_3$ to get another formula by replacing each occurrence of x_1 by τ_1, x_2 by τ_2, and so on.

To capture all of this, we might try defining a map in the following way:

$$\begin{array}{rl} \text{Given} & \langle \tau \mid X \rangle : T_n \longrightarrow T_m, \\ \text{and} & \phi \mid Y \in \mathcal{L}(T_m), \\ \text{let} & \mathbf{S}(\tau \mid X) : \phi \mid Y \longmapsto \phi(\overline{Y}/\tau) \mid X \in \mathcal{L}(T_n). \end{array}$$

This isn't quite right, but it's close. Before pointing out the problems, we pause to note one important detail that this proposal *does* get right: $\mathbf{S}$, so-defined, is *contravariant* in the sense that it turns arrows $T_n \longrightarrow T_m$ into something going in the other direction, from $\mathcal{L}(T_m)$ to $\mathcal{L}(T_n)$.

The problem, which we point out before fixing, is that $\phi(\overline{Y}/\tau)$ may not be a proper substitution instance of ϕ. For example, consider the formula $\forall x_1 R x_1 x_2 \mid x_2 \in \mathcal{L}(T_1)$ and the term $x_1 \mid x_1 : T_1 \longrightarrow T_1$. Following the above construction, $\mathbf{S}(x_1 \mid x_1)(\forall x_1 R x_1 x_2 \mid x_2) = \forall x_1 R x_1 x_2 (x_2/x_1) \mid x_1 = \forall x_1 R x_1 x_1 \mid x_1$. But notice that before the substitution, the second place of the relation R was occupied by a free variable, and after the substitution it is occupied by a variable bound by the quantifier $\forall x_1$.

We solve this by recalling that the arrows of $\mathcal{B}$ are identified with *equivalence classes* of terms rather than with individual terms. Thus, instead of applying $\tau \mid X$ to $\phi \mid Y$, we can instead apply some equivalent term $\sigma \mid Z$ such that $\phi(\overline{Y}/\sigma)$ is a proper substitution instance of ϕ. It turns out there's a nice way to do this:

Definition 7. Let $\phi \mid Y \in \mathcal{L}(T_n)$ and $\langle \tau \mid X \rangle : T_m \longrightarrow T_n$. Let Z be a set of variables with $\mathrm{card}(X) = \mathrm{card}(Z)$ and such that no variable in Z occurs bound in ϕ. Let $\sigma = \tau(\overline{X}/\overline{Z})$. Then $\mathbf{S}\langle \tau \mid X \rangle(\phi \mid Y) = \phi(\overline{Y}/\sigma) \mid Z$.

Lemma 9. If $\langle \tau \mid X \rangle : T_m \longrightarrow T_n$, then $\mathbf{S}\langle \tau \mid X \rangle$ is a Boolean algebra homomorphism $\mathbf{S}(T_n) \longrightarrow \mathbf{S}(T_m)$. That is, $\mathbf{S}\langle \tau \mid X \rangle$ is well-defined on equivalence classes and commutes appropriately with the Boolean operations.

As a corollary to the Lemmas proved so far, we have the following:

Corollary 1 $\mathbf{S}$ *is a contravariant functor that maps each object of $\mathcal{B}$ to a Boolean algebra and each arrow of $\mathcal{B}$ to a Boolean algebra homomorphism.*

There's something a bit funny we need to deal with now. $\mathbf{S}$ does, in fact, map each object of $\mathcal{B}$ to the category of Boolean algebras and each arrow of $\mathcal{B}$ to a Boolean algebra homomorphism. But it's useful (for reasons that will be made clear below) to **not** think of $\mathbf{S}$ as a functor from $\mathcal{B}$ to the category of Boolean algebras and Boolean algebra homomorphisms (call this category $\mathbf{Bool}$. Instead, we will think of $\mathbf{S}$ as a functor from $\mathcal{B}$ to the category of Boolean algebras and *order-preserving* functions (call this category $\mathbf{BoolMon}$. Either way you look at it, the point to observe here is that the functor $\mathbf{S}$ arose very naturally from structure imposed on $\mathcal{L}$ by the relation of $\mathbf{K}$-provability. We now turn to showing that $\mathbf{K}$-provability not only imposes structure *within* $\mathbf{S}(T_n)$, but also *among* the algebras $\mathbf{S}(T_n)$ for different values of n.

