A Defense of Logicism[*]

Hannes Leitgeb
Munich Center for Mathematical Philosophy
Ludwig-Maximilians Universität München
hannes.leitgeb@lmu.de

Uri Nodelman
Philosophy Department
Stanford University
nodelman@stanford.edu

Edward N. Zalta
Philosophy Department
Stanford University
zalta@stanford.edu

Abstract

We argue that logicism, the thesis that mathematics is reducible to logic and analytic truths, is true. We do so by (a) developing a formal framework with comprehension and abstraction principles, (b) giving reasons for thinking that this framework is part of logic, (c) showing how the denotations for predicates and individual terms of an arbitrary mathematical theory can be viewed as logical objects that exist in the framework, and (d) showing how each theorem of a mathematical theory can be given an analytically true reading in the logical framework.

In this paper, we defend logicism, i.e., the claim that mathematics is reducible to logical and analytic truths alone, in the sense that the axioms and theorems of mathematics are derivable from logical truths and analytic truths. We shall assume, in what follows, that the deductive system of second-order logic is a part of logic, both in the usual contemporary sense of *logic* but also in the sense of logic developed later in the paper. This assumption doesn't require a second-order, model-theoretic consequence relation, and so our assumption about second-order logic doesn't require any set theory.

Our defense of logicism advances previous work in the following ways: (1) we precisely formulate the non-modal type-theoretic fragment of object theory;[1] (2) we show how this fragment can be used to analyze not just simple mathematical terms (as in previous work), but also complex relation terms of mathematical theories; (3) we argue that, given a natural definition of 'logic', this fragment of object theory is a logic and that its analysis of mathematics achieves the goals of logicism (as part of the argument, we show how our logicism differs from neo-logicism); and (4) we give a consistency proof by constructing (in the Appendix) a model for this fragment of object theory.

---

[1]This fragment has not previously been precisely formulated though it was used in Zalta 2000, Linsky & Zalta 2006, and Nodelman & Zalta 2014.

The papers cited in footnote 1 presupposed a standard notion of logical truth and assumed that logicism couldn't be true because (i) mathematical theories are often committed to a large, sometimes infinite, ontology, (ii) logic, understood to include second-order logic, requires only one individual and two properties, and (iii) the standard for reducing mathematics to logic is relative interpretability. Given these facts, there is no way to reduce the axioms of mathematical theories that have strong existence assumptions to theorems of logic.

In what follows, however, we argue that logicism is true, and indeed, that it can be given a serious defense. Our defense is based on a more nuanced notion of logical truth. Since the notion of logical truth defined in what follows yields a new body of such truths, this leads us to revise both (ii) and (iii) above. If logic is constituted by our new body of logical truths, then contrary to (ii), logic is committed to more than a non-empty domain of individuals and a 2-element domain of properties; indeed, it may be committed to much more. Contrary to (iii), both the conceptual and epistemological goals of the logicists can be achieved by adopting a notion of reduction other than relative interpretability. We suggest that relative interpretability is the *wrong* notion of reduction and suggest an alternative. We argue that this alternative notion of reduction gives up nothing important when it comes to establishing the most important goals set by the logicists for the foundations of mathematics.

In our defense of logicism, we shall attempt to show that logic includes special domains of individuals, properties, and relations, all of which can be asserted to exist by logical axioms. (Henceforth, we use 'objects' to refer generally to individuals, properties, and relations.) Thus, we agree with the early logicists that logic does have its own special logical objects. But we plan to *justify* this assumption in Section 6, when we defend logicism.

Moreover, when we add certain analytic truths to our background system, we'll be able to assert the existence of new logical objects. Given these new objects, our revised notion of reduction should be familiar: (a) every well-defined individual term of a mathematical theory T is assigned a logical individual as its denotation, (b) every well-defined property or relation term of T is assigned a logical property or logical relation as its denotation, and (c) every theorem of T is assigned a reading stated in terms of these denotations on which it turns out to be analytically true. Thus, we provide precise theoretical descriptions of the entities denoted by the predicates and individual terms of mathematical theories and this provides the means of stating precise truth conditions of the theorems and non-theorems of mathematical theories (Section 5). So once we establish that our background system is part of logic and that we've only extended it with *bona fide* analytic truths (Section 6), we will be defending logicism with respect to a genuine notion of reduction. And with a genuine reduction of mathematics to logic, we achieve the philosophical goals that were foremost in the minds of the early logicists.

But before we can discuss these issues, we start with a brief discussion of the philosophical goals of logicism (Section 1). Then, after discussing some motivating examples (Section 2), we turn to a presentation of our logical framework

(Section 3) and its axioms (Section 4). As noted above, we discuss the application of this logical framework to mathematics (Section 5) and present our argument that this is logicism (Section 6). We conclude by considering some potential objections (Section 7) and by presenting the smallest model of our logical framework in an Appendix.

1   The Goals of Logicism

Why did logicians and philosophers in the very early 20th century, such as Frege (1893/1903) and Whitehead & Russell (1910–1913), set out to establish the logicist thesis that most, if not all, of mathematics is reducible to the laws of logic and analytic truths? If logicism were true, what would be the philosophical benefits?

We take it that there are both conceptual and epistemological benefits. The conceptual benefit is clear: if mathematics is reducible to logic, the conceptual machineries of two *a priori* sciences are reduced to one. The concepts of mathematics become nothing other than concepts of logic. This simplifies the philosophy of mathematics, since (a) logicism would provide an account of all of mathematics, and not just the mathematics that is applied in, or is indispensable for, the natural sciences, and (b) logicism would provide an account of mathematics whether or not the mathematicians conclude that there is only one distinguished, true mathematical theory.

As to the epistemological benefits of logicism, Benacerraf provides one classic formulation:

> The philosophical point of establishing the view was nakedly epistemological: logicism, if it could be established, would show that our knowledge of mathematics could be accounted for by whatever would account for our knowledge of language. And, of course, it was assumed that knowledge of language could *itself* be accounted for in ways consistent with empiricist principles, that language was itself entirely learned. Thus, following Hume, all our knowledge could once more be seen as concerning either 'relations of ideas' (analytic and a priori) or 'matters of fact'.
>
> (Benacerraf 1981, 42–43)

So if logical truths are analytic, and mathematics is reducible to logical and other analytic truths, then we would have an explanation of mathematical knowledge.[2]

In the next section we give a brief overview of the *logical framework* that we shall use and the examples that motivate its application to mathematics. The

---

[2]This doesn't, strictly speaking, rule out the idea that some special faculty of intuition plays a role in our knowledge of mathematics, but only that if there is such a faculty, it is epistemologically innocent, in the sense that it doesn't require that there be a causal mechanism by which abstract mathematical objects give rise to intuitions. We can avoid Gödel's talk (1964, 268) of the analogy with sense perception, but keep the notion of intuition in an enlightened sense. The thesis that intuition provides some means of non-conceptual access to mathematical objects is perfectly consistent with the view that we will be developing here as long as this access is not meant to be causal.

logical framework we use is based on the logic of encoding described in Zalta 1983, 1988, and the applications were sketched in Zalta 2000, 2006. Those familiar with this earlier work and with the general idea can skip directly to the point where the new work in this paper emerges, namely, the detailed formal development of the logical framework in Sections 3 and 4.

2   Some Motivating Examples

Object theory uses the predicates $O!$ and $A!$ to distinguish between *ordinary* and *abstract* objects; in this paper we will take the predicate 'abstract' to be a primitive term of our theory. In a nutshell, abstract entities are individuated by a group of *encoded* properties that they objectify. This understanding of abstractness overlaps with common usage but doesn't coincide with it exactly. So we are happy to regard 'abstract' as a technical term the meaning of which is given more precisely by AXIOMS 3–7 (in Section 4.2) of our theory.

Our logical framework, in its full generality, is developed within a relational type theory. However, after we present the framework, we'll focus only on a certain fragment. To keep the presentation simple, our analysis will focus on those first- or second-order mathematical theories statable in terms of primitive individual constants and primitive 1- and 2-place predicates. Examples of such mathematical theories include Zermelo-Fraenkel set theory (ZF), Peano Arithmetic (PA), real number theory ($\mathbb{R}$), etc.[3] We'll therefore motivate our general logical framework by taking ZF as a typical example. We note that *all* of the philosophically relevant ideas concerning our analysis of mathematics can be understood by examining this basic example, since it should be clear how to extend the framework to analyze mathematical theories requiring more expressive power. That is, the logical framework defined later in Sections 3 and 4 can be further applied in a variety of ways, e.g., to analyze mathematical theories stated in terms of n-place predicates (for arbitrary $n \geq 0$) and not just 1- and 2-place predicates.

We begin by illustrating what our background logical framework must accomplish. We shall take the basic data of mathematics to be *contextualized* mathematical claims, in the same way that sentences with theoretical terms in empirical science only have meaning in the context of scientific theories. For example, a set theorist, given the context of some set theory, might make the following statement:

> No set is an element of the empty set.

Though we shall later formally represent this sentence as stated (see Section 5.2), we begin by making the context explicit and representing the above sentence as the following, where T is some set theory:

> In set theory T, no set is an element of the empty set.

---

[3]We'll assume, for the present purposes, that any functional terms used in the statement of the axioms of these theories have been replaced by predicates and the relevant existence and uniqueness claims.

The sentence displayed above would typically be represented formally as follows, where $\vdash_T$ indicates theoremhood with respect to theory T and 'S' denotes the property of being a set (relative to T) and '$\emptyset$' is a constant of T that denotes the empty set:

$$\vdash_T \neg\exists y(Sy \,\&\, y \in \emptyset)$$

Now to be even more specific, suppose the theory T in question is Zermelo-Fraenkel set theory, formulated with the primitive constant $\emptyset$ and the primitive 2-place relation $\in$. On this formulation, the fact that the Null Set Axiom is a theorem of ZF is expressed as $\vdash_{\text{ZF}} \neg\exists y(y \in \emptyset)$ instead of as $\vdash_{\text{ZF}} \exists x \neg\exists y(y \in x)$.

Now to analyze ZF, PA, $\mathbb{R}$, etc., we shall represent their languages within our (higher-order) framework and so include closed λ-expressions (i.e., with no free variables) such as $[\lambda x\, \varphi]$, $[\lambda F\, \varphi]$, and $[\lambda R\, \varphi]$, $[\lambda xFR\, \varphi]$, all of which are governed by λ-Conversion (and α-Conversion) – see Section 4.1.2 below. In $[\lambda x\, \varphi]$, the λ binds the individual variable 'x' to produce an expression that denotes a property of individuals; in $[\lambda F\, \varphi]$, the λ binds the first-level property variable 'F' to produce an expression that denotes a property of first-level properties; in $[\lambda R\, \varphi]$, the λ binds the first-level 2-place relation variable 'R' to produce an expression that denotes a property of first-level binary relations; and in $[\lambda xFR\, \varphi]$, the λ binds 3 variables (of the types just mentioned).

So where '$\emptyset$' denotes the empty set of ZF, 'S' denotes the ZF property of being a set, and $\in$ denotes the membership relation of ZF, we may infer the following sentences from the above theorem understood now as a theorem of ZF (in which the font sizes of the symbols '$\emptyset$', 'S' and '$\in$' are reduced when they are in argument position):[4]

$$\vdash_{\text{ZF}} [\lambda x \,\neg\exists y(Sy \,\&\, y \in x)]\emptyset$$
$$\vdash_{\text{ZF}} [\lambda F \,\neg\exists y(Fy \,\&\, y \in \emptyset)]S$$
$$\vdash_{\text{ZF}} [\lambda R \,\neg\exists y(Sy \,\&\, yR\emptyset)]\in$$
$$\vdash_{\text{ZF}} [\lambda FRx \,\neg\exists y(Fy \,\&\, yRx)]S\in\emptyset$$

That is, from the fact that it is a theorem of ZF that no set is an element of the empty set, we know that: (a) it is a theorem of ZF that the empty set exemplifies the (first-level) property of being an (individual) x such that no set is a member of x; (b) it is a theorem of ZF that the property of being a set exemplifies the second-level property of being a property F such that nothing exemplifying F is an element of the empty set; (c) it is a theorem of ZF that the membership relation exemplifies the second-level property of being a relation R such that no set bears R to $\emptyset$, and (d) it is a theorem of ZF that the empty set, the property of being a set, and the membership relation stand in the relation: being a property F and relation R, and individual x, such that nothing that exemplifies F bears R to x.

---

[4]In the following examples, we preserve the infix notation for the relation $\in$ by using a formula of the form yRx. However, when we define our logical framework, we will define relational predications in the usual way as having the form Ryx, and the infix variant yRx will be an abbreviation of the former; it is useful for those cases of relation terms such as $\in$ which traditionally appear using infix notation.

Thus, from the single theorem $\exists y(Sy \ \& \ y \in \emptyset)$, we have inferred additional theorems about the properties exemplified by the objects S, $\in$, and $\emptyset$. We shall import *all* of these theorems into our logical framework as *analytic* truths about what is true in ZF. In particular, sentences very much like the following will be analytic truths of our background theory:

$$\text{ZF} \vDash \neg\exists y(Sy \ \& \ y \in \emptyset) \tag{A}$$
$$\text{ZF} \vDash [\lambda x \ \neg\exists y(Sy \ \& \ y \in x)]\emptyset \tag{B}$$
$$\text{ZF} \vDash [\lambda F \ \neg\exists y(Fy \ \& \ y \in \emptyset)]S \tag{C}$$
$$\text{ZF} \vDash [\lambda R \ \neg\exists y(Sy \ \& \ yR\emptyset)]\in \tag{D}$$

These statements have the form $z \vDash p$, in which 'z' is an individual variable and 'p' is a variable for a proposition. Statements of this form will be explicitly defined in terms of one of the primitive logical notions embedded within our logical framework. We shall introduce that definition below, but to complete our examples, notice that if we continue to use 'p' as a variable for propositions, 'F' as a variable ranging over first-level properties, '$\mathcal{F}$' as a variable ranging over second-level properties of first-level properties, and '$\mathcal{R}$' as a variable ranging over second-level properties of first-level relations, then:

Substitute $\neg\exists y(Sy \ \& \ y \in \emptyset)$ for p in $\text{ZF} \vDash p$ to obtain (A).
Substitute $[\lambda x \ \neg\exists y(Sy \ \& \ y \in x)]$ for F in $\text{ZF} \vDash F\emptyset$ to obtain (B).
Substitute $[\lambda F \ \neg\exists y(Fy \ \& \ y \in \emptyset)]$ for $\mathcal{F}$ in $\text{ZF} \vDash \mathcal{F}S$ to obtain (C).
Substitute $[\lambda R \ \neg\exists y(Sy \ \& \ yR\emptyset)]$ for $\mathcal{R}$ in $\text{ZF} \vDash \mathcal{R}\in$ to obtain (D).

Consider that we can now, as a matter of logic, single out all and only those first-level properties F that satisfy the open formula $\text{ZF} \vDash F\emptyset$; single out all and only those second-level properties $\mathcal{F}$ that satisfy the open formula $\text{ZF} \vDash \mathcal{F}S$; and single out all and only those second-level properties $\mathcal{R}$ that satisfy the open formula $\text{ZF} \vDash \mathcal{R}\in$.

We plan to logically objectify each of the groups of properties singled out by these open formulas. To see how, suppose that for any formula φ in which F may or may not be free, there is a unique, abstract individual that *codes up* all and only the first-level properties of individuals satisfying φ. So given the open formula $\text{ZF} \vDash F\emptyset$, there is a unique abstract individual that *encodes* all and only the first-level properties F such that $\text{ZF} \vDash F\emptyset$; later, we'll argue that this individual is a logical object. Using '$A!$' to denote the first-level property of being abstract, and 'xF' to assert that the individual x *encodes* the property F, and definite descriptions of the form $\iota x\varphi$, we could then formulate the following theoretical identification:

$\emptyset_{\text{ZF}} = \iota x(A!x \ \& \ \forall F(xF \equiv \text{ZF} \vDash F\emptyset_{\text{ZF}}))$
The empty set of the mathematical theory ZF is the abstract individual x that encodes all and only those (first-level) properties F such that in ZF, the ZF-empty-set exemplifies F.

Here we are deploying the primitive notion of encoding, x *encodes* F, represented by the formula xF, in which the argument term x is written to the left of the 1-place relation term F. Formulas of the form xF are to be distinguished from the

traditional form of n-place exemplification predication $F^n x_1 \ldots x_n$. The logic of encoding has been described in Zalta 1983, 1988, and elsewhere. Encoding is a *primitive* mode of predication that holds between an abstract object and the properties by which we conceive of it. Encoding is axiomatized rather than defined, and we shall review the axioms governing it below. Thus, in the above example, $\emptyset_{\mathrm{ZF}}$ is the object that encodes all and only the properties that the theory ascribes to it, namely, all and only those Fs such that in the theory ZF, the empty set exemplifies F. In Section 6, we plan to show that this abstract object is in fact a logical object.

Now to extend these ideas to higher types, suppose encoding formulas can be generalized and are well-formed whenever the term on the left signifies an object of some type and the term on the right signifies the (immediately higher-level) property of that type that's encoded. Suppose that for any open formula φ in which $\mathcal{F}$ may or may not be free, there is a unique first-level *abstract property* of individuals that codes up all and only the second-level properties of properties satisfying φ. So if we are now given the open formula ZF $\vDash \mathcal{F}S$ (mentioned above), there is a unique first-level abstract property of individuals that *encodes* all and only the second-level properties $\mathcal{F}$ such that ZF $\vDash \mathcal{F}S$. Using '$A!$' now to denote the second-level property of being abstract, and $F\mathcal{F}$ to assert that the first-level property F *encodes* the second-level property $\mathcal{F}$, we could then formulate the following theoretical identification of the ZF-property of being a set (S_{ZF}):[5]

> $S_{\mathrm{ZF}} = \iota F(A!F \,\&\, \forall \mathcal{F}(F\mathcal{F} \equiv \mathrm{ZF} \vDash \mathcal{F}S_{\mathrm{ZF}}))$
> The ZF-property of being a set is the (first-level) abstract property F that encodes all and only those second-level properties $\mathcal{F}$ of first-level properties such that in ZF, the ZF-property of being a set exemplifies $\mathcal{F}$.

Clearly, one of the second-level properties encoded by S_{ZF} is the property $[\lambda F \,\neg \exists y(Fy \,\&\, y \in \emptyset)]$.

Finally, suppose that for any formula φ in which $\mathcal{R}$ may or may not be free, there is unique first-level *abstract relation* among individuals that codes up all and only the second-level properties of relations satisfying φ. If we are now given the last of the open formulas mentioned above, namely, ZF $\vDash \mathcal{R}\in$, there there is a unique first-level abstract relation that can *encodes* all and only the second-level properties $\mathcal{R}$ such that ZF $\vDash \mathcal{R}\in$. Using '$A!$' to denote the second-level property of being abstract, and '$R\mathcal{R}$' to assert that the first-level relation R *encodes* the second-level property $\mathcal{R}$, we could then formulate the following theoretical identification of the ZF-membership relation $(\in_{\mathrm{ZF}})$:[6]

> $\in_{\mathrm{ZF}} \;=\; \iota R(A!R \,\&\, \forall \mathcal{R}(R\mathcal{R} \equiv \mathrm{ZF} \vDash \mathcal{R}\in_{\mathrm{ZF}}))$
> The membership relation of ZF is the first-level abstract relation R that encodes all and only those second-level properties $\mathcal{R}$ of first-level relations such that in the theory ZF, the ZF-membership relation exemplifies $\mathcal{R}$.

---

[5] Notice that in the encoding formula $F\mathcal{F}$, we've made the italic 'F' slightly smaller in size, so as to make it clear that F is the argument and $\mathcal{F}$ is the second-level property it encodes.

[6] Again, in the encoding formula $R\mathcal{R}$, we've made the italic R slightly smaller in size, so as to make it clear that R is the argument and $\mathcal{R}$ is the second-level property it encodes.

As we shall see, theoretical identifications like the ones described above are an essential component of our reduction of mathematics to logic. It is important here not to regard these theoretical identifications as *definitions* of the expressions on the left-side of the identity sign, for they appear on the right-side as well. Instead, they are to be regarded as theoretical principles of object theory. We are supposing that from a well-defined body of data, i.e., a body of analytic truths of form "In theory T, p", one can 'abstract out' objects that encode all and only the theoretical properties of the individuals and relations denoted by the constants and 1- and 2-place predicates of T. The other essential component of our reduction will be to show how each theorem of T is given a reading on which it is true. This will be the topic of Section 5.2. But first, we present our logical framework in detail.

3 The Language of the Logical Framework

Our logical framework has to be defined so that the foregoing formal representations are well-formed. We therefore start with a relational type theory, so that we can quantify over objects of higher type. To be specific, let us define a *type* as follows:

> i is a type.
> Whenever $t_1, \ldots, t_n$ are any types ($n \geq 0$), $\langle t_1, \ldots, t_n \rangle$ is a type.

We use i as the type for *individuals*, and $\langle t_1, \ldots, t_n \rangle$ as the type for *relations* among objects having types $t_1, \ldots, t_n$, respectively. Henceforth, where t is any type and $n = 1$, we call entities of type $\langle t \rangle$ *properties*. When $n = 0$, we say that $\langle \rangle$ is the type for *propositions*. So properties are 1-place relations and propositions are 0-place relations. We continue to use 'object' to refer to entities of any type.

Constants and variables of every type t will serve as the primitive terms of our language. For any type t, we include the distinguished predicate $A!$, which denotes a primitive property of objects of type t, namely, *being abstract*. We define the language $\mathcal{L}$ by (simultaneously) defining the formulas and terms that constitute the well-formed expressions of $\mathcal{L}$, and we'll assume that the notions of free and bound variables are definable in the usual way.

Simple Terms. Any constant or variable of type t is a (simple) term of type t.

Exemplification formulas. Where $\tau_1, \ldots, \tau_n$ (for $n \geq 0$), are terms of type $t_1, \ldots, t_n$, respectively, and Π is a term of type $\langle t_1, \ldots, t_n \rangle$, then the expression $\Pi \tau_1 \ldots \tau_n$ is an exemplification formula. When $n \geq 1$, we read $\Pi \tau_1 \ldots \tau_n$ as "$\tau_1, \ldots, \tau_n$ *exemplify* Π", and when $n = 0$, we read Π as "Π is true". Truth is the 0-place case of exemplification.

Encoding formulas. Where τ is any term of type t and Π is a term of type $\langle t \rangle$, then the expression $\tau \Pi$ is an encoding formula. We read $\tau \Pi$ as "τ *encodes* Π".

Complex formulas. Where φ, ψ are any formulas and α is any variable, then $\neg \varphi$ ('it is not the case that φ'), $\varphi \rightarrow \psi$ ('if φ, then ψ') and $\forall \alpha \varphi$ ('every α

is such that φ') are complex formulas. We henceforth employ formulas of the form $\varphi \,\&\, \psi$, $\varphi \lor \psi$, and $\varphi \equiv \psi$, as these can be defined in terms of our complex formulas. We define *is a subformula of* φ in the usual way.[7] We say that ψ is a *proper* subformula of φ just in case ψ is a subformula of φ but not identical to φ.

Propositional Formulas. φ is a *propositional formula* iff φ has no encoding subformulas.

Complex terms. There are two kinds of complex terms: definite descriptions and complex relation terms: (1) *Definite descriptions.* Where α is any variable of type $t \neq \langle \rangle$ and φ is any formula, then $\iota\alpha\varphi$ ("the α such that φ") is a complex term having type t.[8] (2) *Complex relation terms.* Where φ is any *propositional* formula that contains no descriptions, then (a) φ is a complex relation term having type $\langle \rangle$, and (b) if $\alpha_1, \ldots, \alpha_n$ ($n \geq 1$) are variables of type $t_1, \ldots, t_n$, respectively, then $[\lambda\alpha_1 \ldots \alpha_n \, \varphi]$ ("being an $\alpha_1, \ldots, \alpha_n$ such that φ") is a complex relation term having type $\langle t_1, \ldots, t_n \rangle$.[9]

Primary Terms. We define τ is a primary term of φ as follows: The primary terms of the exemplification formula $\Pi\tau_1 \ldots \tau_n$ are Π, $\tau_1, \ldots,$ and τ_n. The primary terms of the encoding formula $\tau\Pi$ are τ and Π. The primary terms of a complex formula are the primary terms of its proper subformulas.

Although the foregoing defines the language $\mathcal{L}$ of our logical framework in complete generality, we shall frequently, in what follows, work with only a *fragment* of this language. For example, we often work with abstract objects denoted by terms limited to the following types: i, $\langle \rangle$, $\langle i \rangle$, $\langle i, i \rangle$, $\langle \langle i \rangle \rangle$, and $\langle \langle i, i \rangle \rangle$. (In

---

[7]That is: φ is a subformula of φ; if $\neg\psi$ is a subformula of φ, then ψ is a subformula of φ; if $\psi \to \chi$ is a subformula of φ, then ψ and χ are subformulas of φ; if $\forall\alpha\psi$ is a subformula of φ, then ψ is a subformula of φ; and nothing else is a subformula of φ.

[8]This rules out descriptions such as $\iota p(p \,\&\, \neg p)$. In this example, the description operator would bind a variable of type $\langle \rangle$. We won't need descriptions of this type in our analysis of mathematics.

[9]For purposes of this paper, we won't need terms of the form $[\lambda \, \varphi]$, where the λ doesn't bind any variables. Moreover, the lack of descriptions in λ-expressions is not egregious, since if a description provably denotes something, one can instantiate into a λ-expression with a free variable. The free variables in such λ-expressions will be subject to universal generalization and so denoting terms can be instantiated for them.

This clause also banishes encoding subformulas from λ-expressions (since none occur in propositional formulas). The justification for this is discussed in some detail in Bueno, Menzel, & Zalta 2014 (808–810). Their justification begins with the fact that the theory being developed, intuitively, starts with the domain of (ordinary) relations with which we are all familiar from second-order logic and higher-order logic, in which exemplification is the only form of predication. Then we build abstract objects out of those without attempting to alter the domain of relations. (This intuitive picture is captured by the model construction in the Appendix, which builds the relations from the bottom up before it builds the abstracta from the top down.) This justification continues by considering the fact that η-Conversion, which asserts $[\lambda x_1 \ldots x_n F^n x_1 \ldots x_n] = F^n$, intuitively tells us that ordinary relations are defined by their exemplification positions. See AXIOM 7 below. The reader may consult the cited pages of Bueno, Menzel, & Zalta 2014 for a fuller discussion.

The restriction to propositional formulas in λ-expressions also has the advantage that it avoids paradox. If such expressions as $[\lambda x \, \exists F(xF \,\&\, \neg Fx)]$ were well-formed, then a contradiction would be derivable from the assertion that there is an abstract individual that encodes such a property. (An abstract individual that encodes $[\lambda x \, \exists F(xF \,\&\, \neg Fx)]$ would exemplify this property iff it does not.)

the Appendix, we define an explicit fragment by defining the bounded language $\mathcal{L}_{n,m}$, that includes these and the other types needed in what follows.) Thus, we'll be using the following specific variables:

$x, y, z, \ldots$ are variables of type i
$p, q, r, \ldots$ are variables of type $\langle\,\rangle$
$F, G, H, \ldots$ are variables of type $\langle i \rangle$
$R, S, \ldots$ are variables of type $\langle i, i \rangle$[10]
$\mathcal{F}, \mathcal{G}, \mathcal{H}, \ldots$ are variables of type $\langle\langle i \rangle\rangle$
$\mathcal{R}, \mathcal{S}, \ldots$ are variables of type $\langle\langle i, i \rangle\rangle$

Some observations about our language are in order. First, the propositional formulas are those formulas which are built up out of exemplification formulas and the sentence-forming operations of negation, conditionalization, and quantification described in the definition of complex formulas. Consequently encoding formulas can only make an appearance inside a propositional formula φ if they are buried in a term within some propositional subformula of φ. For example, the formula $Rx\iota y(yG)$ ('x and the y that encodes G exemplify the relation R') and the formula $[\lambda x Rx\iota y(yG)]z$ ('z exemplifies the property of being an x that bears R to the y that encodes G') are well-formed propositional formulas since they have no encoding subformulas. By contrast, the formula $\forall F(xF \rightarrow Fx)$ is not propositional, since it has xF as an encoding subformula.

   Second, since the variables $p, q, \ldots$ are terms of type $\langle\,\rangle$, they are also formulas, by the definition of exemplification formulas. Thus, by the second clause of the definition of complex terms, we can form λ-expressions such as $[\lambda x\, p]$. These denote a properties of individuals, i.e., a property with type $\langle i \rangle$, and we read $[\lambda x\, p]$ as *being such that p*, where p denotes some proposition.

   Finally, by the second clause of the definition of complex terms, we shall be able to formulate λ-expressions such as $[\lambda y\, \varphi]$, $[\lambda F\, \varphi]$, and $[\lambda R\, \varphi]$, when φ is propositional. These will denote, respectively, a property of individuals, a property of first-level properties, and a property of first-level binary relations. Note that the variable bound by the λ need not be free in φ. As we shall see, the resulting expressions behave as expected. For example, it is axiomatic that in the case where the variable y is not free in φ, an individual x exemplifies $[\lambda y\, \varphi]$ iff (the proposition denoted by) φ is true.

   Given these observations about the language of our logical framework, we conclude this section by:

(1) defining the property *being ordinary* as the negation of $A!$,

(2) distinguishing between abstract and ordinary objects of every type by stating their identity conditions, and

(3) defining the conditions under which a proposition p is *true in* an abstract individual x.

---

[10]Notice, here, that we've now used the symbol S in two ways: earlier in the paper we used S as a constant to denote the property *being a set* (and thus an term of type $\langle i \rangle$) and in the above list of variables, we've used S as a variable ranging over first-level relations (and thus an term of type $\langle i, i \rangle$). The context will always make it clear which of these is intended.

Concerning (1). We previously mentioned that where t is any type, then '$A!$' is a distinguished predicate of type $\langle t \rangle$. The symbol $A!$ is a 'typically ambiguous' primitive that denotes a property exemplified by the objects of type t that are *abstract*. And where t is any type and α is a variable of type t, we say that the property *being ordinary* ('$O!$') is *being an α such that α fails to exemplify being abstract*:

$$O! \ =_{df} \ [\lambda\alpha \, \neg A!\alpha]$$

Thus, the typically ambiguous predicate $O!$ is a term of type $\langle t \rangle$, for any type t. The predicates $A!$ and $O!$ consequently partition the domain of each type t into the abstract and ordinary objects of type t. We'll later assert, as an axiom, that any object which encodes a property is abstract.