Before we can observe this structure, however, we again need to introduce a bit of notation. To begin, note that for each variable y, and n-membered set of variables X with $y \notin X$ there is a term-arrow $\langle \overline{X} \mid X \cup \{y\}\rangle : T_{n+1} \longrightarrow T_n$. A natural (and, conveniently, correct) interpretation of term-arrows of the form $\langle \overline{X} \mid X \cup \{y\}\rangle$ is that they are projection onto all-but-one component. From here it's not hard to see the following:

Lemma 10. For $n > 0$ and $y \notin X$, there are exactly n equivalence classes of arrows of the form $\langle \overline{X} \mid X \cup \{y\}\rangle : T_n \longrightarrow T_{n-1}$; one for each component omitted.

Something like the common 'hat' notation to signal omissions is useful here. Usually, one writes $x_1 \ldots \widehat{x_j} \ldots x_n$ to indicate the sequence $x_1 \ldots x_n$, but with x_j omitted. We'll abbreviate further, however, and just write $\widehat{j}$ for this sequence. Thus $\widehat{j} \mid X_1^n$ is shorthand for the term $x_1 \ldots \widehat{x_j} \ldots x_n \mid x_1 \ldots x_n$. It follows that each equivalence class of arrows of the form $\langle \overline{X} \mid X \cup \{y\}\rangle$ has a unique representative of the form $\langle \widehat{j} \mid X_1^n\rangle$.

The maps $\mathbf{S}\langle \widehat{j} \mid X_1^n\rangle$ are very well behaved. To see this, first recall that by Lemma 6 every class in $\mathbf{S}(T_{n-1})$ has a representative with the form $\phi \mid X_1^n - x_j$. But now observe that $\mathbf{S}\langle \widehat{j} \mid X_1^n\rangle[\phi \mid X_1^n - x_j] = [\phi \mid X_1^n]$. Thus, $\mathbf{S}\langle \widehat{j} \mid X_1^n\rangle$ is essentially just the natural *inclusion function* $\mathbf{S}(T_{n-1}) \hookrightarrow \mathbf{S}(T_n)$.

There are equally natural functions going in the other direction:

Definition 8. Let Y be a set of $n > 0$ variables, and y_j be the jth member of $\overline{Y}$. Then $\Pi_j^n[\phi \mid Y] = [\forall y_j \phi \mid Y - y_j]$.

Lemma 11. Π_j^n is an order-preserving function $\mathbf{S}(T_n) \longrightarrow \mathbf{S}(T_{n-1})$

We emphasize that the various Π functions are *not*, in general, Boolean algebra homomorphisms. For example, $(\Pi_1^1[Rx_1 \mid x_1])' = [\neg\forall x_1 Rx_1 \mid \emptyset]$ while $\Pi_1^1([Rx_1 \mid x_1]') = [\forall x_1 \neg Rx_1 \mid \emptyset]$. This explains the bit of funny business mentioned after Corollary 1—the Π functions 'live' in the category of Boolean algebras and order-preserving functions, so if we want to 'see' these functions, it's best to view $\mathbf{S}$ as having this category as its codomain. In the remainder, we will write $\mathbf{BoolMon}$ for the category of Boolean algebras and order-preserving functions and $\mathbf{Bool}$ for the usual category of Boolean algebras and Boolean algebra homomorphisms.

Π_j^n and $\mathbf{S}\langle\widehat{j} \mid X_1^n\rangle$, as noted, point in opposite directions. As it turns out, they are related in a much more surprising way as well:

Lemma 12. Π_j^n is right adjoint to $\mathbf{S}\langle\widehat{j} \mid X_1^n\rangle$.