Concerning (2). Identity is not primitive but is rather defined. Although the definitions in full generality are complex, they are easy to grasp. If x and y are abstract objects of any type t, then x and y are identical whenever x and y encode the same properties having type $\langle t \rangle$. If x and y are any ordinary objects, then we define their identity by cases: (a) ordinary individuals x and y are identical whenever they exemplify the same properties; (b) ordinary properties F and G with type $\langle t \rangle$, where t is any type, are identical just in case they are encoded by the same objects. Identity for ordinary objects of the remaining types are defined in terms of property identity: (c) ordinary propositions p and q of type $\langle \rangle$ are identical just in case the properties *being such that p* and *being such that q* are identical; and (d) ordinary relations F and G of type $\langle t_1, \ldots, t_n \rangle$, where $t_1, \ldots, t_n$ are any types, are identical just in case every way of projecting F and G onto any $n-1$ objects of the appropriate types yields identical properties.[11]

Concerning (3). Using the notion of identity just defined, we may define two more notions that are needed to see how the framework parses the three theoretical identifications in the illustrative examples of Section 2. First, we define a *situation* to be any abstract individual x such that every property x encodes is a property of the form $[\lambda y\, p]$, for some proposition p. Formally, where

---

[11] The foregoing can be captured formally as follows. Where x and y are both distinct variables of type i, and $O!$, $A!$, and F have type $\langle i \rangle$:

$$x = y \ =_{df} \ (O!x \,\&\, O!y \,\&\, \forall F(Fx \equiv Fy)) \vee (A!x \,\&\, A!y \,\&\, \forall F(xF \equiv yF))$$

Where F and G are both of type $\langle t \rangle$ and where $O!$, $A!$, and $\mathcal{H}$ are all of type $\langle\langle t \rangle\rangle$:

$$F = G \ =_{df} \ (O!F \,\&\, O!G \,\&\, \forall x^t(xF \equiv xG)) \vee (A!F \,\&\, A!G \,\&\, \forall\mathcal{H}(F\mathcal{H} \equiv G\mathcal{H}))$$

Where p and q are both of type $\langle \rangle$, x is a variable of type i and $O!$, $A!$, and $\mathcal{H}$ have type $\langle p \rangle$:

$$p = q \ =_{df} \ (O!p \,\&\, O!q \,\&\, [\lambda x\, p] = [\lambda x\, q]) \vee (A!p \,\&\, A!q \,\&\, \forall\mathcal{H}(p\mathcal{H} \equiv q\mathcal{H}))$$

And where F and G are both of type $\langle t_1, \ldots, t_n \rangle$:

$$F^{\langle t_1, \ldots t_n \rangle} = G^{\langle t_1, \ldots t_n \rangle} \ =_{df} \quad (\text{where } n > 1)$$

$$
\begin{aligned}
&O!F \,\&\, O!G \,\&\, \forall y^{t_2} \ldots \forall y^{t_n}([\lambda x^{t_1}\ Fx^{t_1}y^{t_2} \ldots y^{t_n}] = [\lambda x^{t_1}\ Gx^{t_1}y^{t_2} \ldots y^{t_n}]) \,\&\, \\
&\forall y^{t_1} \forall y^{t_3} \ldots \forall y^{t_n}([\lambda x^{t_2}\ Fy^{t_1}x^{t_2}y^{t_3} \ldots y^{t_n}] = [\lambda x^{t_2}\ Gy^{t_1}x^{t_2}y^{t_3} \ldots y^{t_n}]) \,\&\, \ldots \,\&\, \\
&\quad \forall y^{t_1} \ldots \forall y^{t_{n-1}}([\lambda x^{t_n}\ Fy^{t_1} \ldots y^{t_{n-1}}x^{t_n}] = [\lambda x^{t_n}\ Gy^{t_1} \ldots y^{t_{n-1}}x^{t_n}]) \quad \vee
\end{aligned}
$$

$$A!F \,\&\, A!G \,\&\, \forall\mathcal{H}(F\mathcal{H} \equiv G\mathcal{H})$$

x and y are variables of type i, $A!$ has type $\langle i \rangle$, and F is a variable of type $\langle i \rangle$, then:

$$Situation(x) \ =_{df} \ A!x \ \& \ \forall F(xF \to \exists p(F = [\lambda y \ p]))$$

Then where s is any situation, we say p *is true in* s, written $s \vDash p$, iff s encodes the property *being such that* p:

$$s \vDash p \ =_{df} \ s[\lambda y \ p]$$

Note that since $s \vDash p$ is defined in terms of the encoding formula $s[\lambda y \ p]$, it may not appear as a subformula in a propositional formula.

4 The Axioms for the Logical Framework

We can reason using the preceding language by adopting the following groups of principles and rules:

1. The classical axioms and rules of predicate logic, as they are formulated for relational type theory. These are modified only to accommodate the (negative) free logic of definite descriptions. (Thus, a definite definition $\iota\alpha\varphi$ can be instantiated into universal claims only when it is known, by proof or by hypothesis, that $\exists\beta(\beta = \iota\alpha\varphi)$, i.e., that the description is logically proper.[12])

2. An axiom for the substitution of identicals is completely unrestricted.

3. Axioms governing the two kinds of complex terms: definite descriptions and the λ-expressions.

4. Axioms governing the primitive predicate $A!$ and governing encoding predications.

We shall not review (1) and (2) except to say that (a) all of the usual axioms and rules of propositional logic are included, (b) the classical quantifier axioms and rules (suitably modified to accommodate the free logic of definite descriptions) apply to all quantified formulas in which the quantifiers bind variables of any type, and (c) the substitution of identicals governs our *defined* notion of identity for every logical type. In addition to these axioms, we assume only the primitive rules of Modus Ponens and Generalization, and the usual rules that are derivable

---

[12] We shall assume familiarity with the following facts about negative free logic. First, the classical quantifier axiom for universal instantiation is modified so that terms τ can only be instantiated into a universal claim if one knows that $\exists\beta(\beta = \tau)$. Second, for every term τ other than a description, it is axiomatic that $\exists\beta(\beta = \tau)$. Third, and finally, for definite descriptions of the form $\iota\alpha\varphi$ it is an axiom that: $\psi_\alpha^{\iota\alpha\varphi} \to \exists\beta(\beta = \iota\alpha\varphi)$, where ψ is any atomic exemplification or encoding formula in which α occurs as one of the arguments, β doesn't occur free in φ, and $\psi_\alpha^{\iota\alpha\varphi}$ is the result of substituting $\iota\alpha\varphi$ for all the free occurrences of α in φ. This simply captures the principle underlying negative free logic that any atomic formula containing a non-denoting term is false.

from this basis. However, axiom groups (3) and (4) are discussed in subsections below.

It is important to emphasize here, however, that our framework and its application do not semantically presuppose anything more than general Henkin models. The first-level property variables $F, G, \ldots$ need not range over the full power set of the domain over which the individual variables $x, y, \ldots$ range. And, in general, our model in the Appendix shows that the domain of properties having type $\langle t \rangle$ is not the power set of the domain of objects of type t.[13] Nevertheless, in the model described in the Appendix, the axioms discussed below are all true.

In the remainder of this section, then, we describe the axioms that govern the complex terms (Section 4.1) and that govern abstract and ordinary objects of any type t (Section 4.2).

4.1 Axioms Governing the Complex Terms

4.1.1 Definite Descriptions

The principle governing definite descriptions is simply this:

> AXIOM 1 (**Description Axiom**).
> $\beta = \iota\alpha\varphi \equiv \forall\alpha(\varphi \equiv \alpha = \beta)$, provided β is substitutable for α in φ.

This asserts: β is the α such that φ if and only if β is uniquely φ. As a simple example, let α, β be the type i variables x, y, respectively, let Q be a type $\langle i \rangle$ constant, and let φ be the exemplification formula Qx. Then the following is an instance of the Description Axiom:

$$y = \iota x Q x \equiv \forall x(Qx \equiv x = y)$$

This asserts: y is identical to the x such that Qx if and only if y is the unique individual that exemplifies Q. Although we shall not take the time to prove it here, the classical Russell axiom for descriptions is now derivable.[14]

4.1.2 Principles Governing Relations

We employ the standard axiom of λ-Conversion for relations denoted by λ-expressions in which the λ binds one or more variables:

> AXIOM 2: $[\lambda\alpha_1 \ldots \alpha_n \, \varphi]\alpha_1 \ldots \alpha_n \equiv \varphi$

For example, consider the sentence $\neg\exists y Kxa$, where y and a have type i and K is a binary relation constant of type $\langle i, i \rangle$. Then the following consequence of a unary instance of AXIOM 2 asserts that a certain property of relations having type $\langle\langle i, i \rangle\rangle$ is exemplified by K iff $\neg\exists y Kxa$ holds:

---

[13]In general, the domain $\mathbf{D}_t$ of type t is the union of the ordinary objects of type t and the abstract objects of type t. So, $\mathbf{D}_{\langle t \rangle}$ includes all the ordinary properties with type $\langle t \rangle$ and the abstract properties with type $\langle t \rangle$. It will be seen, upon inspection, that this is *not* the power set of $\mathbf{D}_t$.

[14]The proof depends on the facts about the negative free logic of descriptions noted in footnote 12.

$$[\lambda R \, \neg\exists y R y a] K \; \equiv \; \neg\exists y K y a$$

Thus, instances of λ-Conversion simply require that the denotation of the λ-expression be a relation whose exemplification extension consists of the entities that satisfy the λ-expression's matrix. We'll also assume that a principle of α-Conversion, which asserts an identity between alphabetic variants, governs all our λ-expressions; but for simplicity, we won't state this axiom explicitly.

4.2   Principles Governing Encoding

We turn finally to the axioms governing our primitive predicate $A!$ in both exemplification and encoding predications.

4.2.1   What Is Abstract

First, we introduce the axioms that assert the existence of abstract objects of every type. Where α is a variable of type t, F is a variable of type $\langle t \rangle$, and $A!$ is a predicate of type $\langle t \rangle$, we assert:

> AXIOM 3: $\exists\alpha(A!\alpha \,\&\, \forall F(\alpha F \equiv \varphi))$, where φ has no free αs.

Here are three examples (or, rather, example schemes). In the first, x is a variable of type i, while $A!$ and F are of type $\langle i \rangle$. In the second, F is a variable of type $\langle i \rangle$, while $A!$ and $\mathcal{F}$ are of type $\langle\langle i \rangle\rangle$. In the third, R is a variable of type $\langle i, i \rangle$, while $A!$ and $\mathcal{R}$ are of type $\langle\langle i, i \rangle\rangle$:

> $\exists x(A!x \,\&\, \forall F(xF \equiv \varphi))$, where φ has no free xs
> $\exists F(A!F \,\&\, \forall\mathcal{F}(F\mathcal{F} \equiv \varphi))$, where φ has no free Fs
> $\exists R(A!R \,\&\, \forall\mathcal{R}(R\mathcal{R} \equiv \varphi))$, where φ has no free Rs

The first asserts that there exists an abstract individual that encodes all and only the properties of individuals that satisfy φ. The second asserts that there exists an abstract property of individuals that encodes all and only the properties of properties of individuals that satisfy φ. The third asserts that there exists an abstract relation among individuals that encodes exactly the properties of relations among individuals that satisfy φ.

Notice that from any instance of the above, we can derive the existence of a unique such object by an appeal to the definition of abstract object identity. Consider the second example above. There couldn't be two distinct abstract properties of individuals that encode exactly the properties of properties satisfying φ, since distinct abstract properties, by definition, have to differ by one of their encoded properties of properties.

Thus the above principles guarantee that the definite descriptions used in our three illustrative examples of theoretical identifications outlined in Section 2 are logically proper or well-defined (i.e., have denotations), since they are constructed in terms of formulas φ that have no free xs, Fs or Rs, respectively. In other words, the following are theorems:

$$\exists y(y = \iota x(A!x \,\&\, \forall F(xF \equiv \varphi)))$$
$$\exists G(G = \iota F(A!F \,\&\, \forall \mathcal{F}(F\mathcal{F} \equiv \varphi)))$$
$$\exists S(S = \iota R(A!R \,\&\, \forall \mathcal{R}(R\mathcal{R} \equiv \varphi)))$$

We call such descriptions *canonical* since for any formula φ (excluding only those with an inappropriate variable), the descriptions are guaranteed to have a denotation.

Moreover, the following Abstraction Principle is derivable as a theorem schema that governs canonical descriptions:

Abstraction Principle:
$\iota\alpha(A!\alpha \,\&\, \forall F(\alpha F \equiv \varphi))F \equiv \varphi$, where φ has no free αs.

As instances of this principle, we have:

$$\iota x(A!x \,\&\, \forall F(xF \equiv \varphi))F \equiv \varphi$$
$$\iota F(A!F \,\&\, \forall \mathcal{F}(F\mathcal{F} \equiv \varphi))\mathcal{F} \equiv \varphi$$
$$\iota R(A!R \,\&\, \forall \mathcal{R}(R\mathcal{R} \equiv \varphi))\mathcal{R} \equiv \varphi$$

The first theorem asserts that a specific abstract individual, namely the abstract individual encoding just the properties F such that φ, encodes a property F if and only if φ. The second theorem asserts that a specific abstract property of individuals, namely the abstract property encoding just the properties $\mathcal{F}$ such that φ, encodes a property $\mathcal{F}$ if and only φ. The third theorem asserts that a specific abstract relation, namely the abstract relation encoding just the properties $\mathcal{R}$ such that φ, encodes a property $\mathcal{R}$ if and only if φ.

The final axiom of encoding is that objects of type t which encode properties are abstract.[15] Where α is of type t, and F is of type $\langle t\rangle$, this axiom may be formalized as follows:

AXIOM 4: $\exists F\alpha F \to A!\alpha$

This implies, when $O!$ is of type $\langle t\rangle$, that $O!\alpha \to \neg\exists F\alpha F$. So, for example, if we add to our language the name s for the ordinary individual Socrates and take, as a premise, $O!s$, then AXIOM 4 implies that Socrates fails to encode properties.

4.2.2   What Isn't Abstract

Our remaining axioms tell us about what isn't abstract. Intuitively, abstract objects reify, at a lower level, higher-level patterns of properties already present in exemplification logic; they objectify the properties that satisfy higher-level conditions on properties. So, we conceive of abstract relations, of any type, as follows: (a) they encode properties, (b) they exemplify properties of relations and stand in relations among relations, but (c) *nothing* exemplifies them. (c) implies that if a relation is exemplified, it fails to be abstract. Where F is a

---

[15]AXIOM 4 does not say that objects are abstract *if and only if* they encode properties. This is because, for each type, there is a unique *null* abstract object that does not encode any properties.

variable of type $\langle t_1, \ldots, t_n \rangle$, $A!$ has type $\langle \langle t_1, \ldots, t_n \rangle \rangle$, and $\alpha_1, \ldots \alpha_n$ are distinct variables of type $t_1, \ldots, t_n$, respectively, then for $n \geq 0$, it is axiomatic that:

AXIOM 5: $\exists \alpha_1 \ldots \exists \alpha_n F \alpha_1 \ldots \alpha_n \rightarrow \neg A!F$

Note that in the case of the empty type $\langle \, \rangle$, this axiom implies $p \rightarrow \neg A!p$, i.e., that true propositions are not abstract (in the sense of being an abstract object that encodes properties), and hence that abstract propositions are false, i.e., that $A!p \rightarrow \neg p$.

AXIOM 5 implies that both sides of λ-Conversion will be false for abstract properties and abstract relations, since nothing ever exemplifies them. Intuitively, a property like $[\lambda x \, \varphi]$ is something that is exemplifiable by all and only the things satisfying φ, where φ expresses an exemplification pattern. But abstract properties and relations arise by comprehension, i.e., by what they encode, not by what exemplifies them. So λ-constructors build things that are apt for exemplification, whereas entities defined by what they encode aren't things that can be exemplified. Hence λ-expressions don't denote abstract objects. Thus we assert:

AXIOM 6: $\neg A![\lambda \nu_1 \ldots \nu_n \, \varphi]$ $(n \geq 1)$

We leave the formulation of examples of AXIOM 6 to the reader.

Finally, in the special case where F is a variable of type $\langle t_1, \ldots, t_n \rangle$, and $\alpha_1, \ldots, \alpha_n$ are distinct variables of type $t_1, \ldots, t_n$, respectively, and $A!$ has type $\langle \langle t_1, \ldots, t_n \rangle \rangle$, then we also assert that η-Conversion holds for λ-expressions in which φ is atomic and the 'head' relation is not an abstract relation:

AXIOM 7: $\neg A!F \rightarrow ([\lambda \alpha_1 \ldots \alpha_n \, F \alpha_1 \ldots \alpha_n] = F)$ $(n \geq 1)$

5   Application to Mathematics

To develop our logicist account of mathematics, we note first that by 'mathematics' we shall be focusing on *theoretical* as opposed to *natural* mathematics. *Natural* mathematics consists of the ordinary, pretheoretic claims that seem to be about mathematical objects, such as the following:

> The Triangle has 3 sides.
> The number of planets is eight.
> The class of insects is larger than the class of humans.
> Lines a and b have the same direction.

Theoretical mathematics, on the other hand, involves claims that occur in the context of some mathematical theory, whether or not the theory has been explicitly axiomatized, and whether or not the theory has been formalized. Examples of such claims are:

- The empty set is an element of the unit set of the empty set.
  [Said with reference to Zermelo-Fraenkel set theory.]

  - 2 is less than or equal to π.
      [Said with reference to real number theory.]

Though our framework can be applied to the analysis of both natural and theoretical mathematics, our present focus is only on the latter.[16]

  We shall assume, in what follows, that to produce a logicist account of theoretical mathematics, we have to show that arbitrary mathematical theories can be *reduced* to logic plus analytic truths. Our argument divides into two parts: (1) show that an arbitrary mathematical theory T can be reduced to the formal system described in Section 4 when supplemented with analytic truths, and (2) show that the formal system of Section 4 constitutes a logic. To achieve (1), we have to (a) assign the terms and predicates of T denotations that are describable in our framework, and (b) assign the theorems of T a reading in our system, involving those denotations, on which they are analytically true. If (1) and (2) succeed, then we can reap the epistemological benefits of logicism.

  In this section, we explain how the reduction of arbitrary mathematical theories is to be effected, and in the next section we argue that our formal framework is a logic. Though our framework is capable of analyzing mathematical theories of any finite order, recall that for simplicity, we are targeting first- and second-order mathematical theories having only primitive constants, variables, and 1- and 2-place predicates, but without function terms, definite descriptions, or n-ary predicates for $n > 2$. Though our system is set to handle more complex kinds of theories, we need not be distracted here by the extra details involved.

  Our first step shall be to analyze a mathematical theory as a *situation*, which was defined earlier as an abstract object that encodes only propositional properties. This analysis then motivates the definition at the end of Section 3, where we stipulated that p *is true in* T ('$T \vDash p$') means that T encodes the corresponding propositional property $[\lambda y\, p]$. It follows, as a theorem, that a mathematical theory T can be identified as follows:

$$T = \iota x(A!x \;\&\; \forall F(xF \equiv \exists p(T \vDash p \;\&\; F = [\lambda y\, p]))),$$

where we read $\vDash$ as having the smallest scope, so that $T \vDash \varphi \;\&\; \psi$ is understood as $(T \vDash \varphi) \;\&\; \psi$. In other words, a theory T is the abstract individual that encodes exactly the properties F such that there is a proposition p true in T for which F is *being such that p*. Now if we add constants of type i to our logical framework to denote what we pretheoretically judge to be mathematical theories (such as 'ZF' for Zermelo-Fraenkel set theory, '$\mathbb{R}$' for real number theory, 'PA' for Peano Arithmetic, etc.), then statements of the form ZF $\vDash p$, $\mathbb{R} \vDash q$, PA $\vDash r$, etc., become well-formed.[17] Moreover, as an instance of the above theorem, it follows that:

---

[16]For a discussion of the former, and how the second-order modal version of the above framework can analyze such terms as 'The Triangle', 'the number of planets', 'the class of insects', etc., see Pelletier & Zalta 2000, Zalta 1999, and Anderson & Zalta 2004.

[17]Thus, these new constants denote objects that encode the propositional content of the systems that have been formulated, by axioms and rules, in a syntactically second-order language.

$$\text{ZF} = \iota x(A!x \ \& \ \forall F(xF \equiv \exists p(\text{ZF} \vDash p \ \& \ F = [\lambda y \ p])))$$

Similar identifications can be given for other mathematical theories.

The mechanism by which statements of the form $T \vDash p$ become assertible is as follows. Consider an arbitrary mathematical theory T.[18] We import the theorems of T into our framework by appeal to the following Importation Principle (later we argue that the resulting claims are theory-relative analytic truths):[19]

> **Importation Principle**. When φ is a closed theorem of T, then $T \vDash \varphi^*$ shall be an axiom, where (a) φ^* is the result of indexing, to T, all the *closed primary terms*[20] of φ, and (b) whenever τ is a term of T having type t, then the indexed term τ_T is a constant term of the same type as τ.

One consequence of the Importation Principle should be noted: if theoremhood in one of the mathematical theories to be imported is not decidable, the set of axioms resulting from Importation will not be decidable either. Usually, this is to be avoided, since if the set of axioms of the present theory is not decidable, there is no decision procedure for whether a sequence of formulas from the language of our theory is a derivation in the theory. However, since imported axioms don't have the same status as the axioms formulated in Section 4 (as we shall argue, they are analytic rather than logical truths), we simply require the following: in order to use an axiom $T \vDash \varphi^*$ arising from importation within a derivation, one must have a proof of φ within the imported theory T that demonstrates it is a theorem.[21]

Before we explain the formal specifics of the Importation Principle, we would like to highlight from the start what it is going to achieve and what it will *not* achieve. One way of thinking of Importation is as a kind of "positive internalization" principle by which theoremhood in a mathematical theory gets internalized in object theory by means of suitable encoding statements:

$$\text{If } T \vdash \varphi, \text{ then } T \vDash \varphi^* \text{ (that is, } T[\lambda y \ \varphi^*]) \text{ is an axiom.} \tag{PI}$$

However, we will *not* presuppose a corresponding "negative internalization" principle of the form:

---

[18] In what follows, we engage in a harmless abuse of notation; the expression 'T' is used sometimes as a variable ranging over what we pretheoretically judge to be a mathematical theory, while at other times 'T' is used technically as a variable ranging over mathematical theories analyzed object-theoretically. Sometimes the expression is used both ways within the same context, as in the Importation Principle below.

[19] For pre-axiomatized or non-axiomatized theories, the following principle assumes that there is at least some body of truths that constitute the theory. Those are the theorems to be imported.

[20] We defined *primary* terms in Section 3. By saying 'closed', we are excluding the simple variables. We've also assumed that we're formulating mathematical theories with closed λ-expressions, and so there won't be indexed λ-expressions with free variables.

[21] Alternatively, we could address this issue by revising the Importation Principle so that (a) we only import the *axioms* of T and (b) stipulate that $T \vDash \varphi^*$ within object theory is to be closed under the deductive consequences of φ^* that are formulable within the indexed version of the original language of T. This alternative does, however, not provide an analysis of theories that aren't axiomatized.

If $T \nvdash \varphi$, then $T \nvDash \varphi^*$ (that is, $\neg T[\lambda y\, \varphi^*]$) is an axiom. $\qquad$ (NI)

Instead of (NI) we only have the following non-importation principle:

> **Non-Importation Principle**. If $T \nvdash \varphi$, then it is *not* the case that $T \vDash \varphi^*$ is an axiom.

One reason why we do not want to presuppose (NI) is that if a recursively axiomatized theory T is not decidable – in the manner of PA and ZF – then the non-theorems of T are not recursively enumerable. Consequently, we would not have a systematic method of using (NI) to generate negative axioms of the form $\neg(T \vDash \varphi^*)$ in our own theory even if we exploited the derivation of theorems within the imported theory T.

This does not mean that we do not find (NI) – or at least suitable instances thereof, or alternative axioms that might entail suitable instances thereof – to be true and plausible. It is just that, in this paper, we have not formulated such "negative internalization" principles that conform to the same standards of deductive rigor as the Importation Principle. As a result, the axioms of our present theory do not exclude certain unintended interpretations on which an abstract individual (theory) T encodes "more" propositional properties of the form $[\lambda y\, \varphi^*]$ than it should. We leave the systematic study of negative variants of Importation and their consequences to future work.[22] Note that a Carnapian might extend the Importation Principle in a way a Platonist might reject,[23] namely, by the addition of (a) axioms that assert $\tau_T \neq \tau_{T'}$ when $T \neq T'$ and (b) axioms that assert $\varphi^* \neq \psi^*$ when φ^* is imported from T and ψ^* is imported from a distinct theory T'.

Given the Importation Principle as stated, consider the fact that

$$\vdash_{\mathrm{ZF}} \neg \exists y(Sy \,\&\, y \in \emptyset)$$

The closed primary terms of this theorem are S, $\in$, and $\emptyset$, and so the following statement will be an axiom of our framework:

$$\mathrm{ZF} \vDash \neg \exists y(S_{\mathrm{ZF}}y \,\&\, y \in_{\mathrm{ZF}} \emptyset_{\mathrm{ZF}}) \qquad (\mathrm{G})$$

In what follows, we use T-*indexed terms* or *indexed terms* to refer to the terms introduced into object theory by means of the Importation Principle. We assume, for every such theory T and type t other than the type for propositions, there is a distinguished 2-place identity predicate, $=_T$ having type $\langle t, t \rangle$, which applies to the entities of type T in the usual way: it is reflexive and governed by the principle of substitution of identicals.[24] Of course, if T uses a non-standard relation of identity, we defer to the axioms that T uses for this relation. Moreover, we shall have no need of an identity relation on propositions when formulating mathematical theories – it is typically no part of mathematics to be concerned with the identities among propositions.

---

[22]We are grateful to an anonymous reviewer for urging us to comment on this.

[23]For more on Platonism, see the last paragraph of Section 7.1.

[24]See Leitgeb & Ladyman (2008) for an argument that each mathematical structure includes an identity relation specific to that structure.

It is important to pause here to say something about expressivity in the target mathematical theories we are going to analyze. Our goal is to analyze not only the individual terms of mathematical theories but also their relation terms. As far as we know, the best way to identify the relations denoted by the relation terms of mathematical theories is by the properties they exemplify in their respective theories. However, most mathematical theories are not formulated in such a way that they explicitly enable talk about the properties of relations. Consider that, in ZF, you can't talk about the properties of $\in$; e.g., you can't say that, in ZF, $\in$ exemplifies the property $[\lambda R \, \neg\exists y(Sy \,\&\, yR\emptyset)]$. So, in what follows, we shall assume that mathematical theories have been formulated in a formal system that includes (closed) higher-order λ-expressions. Such expressions allow us to talk, within those theories, about the properties of relations. Thus, we will be representing T by way of a conservative extension in which T is formulated with λ-expressions and is closed under the axioms of the relational λ-calculus (including a version of AXIOM 2). From such a formulation, we can abstract out, from T, the properties that are exemplified in T by the relations of T.

Consequently, we may suppose that the following are theorems of ZF:

$\vdash_{\mathrm{ZF}} [\lambda x \, \neg\exists y(Sy \,\&\, y \in x)]\emptyset$

$\vdash_{\mathrm{ZF}} [\lambda F \, \neg\exists y(Fy \,\&\, y \in \emptyset)]S$

$\vdash_{\mathrm{ZF}} [\lambda R \, \neg\exists y(Sy \,\&\, yR\emptyset)]\in$

When these theorems are imported into object theory, we index only the relation term and the argument term:[25]

$$\mathrm{ZF} \vDash [\lambda x \, \neg\exists y(Sy \,\&\, y \in x)]_{\mathrm{ZF}}\emptyset_{\mathrm{ZF}} \tag{H}$$

$$\mathrm{ZF} \vDash [\lambda F \, \neg\exists y(Fy \,\&\, y \in \emptyset)]_{\mathrm{ZF}}S_{\mathrm{ZF}} \tag{I}$$

$$\mathrm{ZF} \vDash [\lambda R \, \neg\exists y(Sy \,\&\, yR\emptyset)]_{\mathrm{ZF}}\in_{\mathrm{ZF}} \tag{J}$$

Thus, not only does object theory become extended with (G), i.e., in ZF, no set is a member of the nullset, but also with the (H), (I), and (J), which assert that: (H) in ZF, the ZF null set exemplifies the ZF property *having no sets as members*; (I) in ZF, the ZF property of being a set exemplifies the ZF property of properties *being an F such that nothing exemplifying F is an element of the null set*; and (J) in ZF, the ZF membership relation exemplifies the ZF property of relations *being an R such that no set bears R to the null set*.

Note here that we have introduced a new kind of λ-expression; these indexed λ-expressions will be treated somewhat differently from the primitive λ-expressions of the language: they do not denote ordinary relations and are not subject to AXIOM 2 (but see below). Instead, they will be subject to the

---

[25]This corrects the procedure in Zalta 2000 and 2006. In the former, the λ-expressions denoting mathematical properties and relations weren't indexed (but should have been). And in both works, we indexed not just the primary terms but also terms inside λ-expressions. In the present work, however, we've come to realize that we need not do so. Indeed, it seems more perspicuous to index only the primary terms, since this way, we don't need to double-index both λ-expressions and the terms inside them. We need not index expressions that themselves include indexed terms, but rather index only terms that could appear in the mathematical theory in question without its index.

Reduction Axiom Schema discussed in the next section, which precisely identi-fies their denotations as abstract relations.[26]

We now turn to the special axioms that identify denotations of the theoretical primitives (predicates and individual terms) of mathematical theories.

5.1   The Denotations of the Terms of T

We can now say what the constants and λ-expressions of T denote in our back-ground theory. Thus far, we've assumed T includes primitive constants, prim-itive 1-place predicates, primitive 2-place predicates (including identity), and λ-expressions. Where τ is any primitive individual constant of T, any primitive 1-place predicate constant of type $\langle i \rangle$ of T, any 2-place predicate constant of type $\langle i, i \rangle$ of T, or any 1-place λ-expression of type $\langle t \rangle$ of T (for any type t), let τ_T be the T-indexed version of τ. By the Importation Principle, these indexed terms have the same type as their non-indexed counterparts. We now turn to the question of what these indexed terms denote.

We shall later argue (Section 6.2) for the view that the meaning of a mathe-matical term τ in theory T is the logical role it has within T. But we here assert an axiom that captures this view by stipulating that term τ of type t in theory T denotes the abstract object of type t that encodes exactly the properties (of type $\langle t \rangle$) that τ_T exemplifies in T. Formally, we assert the following Reduction Axiom Schema, which uses canonical descriptions to identify the denotations of the indexed mathematical terms imported into object theory. Where τ_T and α have type t, $A!$ and F have type $\langle t \rangle$:

> **Reduction Axiom Schema:**
> $\tau_T = \iota\alpha(A!\alpha \ \& \ \forall F(\alpha F \equiv \ T \vDash F\tau_T))$

Note that the instances of this schema are *not* definitions, since the expressions on the left of the identity sign also appear on the right. But they are principles that are analytic, or so we will argue in the next section. Here are some *simple* examples of the above; these tell us exactly which abstract objects are denoted by $\emptyset_{\mathrm{ZF}}$, S_{ZF}, $\in_{\mathrm{ZF}}$ (later we'll discuss some more complex examples):

> **Instances of the Reduction Axiom Schema:**
> $\emptyset_{\mathrm{ZF}} = \iota x(A!x \ \& \ \forall F(xF \equiv \ \mathrm{ZF} \vDash F\emptyset_{\mathrm{ZF}}))$

---

[26] We could have also added, as an example:

> $\mathrm{ZF} \vDash [\lambda FRx \ \neg\exists y(Fy \ \& \ yRx)]_{\mathrm{ZF}} S_{\mathrm{ZF}} \in_{\mathrm{ZF}} \emptyset_{\mathrm{ZF}}$

This asserts that in ZF, the ZF entities S, $\in$, and $\emptyset$ exemplify the ZF relation: being a property F, a relation R, and an object x such that nothing exemplifying F bears R to x. Similarly, we could add as an example from PA:

> $\mathrm{PA} \vDash [\lambda R \ R23]_{\mathrm{PA}}[\lambda xy \ x + y = 5 \ \& \ x < y]_{\mathrm{PA}}$

This asserts that the PA relation, being an x and y such that $x + y = 5$ and such that $x < y$, exemplifies the PA property of being a relation that relates 2 to 3.