Proof. Without loss of generality, let $[\phi \mid X_1^n - x_j] \in \mathbf{S}(T_n)$ and let $[\psi \mid X_1^n] \in \mathbf{S}(T_n)$. Our goal is to show that

$$\mathbf{S}\langle\widehat{j} \mid X_1^n\rangle[\phi \mid X_1^n - x_j] \leq [\psi \mid X_1^n] \quad \Leftrightarrow \quad [\phi \mid X_1^n - x_j] \leq \Pi_j^n[\psi \mid X_1^n]$$

It suffices to show that whenever the statement is well formed, we have

$$\phi \mid X_1^n \vdash_{\mathbf{K}} \psi \mid X_1^n \quad \Leftrightarrow \quad \phi \mid X_1^n - x_j \vdash_{\mathbf{K}} \forall x_j \psi \mid X_1^n - x_j$$

For $\Rightarrow$, assume $\phi \mid X_1^n \vdash_{\mathbf{K}} \psi \mid X_1^n$. Then by Lemma 1, $\phi \to \psi \mid X_1^n \in \mathbf{K}$. Since $\phi \mid X_1^n - x_j$ is well-formed, x_j does not occur free in ϕ. Thus by classical logic, $\phi \to \forall x_j \psi \mid X_1^n - x_j \in \mathbf{K}$. So clearly $\phi \mid X_1^n - x_j \vdash_{\mathbf{K}} \forall x_j \psi \mid X_1^n - x_j$.

For $\Leftarrow$, assume $\phi \mid X_1^n - x_j \vdash_{\mathbf{K}} \forall x_j \psi \mid X_1^n - x_j$. Then by Lemma 2, $\phi \mid X_1^n \vdash_{\mathbf{K}} \forall x_j \psi \mid X_1^n$. Also, since it's clear that x_j is free for x_j in ψ, $\forall x_j \psi \to \psi \mid X_1^n \in \mathbf{K}$. Thus $\phi \mid X_1^n \vdash \psi \mid X_1^n$. $\square$

Note that since $\mathbf{S}\langle\widehat{j} \mid X_1^n\rangle$ is essentially an inclusion of $\mathbf{S}(T_{n-1})$ into $\mathbf{S}(T_n)$, Lemma 12 tells us, in a slogan, that universal quantifications are right adjoint to inclusions. It's worth noting that this result does *not* hold if we think of $\mathbf{S}\langle\widehat{j} \mid X_1^n\rangle$ as a functor whose domain is $\mathbf{Bool}$, for the simple reason that Π_j^n is not a functor whose *codomain* is $\mathbf{Bool}$.

As you might expect, there are dual results for existential quantification:

Definition 9. Let Y be a set of $n > 0$ variables, and y_j be the jth member of $\overline{Y}$. Then $\Sigma_j^n[\phi \mid Y] = [\exists y_j \phi \mid Y - y_j]$.

The proof of the next two lemmas are nice exercises that we encourage the reader to pursue.

Lemma 13. Σ^n_j is left adjoint to $\mathbf{S}\langle\widehat{j} \mid X^n_1\rangle$.

Lemma 14. Without loss of generality, let $[\psi \mid X^n_1] \in \mathbf{S}(T_n)$ and let $[\phi \mid X^n_1 - x_j] \in \mathbf{S}(T_{n-1})$. Then

$$\Pi^n_j(\mathbf{S}\langle\widehat{j} \mid X^n_1\rangle[\phi \mid X^n_1 - x_j] \sqcup [\psi \mid X^n_1]) \leq [\phi \mid X^n_1 - x_j] \sqcup \Pi^n_j[\psi \mid X^n_1]$$

The next relationship we mention is a bit more subtle, so we'll take a minute to flesh it out. To begin, consider the following diagram showing two different roads from the formula $Rx_1x_2x_3 \mid x_1, x_2, x_3$ to the formula $\forall x_2 Rax_2b \mid \emptyset$.

$$\begin{tikzcd}
Rx_1x_2x_3 \mid x_1, x_2, x_3 \arrow[r, "\Pi^3_2"] \arrow[d, "\mathbf{S}\langle ax_2b \mid x_2\rangle"'] & \forall x_2 Rx_1x_2x_3 \mid x_1, x_3 \arrow[d, "\mathbf{S}\langle ab \mid \emptyset\rangle"] \\
Rax_2b \mid x_2 \arrow[r, "\Pi^1_1"'] & \forall x_2 Rax_2b \mid \emptyset
\end{tikzcd}$$