In both of these cases, we need only apply the techniques discussed in the following section in the text, whereby mathematical relations are analyzed as abstract relations that encode just the properties of relations attributed to them in the theory. But we have omitted these examples for simplicity.

$$S_{\mathrm{ZF}} = \iota F(A!F \,\&\, \forall \mathcal{F}(F\mathcal{F} \equiv \mathrm{ZF} \vDash \mathcal{F}S_{\mathrm{ZF}}))$$
$$\in_{\mathrm{ZF}} = \iota R(A!R \,\&\, \forall \mathcal{R}(R\mathcal{R} \equiv \mathrm{ZF} \vDash \mathcal{R}\in_{\mathrm{ZF}}))$$

These all have obvious readings.

We next focus just on the identification of the primitive constants and predicates of a particular theory, so that we can more easily see their consequences. The following Equivalence Theorem Schema is an immediate consequence of our Reduction Axiom Schema, by the Abstraction Principle for abstract objects and substitution of identicals:[27]

Equivalence Theorem Schema:
$$\forall F(\tau_T F \equiv T \vDash F\tau_T)$$

This asserts that a term τ (individual or relation) of theory T encodes exactly the properties τ exemplifies in T. As somewhat more specific examples of this schema, we have:

$$\forall G(\kappa_T G \equiv T \vDash G\kappa_T)$$
$$\forall \mathcal{G}(\Pi^1_T \mathcal{G} \equiv T \vDash \mathcal{G}\Pi^1_T)$$
$$\forall \mathcal{S}(\Pi^2_T \mathcal{S} \equiv T \vDash \mathcal{S}\Pi^2_T)$$

In other words, for any first-level property G, the individual κ_T encodes G iff κ_T exemplifies G in T; for any second-level property of properties $\mathcal{G}$, the property Π^1_T encodes $\mathcal{G}$ iff Π^1_T exemplifies $\mathcal{G}$ in T, and for any second-level property of relations $\mathcal{S}$, the property Π^2_T encodes $\mathcal{S}$ iff Π^2_T exemplifies $\mathcal{S}$ in T.

Clearly, then, the following are instances of the Equivalence Theorem schema:

$$\forall G(\emptyset_{\mathrm{ZF}} G \equiv \mathrm{ZF} \vDash G\emptyset_{\mathrm{ZF}})$$
$$\forall \mathcal{G}(S_{\mathrm{ZF}}\mathcal{G} \equiv \mathrm{ZF} \vDash \mathcal{G}S_{\mathrm{ZF}})$$
$$\forall \mathcal{S}(\in_{\mathrm{ZF}}\mathcal{S} \equiv \mathrm{ZF} \vDash \mathcal{S}\in_{\mathrm{ZF}})$$

That is, the empty set of ZF encodes exactly the properties G that it exemplifies in ZF; the ZF-property of being a set encodes exactly the second-level properties of properties that it exemplifies in ZF; and the membership relation of ZF encodes exactly the second-level properties of relations that it exemplifies in ZF.

Now the axioms introduced by the Importation Principle become salient. For the properties that can be abstracted from those claims can be instantiated into the above universal claims. In particular, the properties that are referenced in (H), (I), and (J) above may be instantiated, respectively, into the above claims to yield the following theorems:

---

[27] Here, for example, is a proof of the second theorem. The Abstraction Principle for abstract properties has the following instance, where Π^1_T is any primitive 1-place predicate of T, and $\mathcal{G}$ is a free variable ranging over second-level properties of properties:

$$\iota F(A!F \,\&\, \forall \mathcal{F}(F\mathcal{F} \equiv T \vDash \mathcal{F}\Pi^1_T))\mathcal{G} \equiv T \vDash \mathcal{G}\Pi^1_T$$

But by the relevant Reduction Axiom, it follows by substitution of identicals that:

$$\Pi^1_T \mathcal{G} \equiv T \vDash \mathcal{G}\Pi^1_T$$

We can now universally generalize on the free variable $\mathcal{G}$ to reach our theorem.

$$\emptyset_{\text{ZF}}[\lambda x \, \neg \exists y (Sy \, \& \, y \in x)]_{\text{ZF}} \; \equiv \; \text{ZF} \vDash [\lambda x \, \neg \exists y (Sy \, \& \, y \in x)]_{\text{ZF}} \emptyset_{\text{ZF}}$$
$$S_{\text{ZF}}[\lambda F \, \neg \exists y (Fy \, \& \, y \in \emptyset)]_{\text{ZF}} \; \equiv \; \text{ZF} \vDash [\lambda F \, \neg \exists y (Fy \, \& \, y \in \emptyset)]_{\text{ZF}} S_{\text{ZF}}$$
$$\in_{\text{ZF}}[\lambda R \, \neg \exists y (Sy \, \& \, yR\emptyset)]_{\text{ZF}} \; \equiv \; \text{ZF} \vDash [\lambda R \, \neg \exists y (Sy \, \& \, yR\emptyset)]_{\text{ZF}} \in_{\text{ZF}}$$

Each primary term in the biconditionals displayed above has been given a formal identification in our theory. Moreover, since the right-hand side of each of the above equivalences is a theorem resulting from the Importation Principle, we have a proof of the following facts about $\emptyset_{\text{ZF}}$, S_{ZF}, and $\in_{\text{ZF}}$:

$$\emptyset_{\text{ZF}}[\lambda x \, \neg \exists y (Sy \, \& \, y \in x)]_{\text{ZF}}$$
$$S_{\text{ZF}}[\lambda F \, \neg \exists y (Fy \, \& \, y \in \emptyset)]_{\text{ZF}}$$
$$\in_{\text{ZF}}[\lambda R \, \neg \exists y (Sy \, \& \, yR\,\emptyset)]_{\text{ZF}}$$

In other words, it is provable in our framework that the empty set of ZF encodes the ZF-property of having no sets as members; the ZF-property of being a set encodes the (second-level) ZF-property of being a property such that nothing exemplifying it is a member of the empty set; and the membership relation of ZF encodes the (second-level) ZF-property of being a relation that no set bears to the empty set.

We conclude this section with two, somewhat more complex, examples. First, recall (H):

$$\text{ZF} \vDash [\lambda x \, \neg \exists y (Sy \, \& \, y \in x)]_{\text{ZF}} \emptyset_{\text{ZF}} \tag{H}$$

We may now use the Reduction Axiom to identify the denotation of the λ-expression as follows:

$$[\lambda x \, \neg \exists y (Sy \, \& \, y \in x)]_{\text{ZF}} =$$
$$\iota F(A!F \, \& \, \forall \mathcal{F}(F\mathcal{F} \equiv \text{ZF} \vDash \mathcal{F}[\lambda x \, \neg \exists y (Sy \, \& \, y \in x)]_{\text{ZF}}))$$

This asserts that the ZF-property *having no sets as members* is identical to the abstract property that encodes all and only those second-level properties that are exemplified, in ZF, by the ZF-property of *having no sets as members*. As an example of such an encoded second-level property, consider *being a property exemplified by the null set* ($[\lambda F \, F\emptyset]_{\text{ZF}}$). It is a fact about ZF that:

$$\vdash_{\text{ZF}} [\lambda F \, F\emptyset][\lambda x \, \neg \exists y (Sy \, \& \, y \in x)]$$

and this gets imported as:

$$\text{ZF} \vDash [\lambda F \, F\emptyset]_{\text{ZF}}[\lambda x \, \neg \exists y (Sy \, \& \, y \in x)]_{\text{ZF}}$$

That is, in ZF, the ZF-property *having no sets as members* exemplifies the ZF-property of properties *being a property exemplified by the null set.* So by the Equivalence Theorem Schema:

$$[\lambda x \, \neg \exists y (Sy \, \& \, y \in x)]_{\text{ZF}} \text{ encodes } [\lambda F \, F\emptyset]_{\text{ZF}}$$

For the final example, recall (J):

$$\text{ZF} \vDash [\lambda R \, \neg \exists y (Sy \, \& \, y \, R \, \emptyset)]_{\text{ZF}} \in_{\text{ZF}} \tag{J}$$

So, where R is a variable of type $\langle i, i \rangle$, α is a variable of type $\langle\langle i, i \rangle\rangle$, $A!$ has type $\langle\langle\langle i, i \rangle\rangle\rangle$, and Γ is a variable of type $\langle\langle\langle\langle i, i \rangle\rangle\rangle\rangle$, the Reduction Axiom Schema implies:

$$[\lambda R \, \neg\exists y(Sy \, \& \, y \, R \, \emptyset)]_{\mathrm{ZF}} =$$
$$\iota\alpha(A!\alpha \, \& \, \forall\Gamma(\alpha\Gamma \equiv \, \mathrm{ZF} \vDash \Gamma[\lambda R \, \neg\exists y(Sy \, \& \, y \, R \, \emptyset)]_{\mathrm{ZF}}))$$

We leave to the reader the formulation of a natural language gloss of this identification. And we leave it to the reader to find examples of properties that are exemplified by $[\lambda R \, \neg\exists y(Sy \, \& \, y \, R \, \emptyset)]$ in ZF. By the Equivalence Theorem, these become encoded by $[\lambda R \, \neg\exists y(Sy \, \& \, y \, R \, \emptyset)]_{\mathrm{ZF}}$.

5.2   Sentence Reduction: True Readings of Mathematical Theorems

Since we won't be using relative interpretability as our standard of reduction, our methodology is to outline an alternative translation procedure that yields a true reading for every unprefixed theorem of each mathematical theory. We therefore show how to assign true object-theoretic readings to the theorems of mathematical theories when we consider those theorems in and of themselves, unprefixed by a theory operator. Platonist philosophers of mathematics believe that the unadorned claims of mathematics, such as '0 is a number', 'the empty set is an element of unit set of the empty set', 'two is less than π', etc., are simply true, while fictionalist philosophers argue that they are false. Our view is that this disagreement is explained by the fact that these claims are ambiguous, for there is an exemplification reading on which they are false and an encoding reading on which they are true.

Take a simple atomic formula, e.g., the statement that '0 is a number', when this is asserted as an axiom of Peano Arithmetic. We've already seen that the prefixed claim "In Peano Arithmetic, 0 is a number" is to be represented as:

$$\vdash_{\mathrm{PA}} N0$$

After importing the above into object theory, our analysis is:[28]

$$\mathrm{PA} \vDash N_{\mathrm{PA}}0_{\mathrm{PA}}$$

Now we want to give a true reading of the unprefixed "0 is a number". But we can *infer* such a reading from an instance of the Equivalence Theorem, since the encoding formula, $0_{\mathrm{PA}}N_{\mathrm{PA}}$, is derivable. As a theorem, object theory regards $0_{\mathrm{PA}}N_{\mathrm{PA}}$ as a *true* reading of "0 is a number". No such argument can be given for the exemplification reading, $N_{\mathrm{PA}}0_{\mathrm{PA}}$, of the unadorned claim "0 is a number". In our framework, this exemplification claim is axiomatically false, by the contrapositive of AXIOM 5 and the fact that N_{PA} is an abstract property

---

[28]Our analysis is not a form of deductivism (if-thenism); the following claim is not a conditional, but rather a categorical (non-conditional) encoding claim of the form 'xF'. Moreover, we provide categorical readings on which unprefixed mathematical claims are true (see below). Such readings are not a part of deductivism.

of individuals. This example shows that predications of the form 'x is F' in natural language are structurally ambiguous, and that in the case at hand, the encoding reading xF is provably true while the exemplification reading Fx is false.

Moreover, we take there to be a structural ambiguity in simple predications of natural language, for the unadorned claim "0 is a number" embodies not only a true atomic fact about a property that 0_{PA} encodes but also a true atomic fact about a property that N_{PA} encodes, namely, $[\lambda F\ F0]_{\text{PA}}$. Once we import $\vdash_{\text{PA}} [\lambda F\ F0]N$ so as to yield the axiom PA $\vDash [\lambda F\ F0]_{\text{PA}} N_{\text{PA}}$, the Equivalence Theorem guarantees that the encoding formula $N_{\text{PA}}[\lambda F\ F0]_{\text{PA}}$ is a theorem of our logical framework. Thus, we have another true reading of our unadorned mathematical claim. In a manner similar to the above, the exemplification reading, that N_{PA} exemplifies $[\lambda F\ F0]_{\text{PA}}$, is axiomatically false, by AXIOM 5 and the fact that $[\lambda F\ F0]_{\text{PA}}$ is an abstract property of properties.

To see how to generalize the procedure for assigning true encoding readings for complex (unadorned) mathematical theorems, let us return to the statement, "No set is an element of the empty set", said in the context of ZF. We've seen how the representation of the claim that this is a theorem of ZF gets imported as the following theorems of our logical framework (the first of which is an axiom by importation, the remainder of which are consequences):

$$\text{ZF} \vDash \neg\exists y(S_{\text{ZF}}y\ \&\ y \in_{\text{ZF}} \emptyset_{\text{ZF}}) \tag{G}$$
$$\text{ZF} \vDash [\lambda x\ \neg\exists y(Sy\ \&\ y \in x)]_{\text{ZF}}\emptyset_{\text{ZF}} \tag{H}$$
$$\text{ZF} \vDash [\lambda F\ \neg\exists y(Fy\ \&\ y \in \emptyset)]_{\text{ZF}}S_{\text{ZF}} \tag{I}$$
$$\text{ZF} \vDash [\lambda R\ \neg\exists y(Sy\ \&\ y\,R\,\emptyset)]_{\text{ZF}}\in_{\text{ZF}} \tag{J}$$

But our question is, what reading should we assign to the unadorned ZF claim that "No set is an element of empty set"? Our answer is that we should assign the *conjunction* of the following three theorems:

$$\emptyset_{\text{ZF}}[\lambda x\ \neg\exists y(Sy\ \&\ y \in x)]_{\text{ZF}}$$
$$S_{\text{ZF}}[\lambda F\ \neg\exists y(Fy\ \&\ y \in \emptyset)]_{\text{ZF}}$$
$$\in_{\text{ZF}}[\lambda R\ \neg\exists y(Sy\ \&\ yR\emptyset)]_{\text{ZF}}$$

These were all proved to be theorems at the end of Section 5.1. We suggest that the *conjunction* of all three theorems captures the facts embodied by the unadorned mathematical claim "No set is an element of the empty set", for this latter claim is not only a fact about the empty set of ZF, but also a fact about the ZF-property of being a set and about the membership relation of ZF.

Indeed, we can, in general, *abbreviate* the conjunction of the above three theorems as a simple, intuitive formula:

$$\emptyset_{\text{ZF}}S_{\text{ZF}}\in_{\text{ZF}}[\lambda xFR\ \neg\exists y(Fy\ \&\ yRx)]_{\text{ZF}}$$

We can intuitively think of this as a ternary encoding formula of the form $xyzH$, where the type of H is that of a ternary relation among entities having the types of x, y, z in that order. Of course, this isn't a primitive formula of our logical framework, but it doesn't need to be, for it simply serves as an abbreviation

of a conjunction of well-formed formulas. In the particular case at hand, three places are needed because there are three primitive theoretical expressions in our target sentence ($\emptyset$, S, $\in$). But it is straightforward to define a function that takes as input an unadorned theorem of ZF and yields as output a n-place encoding formula of the above form, where the output encoding formula is an abbreviation of n well-formed encoding formulas. We omit the details here, though interested readers are directed to Zalta 2000 (250–251).

This, then, is a procedure for assigning a true (T-relative) reading to every theorem of an arbitrary mathematical theory T. It completes the reduction of mathematics to the above axiomatic system. In the next section we will show that each of the axioms of our system counts as logical or analytic.

We note here that the procedure and analysis described above offers a logical reconstruction of mathematical objects and relations as they are given antecedently by some specific mathematical theory. While this provides, for example, a complete reconstruction of ZF-sets, we are not claiming that the reconstruction is necessarily a complete theory of sets since ZF isn't a complete theory of sets. Rather, we are offering the above as an analysis of any claim a mathematician might make in any context in which the mathematician is adopting all and only the assumptions of ZF. And this generalizes to other mathematical theories: in every context in which a mathematician assumes the principles of a mathematical theory, we can use the above methods to analyze their claims.

One final observation, about the completeness of mathematical objects and relations, is in order. Recall that our analysis imports theorems φ of T, i.e., formulas φ such that $\vdash_T \varphi$, as claims of the form $T \vDash \varphi^*$. But for incomplete theories T, a property of the derivability relation $\vdash$ now becomes relevant, namely, that $\vdash_T (\varphi \lor \psi)$ doesn't imply $\vdash_T \varphi$ or $\vdash_T \psi$. Given our methodology above, this extends object theory with claims of the form $T \vDash (\varphi \lor \psi)^*$ but not with claims of the form $T \vDash \varphi^*$ or claims of the form $T \vDash \psi^*$. For example, consider the Continuum Hypothesis (CH), where this is formulated as $2^{\aleph_0} = \aleph_1$. Since $\vdash_{\mathrm{ZF}} (\mathrm{CH} \lor \neg\mathrm{CH})$ doesn't imply $\vdash_{\mathrm{ZF}} \mathrm{CH}$ or $\vdash_{\mathrm{ZF}} \neg\mathrm{CH}$, the object-theoretic claim $\mathrm{ZF} \vDash (\mathrm{CH} \lor \neg\mathrm{CH})^*$ doesn't imply $\mathrm{ZF} \vDash \mathrm{CH}^*$ or $\mathrm{ZF} \vDash (\neg\mathrm{CH})^*$. Moreover, though it follows that $\aleph_{1_{\mathrm{ZF}}}$ encodes $[\lambda x (2^{\aleph_0} = x \lor \neg 2^{\aleph_0} = x)]_{\mathrm{ZF}}$, it doesn't follow either that $\aleph_{1_{\mathrm{ZF}}}$ encodes $[\lambda x\, 2^{\aleph_0} = x]_{\mathrm{ZF}}$ or that $\aleph_{1_{\mathrm{ZF}}}$ encodes $[\lambda x\, \neg(2^{\aleph_0} = x)]_{\mathrm{ZF}}$.

Note that this example highlights an important difference between the standard of relative interpretability and our method of sentence reduction. The relative interpretation of a disjunction $\varphi \lor \psi$ is the disjunction of the respective relative interpretations of φ and ψ, while our reduction of a disjunctive mathematical theorem $\varphi \lor \psi$ does not necessarily coincide with the disjunction of the respective reductions of φ and ψ. Indeed, if neither φ itself nor ψ itself is a theorem, then our reduction will not apply to φ and ψ at all.

These facts will help us to argue that this is a form of logicism – the objects and relations are reified incomplete concepts. Even 'complete' mathematical theories (e.g., the first-order theory of real-closed fields) are, in some sense, about objects that have only mathematical properties and are thus incomplete with respect to what they encode. And even '(deductively) incomplete' theories

are complete in the sense that they completely describe the incomplete entities they are about.

6   Why This is Logicism

In this section, we argue that our analysis of mathematics satisfies the definition of *logicism*, as given below. As part of our argument we establish that our logical framework consists of axioms that are logical or analytic (Section 6.1), and then establish that the axioms needed to assign denotations and truth conditions to mathematical theorems are analytic (Section 6.2). Our usage of 'concept' and 'object' in what follows will not be the standard ones. Traditionally, mathematical individuals are referred to as 'mathematical objects' and mathematical properties and relations are referred to as 'mathematical concepts'. But in our type-theoretic framework, properties, relations, and propositions are also considered as objects, i.e., as entities of which we predicate properties. Moreover our uniform analysis of mathematical individuals and mathematical properties and relations allows us to talk about all of these mathematical objects as mathematical concepts. Similarly, we will use the terms 'logical object' and 'logical concept' interchangeably.

Logicism, historically, is the claim that every true mathematical proposition is derivable from the laws of logic extended with analytic truths such as definitions.[29] Since we are focusing only on theoretical mathematics, logicism can be restated as the following clearer and simpler thesis: all mathematical theorems are derivable from the laws of logic extended with analytic truths, where by 'mathematical theorems' we mean any statement that is part of or derivable from any mathematical theory.

Moreover, in the Frege-Russell tradition, logicism consisted of an additional claim, to the effect that mathematical concepts are (analyzable in terms of) logical concepts.[30] Thus, we may understand logicism as the conjunction of the following two theses (Carnap 1931):

LC Logicism about Mathematical Concepts: Every mathematical concept denoted by a mathematical term is (identical to) a logical concept denoted by a logical term.

LT Logicism about Mathematical Theorems: For every mathematical theorem, if each mathematical term denoting a mathematical concept in any such theorem is replaced by a logical term denoting the logical concept identical to the mathematical concept, then the resulting theorem is logically or analytically true.

---

[29]Some philosophers, e.g., Roeper 2015 and Klev 2017, take logicism to be the narrower claim that *arithmetic* is reducible to logic, but we regard logicism to be more broadly conceived.

[30]Here we shall be talking about *well-defined* mathematical concepts. We take the *well-defined* mathematical concepts of a theory T to be those represented by a term (i.e., an individual term or a predicate) of T that is either primitive or uniquely definable in T.

Clearly, as we have formulated these two principles, LT presupposes LC.

Logicism has traditionally been formulated primarily as a matter of logical truth, and not logical consequence, since the emphasis has been on reducing mathematical claims to logical truths and not on showing that the consequences mathematicians infer from mathematical principles are purely logical. But we want our conception of logicism to extend to the idea that mathematical practice involves a body of inferences, so that logicism also encompasses the idea that mathematical truths can be derived as logical consequences of a logic (cf. Rayo 2005, 204). However, in what follows, we focus primarily on LC and LT and, along the way, note how our understanding yields logicism with respect to the notion of logical consequence defined below.

Our plan, then, is as follows:

- First, we argue that the axioms presented in Section 4 are all either logical truths or analytic. In the case where we take the axioms to be logically true, we shall argue for their logicality by putting forward what we take to be a correct conception of logical truth, and then showing that under that conception, these axioms are logical truths. It is a consequence of our argument that the abstract objects picked out by our canonical descriptions are logical objects.

- Second, we argue that the additional axioms, put forward in Section 5 for the analysis of mathematical language, are all analytic.

- From the conclusions of these two arguments, we may then immediately conclude that the theorems of mathematics—represented in object theory as explained in Section 5.2—are logical or analytic truths, since logical consequences of logical and analytic truths are either logical or analytic. Moreover, we shall see (in Section 6.2) that LC follows as well, namely, that every mathematical concept denoted by a mathematical term is (identical to) a logical object.

6.1   Our Framework Axioms Are Logical or Analytic

In this section, we shall not argue, but rather assume, that the principles of classical logic, the substitution of identicals, the axiom governing descriptions, and the principle of λ-Conversion (i.e., the axioms discussed in Section 4 prior to Section 4.2) are logically true (as this notion is defined below) or analytic, where analyticity is defined in the usual way as truth in virtue of meaning (in this case, of the logical symbols). The principles of classical logic and the substitution of identicals have traditionally been regarded as logical. We add the law governing descriptions and λ-Conversion (AXIOMS 1 and 2) to this list of logical truths on the grounds that they are true in virtue of the meaning of the expressions *the* (represented by the ι) and *being such that* (represented by the λ).[31]

---

[31]We recognize that λ-Conversion has existential import; one can derive existence claims from it, namely, the comprehension principle for relations. That is, one can derive $\exists F \forall x (Fx \equiv \varphi)$ from $[\lambda x\, \varphi]x \equiv \varphi$. But having existential import hasn't disqualified the laws of classical

Moreover, AXIOMS 4–7 (in Section 4.2) stipulate what is meant by the property of *being abstract* and, as such, are nothing more than meaning postulates. Once we take *abstract* objects to be those entities that are individuated by the properties they encode, these axioms articulate the conception of such objects in formal detail and, hence, are analytically true. Thus, we see the burden of the present paper as showing that the comprehension principle for abstract objects (AXIOM 3) is a logical truth, despite the fact that it baldly asserts the existence of abstract objects.

We begin with the observation that the classical understanding of the model-theoretic interpretation of the predicate calculus has overlooked one key feature of such interpretations. In particular, model-theoretic interpretations *should* include, in the domain of interpretation of the variables, everything that is required for the very possibility of predication, logically complex thought (including abstract mathematical thought), and logical consequences of those thoughts. That's the point of (a) thinking that the predicate calculus is a fundamental system for expressing our thoughts and validating inferences, and (b) thinking that an interpretation of that system will give us an insight into what's required for the possibility of having those thoughts and making those inferences.

To approach our thesis, let's reconsider why λ-Conversion is a logical truth. Consider one of its consequences, which is logically complex not only because it involves the λ-expression but also because it involves the negation symbol:

$$[\lambda y\,\neg Gy]x \equiv \neg Gx$$

This holds for any property G: something exemplifies the negation of G iff it fails to exemplify G. There exists a logical exemplification pattern that underlies this fact, one that everything that fails to exemplify G has in common! After all, entities in the world do divide up into those that exemplify G and those that do not, and without the existence of the negation of G, we could not express that thought. How could two individuals a and b which both fail to be G not share the pattern of what is most-easily described as "exemplifying not-G"? If we treat this property of exemplifying not-G as reifying or representing this exemplification pattern, then it is required in any domain that contains the entities needed for truth of multiple predications of the form "x exemplifies not-G" ($[\lambda y\,\neg Gy]x$). And λ-Conversion also provides the *logical* justification as to why it is correct to infer one side of the biconditional from the other.

This same argument now applies to other instances of λ-Conversion, e.g., those involving other complex formulas such as conjunctions, disjunctions, conditionals, etc. The instances of λ-Conversion are true in any domain that contains the entities needed for the truth of predications involving complex predicates such as "x is G-and-H" ($[\lambda y\, Gy\,\&\,Hy]x$), etc., and provide the logical justification for inferences to and from such predicates, such as the inference from

---

logic from being considered logical. The existential import of λ-Conversion is analogous to that of the quantifier laws; just as the latter imply that the domain of individuals must be non-empty, the principle of λ-Conversion implies that each higher-order domain is not just non-empty but has at least two properties, one that everything exemplifies and one that is empty. Both the quantifier laws and λ-Conversion make minor existence demands.

"*x* is *G*-and-*H*" to "*x* is *G*" (i.e., $[\lambda y\ Gy\,\&\,Hy]x \vdash Gx$). The point also applies to relations and relational λ-expressions. Complex reasoning about the converse of relation R ($[\lambda xy\ Ryx]$)[32] and relations like unrequited love ($[\lambda xy\ Lxy\,\&\,\neg Lyx]$) assumes that the domain contains such relations. And, in general, the comprehension principle for relations, which we noted is derivable from λ-Conversion (footnote 31), is logically true precisely because it postulates the entities that are required for such complex relational reasoning.

This leads us to a somewhat different philosophical conception of logical truth and logical consequence. Let $\mathcal{L}$ be any language that is an extension of the language of object theory. Then we say: a formula φ in $\mathcal{L}$ is *logically true* if and only if φ is true in every model of $\mathcal{L}$ that includes all the entities required for the possibility of thinking thoughts expressible in $\mathcal{L}$. For any formula ψ of $\mathcal{L}$, the phrase "possibility of thinking" that ψ refers to the activity of having the particular thought that ψ, i.e., entertaining the particular propositional content that ψ. So, by saying "required for the possibility of thinking" that ψ, we also mean required for the existence of the propositional content that ψ.[33] Moreover, logic is committed to the existence of whatever entities are required for the possibility of drawing inferences when reasoning theoretically.[34] Consequently, we also say that a formula φ in $\mathcal{L}$ is a *logical consequence* of a set $\Gamma \subseteq \mathcal{L}$ if and only if φ is true in every model of $\mathcal{L}$ that (i) includes all the entities required for the possibility of thinking thoughts expressible in $\mathcal{L}$ and (ii) makes every member of Γ true.

The Tarskian and Fregean conceptions of logical truth dovetail in our conception: from Tarski we take the idea that logical truth is truth in all models of some given language $\mathcal{L}$ (Tarski 1936 [1983]); from Frege we take the conception of logic as providing constitutive norms of thought and reasoning as such (MacFarlane 2002), including constitutive norms for logically complex thought and reasoning.[35] By combining the two, we end up with a definition of the logical truth of a formula φ in $\mathcal{L}$ as φ's truth in all models that include everything that is required for the possibility of having the (logically complex) thoughts expressible in $\mathcal{L}$.[36] Under this conception, not only is λ-Conversion a logical

---

[32]For example, arguing that a non-symmetric relation is distinct from its converse.

[33]If ψ contains an empty term and doesn't denote a proposition, it still has truth conditions. In that case, the phrase "possibility of thinking" that ψ should be taken to mean: required for the possibility of entertaining the truth conditions of ψ, i.e., required for the existence of ψ's truth conditions.

[34]Though the natural ontological interpretation of object theory's quantifiers is that they commit one to the existence of abstract objects, others (e.g., Azzouni 2004, Priest 2005 [2016]) have suggested that one can interpret quantifiers without any ontological commitment (e.g., read $\exists$ as "some", rather than "there exists" or "there is". For the purposes of this paper, what matters is the reduction of mathematics to object theory, and we leave it an open question as to how one should interpret the quantifer $\exists$ that is used to express the theory. We'll give it a natural interpretation, but others are possible.

[35]For other views on Frege's conception of logic, see Goldfarb 2001, Linnebo 2003, and Blanchette 2012.

[36]This is consistent with the idea of logical truths as those that are constitutive of thought in general, and which are thus constitutively *a priori* in the sense discussed by Friedman 1994 [1999]. (A sentence φ is constitutively *a priori* for a theory T just in case it is presupposed by T.) Indeed, one might perhaps think of the argument to be given in the present section as

truth, but as noted in the example discussed above, the inference from $\neg Gx$ to $[\lambda y \, \neg Gy]x$ is one of logical consequence.

Let's then see how this conception can be used to understand why the lowest level instances of AXIOM 3 (i.e., the instances asserting the existence of abstract individuals) are logically true. The early Greek mathematicians adopted a method that has persisted until this day: they attempted to characterize objects whose *only* properties are the properties given by the defining principles.

Consider a simple example, such as discussions of The Equilateral Triangle in some language $\mathcal{L}_E$ of Euclidean geometry. Take any model of $\mathcal{L}_E$ that includes all the entities required for the possibility of thinking thoughts expressible in $\mathcal{L}_E$. A mathematician might have thoughts expressible in $\mathcal{L}_E$ such as:

> The Equilateral Triangle has sides of equal length.

In thinking about this object in the abstract, the mathematician might *logically infer* that The Equilateral Triangle is not scalene.[37] Here we have a logical conclusion in the form of a simple predication about The Equilateral Triangle, and the domain must have an object that exemplifies or encodes being an equilateral triangle for the thought (i.e., the propositional content) to exist and for the inference to be valid.

Thus, the mathematician has defined, objectified, and drawn inferences about a pattern of properties of individuals. The assertion that this pattern exists as an individual is true in the given model of $\mathcal{L}_E$, since models include everything required for having complex thoughts and making inferences expressible in $\mathcal{L}_E$ and thus include the relevant pattern of properties of individuals. It is important to emphasize here that the existence of this pattern doesn't commit us to saying that there is an object that exemplifies the properties defining the pattern. In fact, we have two options that avoid this commitment but which offer an object of thought: either treat the pattern as a property of properties in 3rd-order logic, or treat it as an *abstract* individual that encodes the properties in question. But the assertion of simple predications in $\mathcal{L}_E$ like the one displayed above suggests that the mathematician has conceived of The Equilateral Triangle as an abstract individual. AXIOM 3 is therefore a logical truth because it is true in every model that includes the entities required for having thoughts expressible in $\mathcal{L}_E$.