It's clear enough, in fact, that the two roads are the same not only at the level of formulas, but also at the level of equivalence classes. That is, the following also commutes:

$$\begin{tikzcd}
{[Rx_1x_2x_3 \mid x_1, x_2, x_3]} \arrow[r, "\Pi^3_2"] \arrow[d, "\mathbf{S}\langle ax_1b \mid x_1\rangle"'] & {[\forall x_2 Rx_1x_2x_3 \mid x_1, x_3]} \arrow[d, "\mathbf{S}\langle ab \mid \emptyset\rangle"] \\
{[Rax_1b \mid x_1]} \arrow[r, "\Pi^1_1"'] & {[\forall x_2 Rax_2b \mid \emptyset]}
\end{tikzcd}$$

There's nothing special about the particular formulas here, we have as a more general fact that

Lemma 15. The following diagram commutes:

$$\begin{tikzcd}
\mathbf{S}(T_3) \arrow[r, "\Pi^3_2"] \arrow[d, "\mathbf{S}\langle ax_1b \mid x_1\rangle"'] & \mathbf{S}(T_2) \arrow[d, "\mathbf{S}\langle ab \mid \emptyset\rangle"] \\
\mathbf{S}(T_1) \arrow[r, "\Pi^1_1"'] & \mathbf{S}(T_0)
\end{tikzcd}$$

Proof. Without loss of generality, let $[\phi \mid x_1, x_2, x_3] \in \mathbf{S}(T_3)$. Following the 'top' path is easy and takes us to $[\forall x_2\phi(x_1/a, x_3/b) \mid \emptyset]$. Following the 'bottom' path we first get to $[\phi(x_1/a, x_2/x_1, x_3/b) \mid x_1]$, and from there to $[\forall x_1\phi(x_1/a, x_2/x_1, x_3/b) \mid \emptyset]$. But clearly this class intersects to $[\forall x_2\phi(x_1/a, x_3/b) \mid \emptyset]$. So since equivalence classes are disjoint, these are in fact the same class. $\square$

The general result being exemplified here is the following:

Lemma 16. Without loss of generality, let $\langle\tau \mid X_1^n\rangle : T^n \longrightarrow T^m$ be a term. Then the following commutes:

$$\begin{tikzcd}
\mathbf{S}(T_{m+1}) \arrow[r, "\Pi_j^{m+1}"] \arrow[d, "\mathbf{S}\langle\overline{\tau_{<j}}x_{n+1}\overline{\tau_{\geq j}}|X_1^{n+1}\rangle"'] & \mathbf{S}(T_m) \arrow[d, "\mathbf{S}\langle\tau|X_1^n\rangle"] \\
\mathbf{S}(T_{n+1}) \arrow[r, "\Pi_{n+1}^{n+1}"'] & \mathbf{S}(T_n)
\end{tikzcd}$$

Proof. Without loss of generality, let $[\phi \mid \overline{X_1^{m+1}}] \in \mathbf{S}(T_{m+1})$. Via the top path, this class gets sent to $[\forall x_j\phi(\overline{X_1^{m+1}} - x_j/\tau) \mid \overline{X_1^n}]$. Via the bottom path, it gets sent to $[\forall x_{n+1}\phi(\overline{X_1^{j-1}}/\tau_{<j}, x_j/x_{n+1}, \overline{X_{j+1}^{m+1}}/\tau_{>j}) \mid \overline{X_1^n}]$. To see these are the same class, observe that the first can be rewritten as $[\forall x_j\phi(\overline{X_1^{j-1}}/\tau_{<j}, x_j/x_j, \overline{X_{j+1}^{m+1}}/\tau_{>j}) \mid \overline{X_1^n}]$. With by-now-familiar tricks, we then see that if y occurs in neither representative formula, then both prove $\forall y\phi(\overline{X_1^{j-1}}/\tau_{<j}, x_j/y, \overline{X_{j+1}^{m+1}}/\tau_{>j}) \mid \overline{X_1^n}$. Equally clearly, and by the same tricks, this formula proves both representatives. Thus the classes intersect, so are identical. $\square$

3 Hyperdoctrines

That was a lot of information. Here are what we take to be the important bits:

Corollary 1 $\mathbf{S}$ is a contravariant functor from $\mathcal{B}$ to $\mathbf{BoolMon}$ whose image is in $\mathbf{Bool}$.

Lemma 12 Each arrow of the form $\mathbf{S}\langle\widehat{j} \mid X_1^{n+1}\rangle$ has a right adjoint Π_j^{n+1}.