More generally, we may reason about any other combinations of properties in $\mathcal{L}$ that might be of interest where these combinations could be considered as individuals. AXIOM 3 is a logical truth given that these individuals must be present in every model of $\mathcal{L}$. There is an analogy with λ-Conversion; if one is willing to accept λ-Conversion as logical, on the grounds that, for any language $\mathcal{L}$, λ-Conversion is true in every model that includes the relations needed to express exemplification predications in $\mathcal{L}$, then one should likewise be willing to accept AXIOM 3 as logical. In other words, if one recognizes that

---

a kind of transcendental argument.

[37] This example is representative of modern mathematicians as well. Consider Dedekind, who defined his simply infinite systems as consisting of objects whose *only* properties were those given by the axioms in his 1888 (§71).

second-order comprehension is logical because it merely expresses the existence of entities presupposed for higher-order thinking and reasoning, then one should also recognize that comprehension over abstract individuals is logical because it (analogously) merely expresses the existence of entities presupposed for the possibility of such activities.

Similar conclusions now apply to higher-level λ-Conversion and higher-level abstracta. For take the example:

$$[\lambda R \, \neg \forall x Rxx]S \equiv \neg \forall x Sxx$$

This asserts: relation S exemplifies being a non-reflexive relation iff S fails to be reflexive. There is a pattern of which every relation that fails to be reflexive is a part! Clearly, relations in the world do divide up into those that are reflexive and those that are not, and without the existence of the negation of the property of reflexivity $[\lambda R \, \neg \forall x Rxx]$, how could two relations S and S' which both fail to be reflexive not share the pattern of what is most-easily described as "being a non-reflexive relation"? If we treat this property of being a non-reflexive relation as reifying or representing this pattern, then it is required in any model that contains the 2nd-level properties of relations needed for the truth of multiple predications of the form "S is non-reflexive" ($[\lambda R \neg \forall x Rxx]S$). And so on to other instances of higher-order λ-Conversion. Thus, higher-order λ-Conversion is logical because higher-order properties like *being non-reflexive* exist in every model that includes the entities needed to express such thoughts as "S is non-reflexive" and for reasoning to the conclusion that a non-reflexive relation is not an equivalence relation.

And this is likewise the case for higher-level comprehension over abstract entities. Thus, let's consider why a particular instance of the comprehension principle for abstract relations, as applied to mathematical relations, is a logical truth. Consider the less-than relation ($<_D$) as given by a language $\mathcal{L}_D$ and the theory of dense linear orderings without endpoints. This relation is given by the following theory T_D, in which $<$ is not indexed:

$$\forall x, y, z(x < y \, \& \, y < z \to x < z) \qquad \text{(Transitivity)}$$
$$\forall x(x \not< x) \qquad \text{(Irreflexivity)}$$
$$\forall x, y(x \neq y \to (x < y \, \lor \, y < x)) \qquad \text{(Connectedness)}$$
$$\forall x, y \exists z(x < z < y) \qquad \text{(Dense)}$$
$$\forall x \exists y \exists z(z < x < y) \qquad \text{(No Endpoints)}$$

The theorems derivable from these axioms constitute the theory T_D. What is more, we think of the $<_D$ relation itself as being constituted by that theory as well. The world itself doesn't contribute any facts about $<_D$ and there is no guarantee that a relation exists that exemplifies the properties of the $<_D$ relation — all there is to $<_D$ are the properties it has been assigned in this theory. In other words, the truths that ground all the facts about $<_D$ are facts of the form "In the theory T_D, $\mathcal{R}<_D$", where $\mathcal{R}$ ranges over properties of relations. The theorems in the scope of the operator "In the theory T_D, ..." ascribe to relation $<_D$ various properties of relations, such as the properties

of being transitive, irreflexive, connected, dense, and having no endpoints, and those that follow from them. There exists a pattern of predications, embedded within the theorems of T_D governing $<_D$, that we may articulate as a pattern of properties of relations, namely, the pattern: $T_D \vDash \mathcal{R}<_D$. $<_D$ just *is* that pattern of properties of relations, but instead of representing this pattern as a property of type $\langle\langle\langle i,i\rangle\rangle\rangle$ (property of properties of relations), encoding predication turns that pattern into an abstract relation of type $\langle i,i\rangle$ that encodes the properties of relations $\mathcal{R}$ that satisfy the pattern $T_D \vDash \mathcal{R}<_D$. (This is expressed in terms of our Reduction Axiom Schema, discussed in Section 5.1.)

Indeed, that relation must exist for us to have a mathematical thought, and draw inferences, about the relation $<_D$. Hence, the notions of logical truth and consequence defined above have the following application: a sentence φ of $\mathcal{L}_D$ containing the term $<_D$ is *logically true* if and only if φ is true in all interpretations that include all the (abstract) objects required for the possibility of having thoughts expressible in $\mathcal{L}_D$. $<_D$ is required for the possibility of having thoughts expressible in $\mathcal{L}_D$. Thus, the following claim, which expresses the existence of $<_D$, is logically true:

$$\exists R(A!R \,\&\, \forall \mathcal{R}(R\mathcal{R} \equiv T_D \vDash \mathcal{R}<_D))$$

This asserts: there is an abstract relation that encodes all and only the properties of relations exemplified by $<_D$ in T_D. And, generally, for any relation S of mathematical theory T, to have a thought about S, the following must be true:

$$\exists R(A!R \,\&\, \forall \mathcal{R}(R\mathcal{R} \equiv T \vDash \mathcal{R}S))$$

Notice the theory T_D is a simple case in which the only distinguished term of the mathematical theory is a relation term. More complex mathematical theories involve both distinguished relation terms and distinguished individual terms.

For example, Peano Arithmetic has as primitives: the property *being a number*, the relation *successor*, and the individual *Zero*. In this case, the existence of the abstract property *being a number*$_{\mathrm{PA}}$, of the abstract relation property *successor*$_{\mathrm{PA}}$, and the abstract individual *Zero*$_{\mathrm{PA}}$ are asserted by the relevant instances of comprehension AXIOM 3. Thus, when our analysis is applied to Peano Arithmetic:

- There are at least *three* kinds of exemplification patterns that exist in the sentences prefaced by the operator "In Peano Arithmetic, ...", namely, patterns of properties of the property of *being a number*, patterns of properties of the *successor* relation, and patterns of properties of the individual *Zero*. (There are, additionally, patterns of properties of both the relations and properties definable in PA, but we'll discuss those below.)

- These particular patterns induce and ground *three* corresponding kinds of encoding patterns that exist in the data of the form "In Peano Arithmetic, ...", namely, patterns of properties of the property of *being a number* in PA, patterns of properties of the *successor* relation in PA, and patterns of properties of the individual *Zero* in PA.

So, it follows that the instances of AXIOM 3 that assert the existence of the entities needed for the the analysis given by the Reduction Axiom Schema are all logical truths.[38]

In this section, we have argued for the logicality of axioms that assert the existence of two general kinds of logical entities:

- Those which exist as exemplification patterns among individuals, properties and relations and which, by comprehension, are logical objects within the domain of (higher-order) properties, i.e., those whose existence is asserted by AXIOM 2.

- Those that exist as predication patterns (either exemplification or encoding patterns) among properties and relations, that, by comprehension for abstract individuals, comprehension for abstract properties, and comprehension for abstract relations, are logical objects within the respective domains, i.e., those whose existence is asserted by AXIOM 3.

Both kinds of entities are *logical* in so far as they are patterns of predications. The entities asserted to exist by AXIOMs 2 and 3 are *abstracted* from *pure logical patterns* formulable solely in terms of predications generally in our language. Given that the conditions under which they are asserted to exist correlate with pure logical patterns that exist in our language, what else could they be but logical objects? So in what follows, we'll refer to the entities denoted by canonical descriptions as logical objects.

Note that if our axioms are logical, then any *theorem* we can derive is logical.[39] Now if we can show that the claims of mathematics prefixed by the theory operator, which are imported when we apply object theory, are analytic, then it will follow that the *unprefixed* encoding claims of mathematics derived from the prefixed claims and object theory (at the end of Section 5.1) become logical or analytic. So we now turn to a defense of the idea that when we extend object theory in the application to mathematics, the new axioms are analytic.

6.2 The Additional Axioms for Mathematics are Analytic

Our goal in this section is to show that the axioms added to object theory in Section 5, namely, those introduced by the Importation Principle and the Reduction Axiom, are analytic. These are principles that underlie our analysis of mathematics.

---

[38]This approach to logicism advances the ideas in Hodes 1984 (143) in several ways: his idea that the theory of natural numbers is an "encoding of a fragment of third-order logic" has been worked out in a systematic way, with the notion of *encoding* made rigorous. Moreover, we've applied the same technique to mathematical relations.

[39]This point assumes that rules of inference preserve analyticity and logicality. This is clear in the case of analyticity. But we think it holds even of our new notion of logicality. We claim that if axioms φ and $\varphi \to \psi$ are logical in virtue of being required for the possibility of abstract thought and reasoning, then ψ is logical for the same reason.

We take it to be uncontroversial to claim that axioms introduced by the Importation Principle are analytic: we can put aside the controversial question of whether "$\emptyset \in \{\emptyset\}$" is analytic, and yet still claim that "In ZF, $\emptyset_{\mathrm{ZF}} \in_{\mathrm{ZF}} \{\emptyset\}_{\mathrm{ZF}}$" is. The latter is true in virtue of the meaning of the terms 'ZF', '$\emptyset_{\mathrm{ZF}}$', '$\in_{\mathrm{ZF}}$', and '$\{\emptyset\}_{\mathrm{ZF}}$'. Since 'ZF' denotes a theory, and a theory encodes its theorems, 'In ZF, $\emptyset_{\mathrm{ZF}} \in_{\mathrm{ZF}} \{\emptyset\}_{\mathrm{ZF}}$', when represented as $\mathrm{ZF} \vDash \emptyset_{\mathrm{ZF}} \in_{\mathrm{ZF}} \{\emptyset\}_{\mathrm{ZF}}$, is true in virtue of the meaning of all the terms used in the representation.

It remains to argue that axioms introduced by the Reduction Axiom are analytic. To do this, we argue that the meanings of the terms flanking the identity sign in instances of the Reduction Axiom are identical, i.e., that the meaning m_τ of a mathematical term τ is identical to the meaning of the canonical description used to identify the denotation of τ. As we shall see, this conclusion almost immediately implies LC.

So, to argue that the instances of the Reduction Axiom are analytic, consider any mathematical concepts and the mathematical theories in which they occur. As examples, we again consider the concepts and the axioms of ZF set theory. Recall, first of all, that we formulated the following instances of the Reduction Axiom Schema:

$$\emptyset_{\mathrm{ZF}} = \iota x(A!x \,\&\, \forall F(xF \equiv\ \mathrm{ZF} \vDash F\emptyset_{\mathrm{ZF}}))$$
$$S_{\mathrm{ZF}} = \iota F(A!F \,\&\, \forall \mathcal{F}(F\mathcal{F} \equiv\ \mathrm{ZF} \vDash \mathcal{F}S_{\mathrm{ZF}}))$$
$$\in_{\mathrm{ZF}} = \iota R(A!R \,\&\, \forall \mathcal{R}(R\mathcal{R} \equiv\ \mathrm{ZF} \vDash \mathcal{R}\in_{\mathrm{ZF}}))$$

The right-hand side of these identity statements involve canonical definite descriptions (we call them canonical T-based descriptions below). These descriptions are formulated with the new indexed terms introduced when the mathematical theories are imported into object theory.

By referencing these descriptions, we may give the following argument for the claim that the instances of our Reduction Axiom Schema are analytic (and once we establish that, we give an argument for the thesis LC of Logicism about Concepts). Let τ be any unambiguous mathematical term used in a mathematical theory T, where τ is either an individual term or a predicate of T:

(P1) The meaning, m_τ, of a mathematical term τ is the inferential role of τ in the theory T.

(P2) The inferential role of τ in the theory T is the logical object denoted by the canonical T-based description for τ.[40]

(P3) The logical object denoted by the canonical T-based description is also the meaning of the canonical T-based description.

---

[40]It may come as a surprise to some that we have identified an inferential as a logical object. Traditional analyses of inferential roles do not generally assign a distinct, meaning-constituting role for each distinct non-logical symbol; for example, an inferential role analysis of number theory using the Peano *rules* (Warren 2020, 200) doesn't specify, in theoretical terms, *the meaning* of the constant symbol '0' or of the predicate symbol 'N'. But we are capturing the inferential role of the mathematical terms of T by objectifying their pattern within the body of theorems of T. We'll argue for this below.

So by transitivity of identity, the meaning m_τ of a mathematical term τ is identical to the meaning of the canonical T-based description for τ. And by the uncontroversial principle that if the meanings of τ and τ' are identical, then $\tau = \tau'$ is analytic, it follows that:

(A) The instances of the Reduction Axiom Schema are analytic.

Here, then, is our support for the premises of the above argument.

Concerning (P1). We think it is reasonable to suppose that a mathematical concept m_τ is constituted by the systematic use of the mathematical term τ. In turn, the systematic use of τ can be grounded in a system of axioms and inferences in which that term appears. Thus, m_τ can be identified with the inferential role of τ in T. For example, $m_\emptyset$, m_π, $m_\in$, $m_<$, etc., are identical to the inferential roles of $\emptyset$, π, $\in$, $<$, etc., in their respective theories.

We take (P1) to be a principle that is consistent with the following historical antecedents. It is one way to spell out Wittgenstein's meaning-as-use doctrine (1953), as well as the work of inferential role theorists such as Sellars (1980) and Brandom (1998). (P1) is also related to Schlick's and Carnap's view of theoretical terms in science, which in turn were influenced by Hilbert's view of geometry in which the meanings of mathematical terms are determined completely by the theories in which they figure.[41] (But P1 should be even easier to accept than the corresponding view of theoretical terms in science since, unlike the latter, mathematical terms aren't introduced with the idea of representing some empirical entity.) Finally, (P1) is consistent with Frege's Context Principle, except that the meaning for a mathematical term is not given by any single reference-fixing sentence but rather by a whole theory.

Note also that by identifying the meaning of a mathematical term with its inferential role, (P1) doesn't require us to invoke either the notion of an intension (in Carnap's [1947] sense) or the notion of a concept (in Church's [1951] sense). These notions are not needed in the semantics we give in the Appendix: meaning there is represented in terms of denotation. For any term τ, the meaning of that term is simply its denotation relative to (our interpretation and) an assignment to the variables, i.e., $\boldsymbol{d_f}(\tau)$. Moreover, we are not assuming a modal framework, and so the notion of intensionality, i.e., functions from worlds to (sets of) entities, doesn't apply. The rest of our argument will then be aimed at explicating the meaning, now identified as an inferential role, in terms of objects that will turn out to be logical.