Lemma 14 Whenever all of it makes sense, we get that

$$\Pi_j^n(\mathbf{S}\langle\widehat{j} \mid X_1^n\rangle[\phi \mid X_1^n - x_j] \sqcup [\psi \mid X_1^n]) \leq [\phi \mid X_1^n - x_j] \sqcup \Pi_j^n[\psi \mid X_1^n]$$

Lemma 16 Whenever all of it makes sense, the following diagram commutes:

$$\begin{tikzcd} \mathbf{S}(T_{m+1}) \arrow[r, "\Pi_j^{m+1}"] \arrow[d, "\mathbf{S}\langle\overline{\tau_{<j}}x_{n+1}\overline{\tau_{\geq j}}|X_1^{n+1}\rangle"'] & \mathbf{S}(T_m) \arrow[d, "\mathbf{S}\langle\tau|X_1^n\rangle"] \\ \mathbf{S}(T_{n+1}) \arrow[r, "\Pi_{n+1}^{n+1}"'] & \mathbf{S}(T_n) \end{tikzcd}$$

We define a Boolean hyperdoctrine to be a functor that has 'the same structure' as the functor $\mathbf{S}$:

Definition 10. A *Boolean hyperdoctrine* is a contravariant functor $H : \mathcal{B} \longrightarrow \mathbf{BoolMon}$ such that[4]

BH1 The image of H is in $\mathbf{Bool}$

BH2 Each arrow of the form $H\langle\widehat{j} \mid X_1^{n+1}\rangle$ has a right adjoint Π_j^{n+1}.

BH3 Whenever all of it makes sense, we get that

$$\Pi_j^n(H\langle\widehat{j} \mid X_1^n\rangle[\phi \mid X_1^n - x_j] \sqcup [\psi \mid X_1^n]) \leq [\phi \mid X_1^n - x_j] \sqcup \Pi_j^n[\psi \mid X_1^n]$$

BH4 Whenever all of it makes sense, the following diagram commutes:

$$\begin{tikzcd} H(T_{m+1}) \arrow[r, "\Pi_j^{m+1}"] \arrow[d, "H\langle\overline{\tau_{<j}}x_{n+1}\overline{\tau_{\geq j}}|X_1^{n+1}\rangle"'] & H(T_m) \arrow[d, "H\langle\tau|X_1^n\rangle"] \\ H(T_{n+1}) \arrow[r, "\Pi_{n+1}^{n+1}"'] & H(T_n) \end{tikzcd}$$

When it matters, we will distinguish the elements of and operations on the various algebras $H(T_n)$ by subscripting them. E.g. if necessary we will write $\leq_n$ for the partial order in $H(T_n)$ or $\top_n$ for its top element. Two key differences between arbitrary Booleans hyperdoctrine and the *syntactic hyperdoctrine* $\mathbf{S}$ are worth noting explicitly:

[4]There is a more general version of hyperdoctrines that take as their domain *any* category with enough structure to interpret the types of the language in question. This added complication adds little of importance in the case at hand, so is ignored.

- In $\mathbf{S}$, the algebras $\mathbf{S}(T_n)$ are always algebras of formulas. In an arbitrary Boolean hyperdoctrine H, $H(T_n)$ can be any Boolean algebra whatsoever.
- In $\mathbf{S}$, the homomorphisms $\mathbf{S}\langle\tau \mid X\rangle$ always arise via substitution. In an arbitrary Boolean hyperdoctrine H, $H\langle\tau \mid X\rangle$ can be any Boolean algebra homomorphism whatsoever.

Definition 11. If H is a Boolean hyperdoctrine, then an *interpretation* of $\mathcal{L}$ in H is a function $[\![-]\!]$ that assigns a member $[\![R]\!] \in H(T_n)$ to each n-ary predicate R. An interpretation induces an assignment of a semantic value $[\![\phi \mid X]\!]$ to each formula in the following way:

- If R is m-ary and $\tau \mid X : T_n \longrightarrow T_m$, then $[\![R\tau \mid X]\!] = H[\![\tau \mid X]\!][\![R]\!]$.
- $[\![\neg\phi \mid X]\!] = [\![\phi \mid X]\!]'$
- $[\![\phi \wedge \psi \mid X]\!] = [\![\phi \mid X]\!] \sqcap [\![\psi \mid X]\!]$
- $[\![\phi \vee \psi \mid X]\!] = [\![\phi \mid X]\!] \sqcup [\![\psi \mid X]\!]$
- $[\![\phi \rightarrow \psi \mid X]\!] = [\![\phi \mid X]\!]' \sqcup [\![\psi \mid X]\!]$
- If $\overline{Y} = y_1 y_2 \dots y_n$, then $[\![\forall y_j \psi \mid Y]\!] = \Pi_j^n [\![\psi \mid Y - y_j]\!]$.

Definition 12. The identity interpretation for the syntactic hyperdoctrine—written $[\![-]\!]_{\text{id}}$— is the assignment $[\![R]\!]_{\text{id}} = [R\overline{X_n^1} \mid X_n^1]$.

Lemma 17. $[\![\phi \mid X]\!]_{\text{id}} = [\phi \mid X]$ for all formulas $\phi \mid X$.

Lemma 18. For any Boolean hyperdoctrine H and interpretation $[\![-]\!]$,

$$[\![\langle\tau \mid X\rangle(\phi \mid Y)]\!] = H\langle\tau \mid X\rangle[\![\phi \mid Y]\!]$$

Definition 13.

- We say that $\phi \mid X$ is $\mathbf{K}$-valid in H relative to the interpretation $[\![-]\!]$ when $[\![\phi \mid X]\!] = \top_{\text{card}(X)}$.
- We say $\phi \mid X$ is $\mathbf{K}$-valid in H when $\phi \mid X$ is $\mathbf{K}$-valid in H relative to every interpretation.
- We say $\phi \mid X$ is $\mathbf{K}$-valid when $\phi \mid X$ is $\mathbf{K}$-valid in H for every H.

Theorem 1. If $\phi \mid X$ is $\mathbf{K}$-valid, then $\phi \mid X \in \mathbf{K}$.

Proof. We prove the contrapositive. If $\phi \mid X \notin \mathbf{K}$, then by Lemma 8, $[\phi \mid X] \neq \top_{\text{card}(X)}$. So by Lemma 17, $\phi \mid X$ is not $\mathbf{K}$-valid in the syntactic hyperdoctrine equipped with the identity interpretation. Thus $\phi \mid X$ is not $\mathbf{K}$-valid in the syntactic hyperdoctrine. So $\phi \mid X$ is not $\mathbf{K}$-valid. □

Corollary 2 *If $[\![\phi \mid X]\!] \leq [\![\psi \mid X]\!]$ in every interpreted Boolean hyperdoctrine, then $\phi \mid X \vdash_{\mathbf{K}} \psi \mid X$.*

It follows from Theorem 1 and Corollary 2 that the structural features we've identified in B1-B4 are satisfied by enough structures to falsify every nontheorem and to counterexample every nonentailment. But the syntactic hyperdoctrine has many features that we *didn't* include in B1-B4. As an example, not only is it true that the various algebras $\mathbf{S}(T_n)$ are Boolean algebras, it also happens to be the case that they are all countably infinite Boolean algebras generated by a countable infinity of atoms. But B1 doesn't require *any* of the $H(T_n)$'s to have either of these features. So B1-B4 allow us to interpret $\mathcal{L}$ in structures that look quite different from the prototypical example of a Boolean hyperdoctrine, $\mathbf{S}$.

Of course, admitting Boolean hyperdoctrines that are quite different from $\mathbf{S}$ runs the risk of admitting Boolean hyperdoctrines in which we can interpret some *theorems* as false and/or counterexample some entailments. That this *doesn't* happen is, we think, somewhat surprising.

Theorem 2. If $\phi \mid X \in \mathbf{K}$, then $\phi \mid X$ is $\mathbf{K}$-valid.

The proof, which we omit, is a straightforward induction on the length of the proof witnessing that $\phi \mid X \in \mathbf{K}$.

Corollary 3 *If $\phi \mid X \vdash_{\mathbf{K}} \psi \mid X$, then in every interpreted Boolean hyperdoctrine $[\![\phi \mid X]\!] \leq [\![\psi \mid X]\!]$.*

4 Metaphysical interpretation

We claimed at the outset that the story we were telling here would be of interest to metaphysicians. In this section we finally make good on that promise. The novelty of the presentation above is this: we have presented a respectable semantics for quantified first-order logic without any appeal to things that are being quantified over. We have, if you like, Being ($\exists$, that is) without beings.

Ordinary model theory is emphatically not like this. In an ordinary model (of the signature Σ, say) one has first of all, a domain of individuals. One then has the interpretations of the various symbols of Σ: interpretations of predicates are sets of individuals, interpretations of constants are individuals, and so on. Thus, the whole model-theoretic edifice *grounds out*, in some sense, at the level of individuals.

The grounding going on here, whatever it might be, is a fairly robust matter. To begin, there is the obvious dependence of sets on their members. But even putting that to the side, the comparison of structures ultimately comes down to the sets of individuals in their domains; maps of Σ-structures are simply maps between the underlying domains of those structures that happen to have certain further properties. The grounding of the model-theoretic world on the world of individuals and particulars further reveals itself on even a casual examination of many of the classical results of the subject. As often as not, said results are either statements about possible cardinalities for structures, or statements about how many structures there are (up to isomorphism) of a certain cardinality. Making generalizations about the psychology of workers in a scientific field is a risky business, but it seems fair enough to say that the models are fundamentally understood to be decorated sets (like groups, fields, and other objects in concrete categories), and that their underlying sets and the individuals that inhabit them are fundamental to the subject.

This incursion of set theoretic concepts into model theory, and from there into metaphysics, where sets are smuggled in as indispensable for semantics, has deeply colored contemporary analytic philosophy, both subtly and overtly. We can give some examples of both kinds of coloring.

Among the overtly colored subjects, we have, for example, the family of problems related to absolute generality. Parsons describes one of these problems in the following way:

> The universe of [the] metaphysician's purview surely includes everything, with no restriction tacit or otherwise. Logic might seem at first sight to envision only restricted generalization. We interpret the language of quantified logic with respect to a domain or 'universe of discourse'... Typically the domain is a set, and set theory tells us how, given a set, to describe a set containing elements not in the first set. In a sense, the received way of interpreting quantificational logic takes all quantifiers to be restricted. (Parsons, 2006, p203)

This is to put everything a little plainly, and in the literature one finds other approaches to quantification—holding on to the standard semantic machinery but replacing sets with classes, properties, pluralities, or some other kind of collection, or finding some way to think of many domains as being stitched together into one, perhaps via the semantics of modal logic—but the basic model-theoretic flavoring is clear. If one does without the idea of a domain (as we do) then it is not clear that the problem of absolute generality is even expressible, let alone a problem.

One finds overt model theoretic flavoring in certain versions of the bad company objection, familiar to Neologicists (Boolos, 1987). Here, the problem is roughly as follows. Certain axioms that Neologicists would like to have can only be true in models with domains of certain cardinalities. Sometimes, the cardinalities allowed by one axiom do not overlap with the cardinalities allowed by another. At this point, it's generally taken to be clear that the axioms are incompatible in some metaphysically deep sense, even though it's quite open (since we are dealing with second-order model theoretic semantics, for which there is no completeness theorem) that the axioms are perfectly consistent with one another in spite of not being jointly satisfiable.

A third example might be found in the debate over Putnam's model-theoretic argument for anti-realism (Putnam, 1980), and more generally in the discussion surrounding Skolem's paradox. Putnam argues roughly that, since by standard model theoretic results, there are many first-order models (of varying cardinalities) in which any set of platitudes or observations we might put forward would be satisfied, much of our mathematical language cannot have a fixed "intended interpretation", to the point where statements independent of the Zermelo-Frankel axioms cannot have truth values. Putnam proposes a "non-realist semantics" to get around this. The point at hand thought is that Putnam simply *assumes* that realism entails some form of broadly model-theoretic semantics for natural language.

As for less overt colorings, one could multiply examples endlessly. Lewis' metaphysics has a broadly model-theoretic flavor from the identification of properties with sets of (possible) individuals in On the Plurality of Worlds up through the fairly explicit picture of language in General Semantics.[5] And within metaphysics, Lewis casts a long shadow. More generally, the idea that the ground floor of metaphysics should somehow be a set of discrete individuals of some kind (whether mereological fusions, simple substances, space-time points, events, tropes, or some other kind of thing), over which

[5] See (Lewis, 1986) and (Lewis, 1970).

our best theory quantifies, is ubiquitous. And it's this picture to which we are offering an alternative.

What is the alternative? It's a world in which the ground floor is not a bunch of things over which we quantify, but a bunch of propositions, upon which the quantifiers act, transforming them from type to type. The world, at least as far as logic is concerned, is a totality of facts, not of things. Or, more precisely, the world is a totality of propositional functions organized into families by type. All that remains is determining how these families hang together. So—and this is the fun part—objects are in an important sense *secondary* features of the world, emergent from the underlying propositional structure. More to the point, objects on the hyperdoctrinal perspective are homomorphisms of algebras of propositional functions rather than members of a domain of quantification. Thus it is the *algebras* and not the *objects* that are taken as primitives—the whole edifice, that is, grounds out at the level of the algebra of propositional functions, not at the level of individuals. Objects, of course, still play an important role in our thinking about how the families of propositional functions hang together. But in the same way you would want to say that an isomorphism (a function) between two structures exists because of the way the structures are, rather than explaining the way the structures are by appealing to the existence of a certain function, one can say that objects (regarded as a certain type of homomorphism) depend on the algebraic structure of propositions, rather than the other way around.

One might object here that, even if hyperdoctrinal semantics lets us do first-order logic without a commitment to objects, it still commits us to a whole zoo of categorial machinery. The exposition of the theory commits us to functors, to adjoint pairs, to categories themselves… This objection however, misunderstands the nature of the machinery. It's the following kind of mistake: Imagine a nominalist who, upon discovering that his pet bird was in fact, an African swallow, bemoaned his new ontological commitment to African swallows. When we call a bird a swallow, we just give it a name that conceptualizes it as part of an orderly scheme for classifying organisms—we don't postulate a new thing. Analogously, functors, adjoint pairs, and categories are just ways of organizing and conceptualizing familiar structure, not new categories of beings that we here postulate. Our basic commitments are only to the consequence relation, to the propositional functions it orders, and to some basic operations on propositional functions, like negation and quantification.

Furthermore, hyperdoctrinal semantics is neutral on the nature of consequence, and on how to account for the algebraic relations of propositional functions. So, there is room here for a variety of different metaphysical pictures; perhaps one could return to the idea of objects inhabiting a domain of quantification if this seems the best way to explain what it means for one propositional function to entail another. The point is that this layer of metaphysical structure is not at all required for a precise metatheory of first-order logic. Instead, it is up to the metaphysician to motivate it, or reject it, on grounds internal to their practice rather than by appeal to some alleged logical necessity.

And there are more degrees of freedom here than just the recovery or abandonment of standard model-theoretic semantics, because there's no particular reason to restrict to the usual metaphysical data. Categories are agnostic about the structure of the objects they're made up of. This agnosticism lifts, in the case at hand, to an agnosticism about how ontology ought to be done. Thus, if agnosticism appeals to you, you ought to find hyperdoctrinal semantics a welcoming space. If you are agnostic about agnosticism, you're still likely to find hyperdoctrinal semantics useful, as a way of disentangling your gnostic ruminations from your logical commitments, and freeing up more space for you to explore.

References

Boolos, G. (1987). The consistency of Frege's foundations of arithmetic. In Thompson, J. J., editor, *On Being and Saying: Essays in Honor of Richard Cartwright*, pages 3–20. MIT Press, Cambridge, UK.

Lewis, D. (1986). *On the plurality of worlds*, volume 322. Blackwell, Oxford, UK.

Lewis, D. K. (1970). General semantics. *Synthese*, 22(1-2):18–67.

Parsons, C. (2006). The problem of absolute universality. In Rayo, A. and Uzquiano, G., editors, *Absolute Generality*. Oxford University Press, Oxford, UK.

Putnam, H. (1980). Models and reality. *Journal of Symbolic Logic*, 45:464–482.

Shay Logan
Kansas State University, Department of Philosophy
United States
E-mail: salogan@ksu.edu

Graham Leach-Krouse

Kansas State University, Department of Philosophy
United States
E-mail: gleachkr@ksu.edu