Concerning (P2). We argue for P2 by considering examples such as the following: the inferential role of '$\emptyset$' in ZF is properly identified by the canonical

---

[41] See, for example, Friedman 1994 [1999, 26], where we find:

> In *General Theory of Knowledge*, his [Schlick's] starting point is Hilbert's *Foundations of Geometry* and the notion of axiomatic or *implicit* definition [...] According to the conception that Schlick derives from Hilbert, the primitive terms of geometry require no intuitive meaning or content. All we need to know about these primitives for the purposes of pure geometry are their mutual logical relationships set up explicitly in the axioms. Points, lines, and planes are any system of objects whatsoever that satisfy these axioms.

description, $\iota x(A!x \,\&\, \forall F(xF \equiv \mathrm{ZF} \vDash F\emptyset_{\mathrm{ZF}}))$. This holds because the abstract object denoted encodes all and only the properties of the null set derivable in ZF.[42] For instance, it is derivable in ZF both that $\emptyset \in \{\emptyset\}$ and that $[\lambda x\, x \in \{\emptyset\}]\emptyset$. The latter gets imported into object theory as $\mathrm{ZF} \vDash [\lambda x\, x \in \{\emptyset\}]_{\mathrm{ZF}}\emptyset_{\mathrm{ZF}}$. As such, the property $[\lambda x\, x \in \{\emptyset\}]_{\mathrm{ZF}}$ is one of the properties that satisfies the formula $\mathrm{ZF} \vDash F\emptyset_{\mathrm{ZF}}$ in the matrix of the description $\iota x(A!x \,\&\, \forall F(xF \equiv \mathrm{ZF} \vDash F\emptyset_{\mathrm{ZF}}))$. In object theory, the inferential role of the symbol $\emptyset$ in ZF is constituted by the object that encodes the totality of such properties. Its representation, $\emptyset_{\mathrm{ZF}}$, as identified by our canonical ZF-based description, captures the inferential role.

Concerning (P3). Our argument for this premise begins with the inspection of the semantics of our formalism, which reveals that the terms of our formalism are assigned only one semantic value. We claim that this semantic value serves both as the denotation and meaning of the terms of our formalism. Our semantics assigns meanings by assigning denotations. Indeed, we take it that *for our formalism*, the distinction between the denotation and meaning of its terms just collapses.[43] One doesn't have to build a formal language with terms having both intensions and extensions in order to model the intensions and extensions of the terms of natural language. One simply needs to have (a) terms in the formal language that can represent the extensions of the terms of natural language as well as (b) terms in the formal language that can represent the intensions of the terms of natural language. That is what our system does.[44]

Given P1 – P3, therefore, we've established (A), i.e., that the instances of our Reduction Axiom Schema are analytic. This completes our argument that the additional axioms added to object theory for the analysis of mathematics are analytic, and so we have established logicism in the form of LT. Since the theorems of mathematics are representable as theorems of extended object theory, and the axioms of extended object theory are all either logical truths or analytic truths, the theorems of mathematics are themselves either logical or analytic. This of course assumes that the rules of inference in classical logic

---

[42] Again, our analysis is one way of developing a *proof-theoretic semantics*, since we are generating term meanings by abstracting over the proof-theoretic roles of the relevant terms. See Prawitz 2006, Schroeder-Heister 2006, Francez & Dyckhoff 2010.

[43] As a consequence of this fact, it doesn't matter whether we say that a term "denotes" or "expresses" its semantic value.

[44] In the Appendix, we build the smallest *extensional* model of object theory. For purposes of showing consistency, this suffices. But one can build models in which the denotation of an n-place predicate is not just a set (or a truth-value, in the case of a 0-place predicate). Indeed, we take it that in the intended models, n-place predicates denote n-place relations, where the latter are then systematized by the principles for n-place relations offered by object theory (i.e., the principles laid down in Sections 4.1.2 and the definitions for the identity of relations given in Section 3). The result will still be an 'extensional' model but with primitive 'intensional' entities or urelements populating the relational domains. Moreover, we need not represent abstract individuals and abstract relations in these more fine-grained models as sets of properties intensionally-conceived. Instead, they too, may be regarded as intensionally-conceived urelements, indeed, hyperintensionally-conceived urelements, given that they correlate with sets of intensionally-conceived properties. So the denotations of our canonical T-based descriptions may themselves be conceived as highly intensional entities. We see no need to add a second semantic value (i.e., an intension) for the terms of our formal language.

preserve logicality and analyticity. We shall not argue for this claim.

Furthermore, premises P1 and P2 imply LC in the following form:

LC′ The meaning of a mathematical term τ, m_τ, is identical to a logical object.

LC follows from LC′ by taking the meaning of a term to be its denotation.

7 Objections and Observations

7.1 Objections

One objection that might be raised is whether we have offered an analysis that does 'too much', in that it would give us a means of reducing theoretical terms in natural science to logic! The objection argues that our very same procedure, as outlined above, would give us denotations for theoretical terms like 'electron', namely, the abstract property that encodes exactly the properties of properties attributed to this property by our best available physics. But, here, we argue, there is a disanalogy that prevents one from properly applying the above analysis to theoretical terms of natural science. The disanalogy is that in natural science, the theoretical properties like being an electron are natural properties, whereas the theoretical properties of mathematics are not. Thus, in the case of natural science, one might distinguish the natural property from our various concepts of that property, as these concepts change from scientific theory to scientific theory. The property of being an electron, for example, is something there in the world, though our theories of the electron reflect our evolving concept of this property. The concept, but not the property, is tied to the inferential role.

Given this distinction, we would argue that our analysis above could not be applied to analyze the property of being an electron (though it might be applied to the concept *electron* as this might be embodied by some scientific theory). Thus, P1 fails in the case of the natural properties of physics: "the meaning of theoretical term τ in a physical theory" is a natural property, not a physical concept. Hence P1 is false. Whereas the physical concept might well be identical to an inferential role, as (the corresponding version of) P1 would have it, the physical property is not an inferential role at all. By contrast, in the case of mathematical properties, there is no distinction to be drawn between our concepts of a mathematical property and the property itself. Either the mathematical properties and our concepts of them collapse, since the former are not given by anything over and above the concepts, or there are no mathematical properties beyond our mathematical concepts. In the former case, we use the above analysis to identify both the property and the concept, collapsing the two; whereas in the latter case, we use the above analysis solely for understanding our mathematical concepts (in which case there is nothing else to understand).

Another objection might run as follows. Our analysis assumes that mathematical objects are identified in terms of actual theories, i.e., theories that someone has actually developed or asserted. Doesn't this imply that the abstract realm of mathematical objects depends on the contingent actions of humans? To this, we may reply that by showing how all axiomatically developed

mathematics consists of logical/analytic truths, we have shown a striking fact that achieves the goals of logicism. But a deeper response is also available, since the objection suggests that the theorems of mathematics are contingent claims.

In fact, they are not. To see why, note that we've analyzed the theorems of mathematics as encoding truths about the individuals and relations of mathematics. Though we didn't develop the modal version of object theory here, one modal principle included in the theory is the claim that $\Diamond xF \rightarrow \Box xF$, i.e., that if possibly an abstract object encodes a property, then it does so necessarily.[45] So, though it may be a contingent fact as to what mathematicians have asserted by way of mathematical axioms, the theory-prefixed claims of the form "In theory T, p" are not contingent; they are analytic truths and given claims of that form, neither the theorems of T (as we've analyzed them) are contingent claims nor are the objects of the theory contingent objects.[46]

Thus, the realm of mathematical objects is not so closely tied to the contingent actions of human mathematicians. Though we haven't developed *modal* object theory in this paper, it would be trivial to add a modal operator. If we adopt the axioms for S5 modal logic, then these possibility claims about possible authors are in fact necessary, and thus the realm of mathematics becomes defined on our view in terms of objects with no air of contingency about them.

Finally, a Platonist might object that when we extend a theory like ZF to ZFC, the mathematician is not talking about a different realm of sets, while our approach implies that they are. But in fact this is not the case: our approach is consistent with assigning "the" right denotations to set-theoretic terms and predicates and that perhaps these denotations are only incompletely described by both ZF and ZFC. Note that the Platonist claim presupposes both that sets exist independently of our theories of them and that when we move from ZF to ZFC, the new theory is simply characterizing the objects of ZF further. So, for the sake of argument, suppose that the sets do exist independently of our theories of them and that there is consequently a complete body of all set-theoretic truths. Introduce a proper name, say '𝔊', for that body of truths and replace 'ZF' in our reduction axioms from above by '𝔊'. Then everything should go as before, and we should be able to reconstruct the mathematical terms of set theory as logical expressions. That is, we can plug 𝔊 into the machinery that we described above and the result, we claim, is a logicist reconstruction of *the* concept of set. Given this reconstruction, the theorems of ZF and ZFC will be true of the denotations that we have assigned to the terms and predicates of set theory 𝔊 since 𝔊 includes these theorems (assuming Choice is included in 𝔊). Of course, this is all completely hypothetical: we know that if *the* body of truths of set theory exists, it is not recursively axiomatizable, so we will never

---

[45]See the modal applications of object theory beginning with Zalta 1983, Ch. III.

[46]Our analysis extends to *any possible* mathematical theory (we include inconsistent theories of mathematics; see Zalta 2024, Section 3.2). Define a *possible mathematical theory* as any mathematical *situation* that possibly has an author, where we leave it to the mathematicians to say which propositions are mathematical. Then our analysis also applies to possible mathematical theories. Of course, we cannot import the theorems of those possible mathematical theories into object theory until a mathematician actually asserts a theory.

be "given" that body of truths in the form of a complete axiomatic system; nor does there seem to be any alternative manner in which we could be "given" that body of truths in a literal sense. Indeed, since Importation does not include the negative internalization principle (NI) as noted Section 5, our axiomatic theory does not include any axioms that would rule out that ZF, ZFC, and 𝔊 are pairwise identical. So none of this affects the principal logicist point that we want to make.

7.2   Comparison with Other Approaches

At the present time, we know of no other successful version of logicism, i.e., no successful attempt to establish LC and LT. While Frege's version of logicism failed due to the inconsistency of his Basic Law V, Whitehead and Russell's account of logicism was based on principles, such as the Axiom of Reducibility and the Axiom of Infinity, whose status as logical truths were unclear at best. Similarly, efforts by Hodes (1991) and Tennant (2004) both require an appeal to non-logical, or even non-analytic, axioms of infinity,[47] and it is not clear how the methodology in Tennant (1987, 2022) can be extended to the logicist analysis of an arbitrary mathematical theory T (i.e., it is not clear how, for an arbitrary theory T, to state introduction and elimination rules for the terms and predicates of T in a way that yields all and only the theorems of T).[48]

In recent years, neologicist theories have been developed that rely on abstraction principles.[49] These neologicist theories add new abstraction principles for each new kind of mathematical object introduced and each of these new 'double abstraction-identity principles' (like Hume's Principle, which introduces two abstractions, $\#F$ and $\#G$, in the same principle) *combines* both a comprehension (or existence) claim and an identity claim for the new kind of mathematical object. Clearly, any reduction of mathematics to logic will have to use some definitions or principles for identifying the mathematical objects as logical objects. On our view, however, only a single comprehension principle for objects is needed and, moreover, a single identity principle for objects is specifiable independently of comprehension.

Our approach differs from previous approaches in the following ways. (a) We appeal only to principles that are arguably logical or analytic and, in particular, we don't appeal to any non-logical axiom of infinity. Our unapplied and purely logical theory still has a finite model (described in the Appendix). When we apply the theory to mathematics, we sometimes import an infinite number of theorems into our own theory in the form of analytic truths. The infinity of mathematical entities that results from this extension is a presupposition of

---

[47]So our worry concerning such axioms of infinity is not due to them postulating the existence of some kind of infinite object—which would be fine, as far as we are concerned, as long as the object in question is a logical object—rather what we worry is about is whether one can argue that these axioms of infinity are logical or, at the very least, analytic.

[48]We also note the following difference between the present analysis and that in Tennant 1987, 2022: we offer a *theoretical identification* of the well-defined mathematical terms of T.

[49]See the work of Wright 1983, Hale 1987, Boolos 1986/87, Cook 2003, etc., and for an overview, see Linsky & Zalta 2006.

mathematical thought and thus counts as logical in our understanding of the term. (b) We don't have to continually re-prove our system is consistent since we use a uniform method for analyzing every kind of mathematical object. The model we have proposed in the Appendix—even though it is merely a minimal model for our background logical theory—grounds our conjecture that no special steps need be taken to guarantee consistency each time a new part of mathematics is analyzed in the manner outlined in Section 5. Though early forms of neologicism faced the 'bad company' and 'embarassment of riches' objections, recent forms have employed more general methods to guarantee consistency.[50] (c) Our approach is not subject to the Julius Caesar problem.[51] (d) Our analysis is prepared even for not-yet-formulated mathematical theories and new kinds of mathematical objects. Finally, (e) our approach gives an account of the denotations of both the individual terms as well as the predicates of mathematical theories.

Finally, our logicist proposal also differs from Linnebo's (2018) recent theory of thin objects. Linnebo develops another closely related follow-up project to Frege in which a comprehensive class of abstraction principles is employed to introduce abstract objects that are "thin" in the sense that their existence does not make any substantial demand on the world. There are three main differences to Linnebo's approach. (i) His abstraction principles are dynamic and predicative ("new" domains of objects are being introduced by iterated abstraction from "old" or given domains of objects), whereas the abstraction principles that are derivable from our theory are of the more traditional static and impredicative type.[52] (ii) We defend our basic principles to be logically or analytically true, while Linnebo does not and instead argues that his abstraction principles don't

---

[50]For the bad company objection, see Boolos 1990 (214), Field 1989 (158), and Dummett 1991 (188–189). For the embarassment of riches objection, see Heck 1992 and Weir 2003 (16). For recent work on the bad company objection, see the papers in Linnebo (ed.) 2009. See MacBride 2003 and Linnebo 2009 for a summary of the problem.

The neologicist could adopt some constraints, such as those proposed in Fine 2002, and suggest that any abstraction principle meeting those constraints would be 'safe' and not require a new consistency proof. But as far as we can tell, by placing those constraints on abstraction principles, one couldn't then derive all parts of mathematics from Fregean biconditionals. See Burgess 2003, 2005. Moreover, even if we adopt the constraints on abstract principles described in Fine 2002, it is not clear what final set of abstract principles emerge as the ones to be adopted. Which abstraction principles are to be asserted? A satisfying answer might be supplied by Cook (2012) who conjectures the class of acceptable abstraction principles to be the so-called class S-STB, which, other than not being affected by the bad company objection, has the highly attractive formal properties of being (i) field-conservative and (ii) maximally strictly logically symmetrically class conservative. (For the definitions of these notions and of the class S-STB itself, see Cook 2012; for some conceptual amendments, see Cook & Linnebo 2018.) We will have to leave a more detailed comparison with that neo-logicist proposal to future work.

[51]This is the problem that Frege himself raised for his own view: when abstraction principles like Hume's Principle ($\#F = \#G \equiv F \approx G$) are added to second order logic as the basis for identifying the numbers, identity is given only when two numbers are given in the form $\#F$ and $\#G$. The condition '$x = \#F$' is left undefined, and so the analysis yields no answer to questions like, "Is Julius Caesar identical to the number of Fs". In our system, '$x = \#F$' is always defined, since '$x = y$' is defined for every x, y.

[52]For the definition of '(im)predicative abstraction principle', see Linnebo 2018 (97).

make a substantial demand on the world. Indeed, Linnebo does not argue that his theory is a version of logicism. (iii) While Linnebo offers solutions for the standard problems for neo-logicism, we may not know how to show that some present or future mathematical theories are derivable from any of his abstraction principles. If this is right, our approach would seem preferable, as it is based on the actual presentations of mathematical theories in mathematical practice. We will have to leave a more substantial comparison with Linnebo's theory of thin objects to future work.

The principles of object theory are *general* in the sense accepted by both Kant and Frege (as described in MacFarlane 2002), namely, they are constitutive of, and provide a norm for, the possibility of having complex logical thoughts, including abstract mathematical thought. That is very different from the more standard conception of logic, since our conception allows that some existence claims can be logical truths. Indeed, we suggest that the argument (in the previous section) for the logicality of both of our comprehension principles (one for relations of every higher-order type and one for abstract objects of every type) justifies the early logicist view that logic may endorse existence claims, namely, those that assert the existence of the logical objects that Frege, Russell, and Whitehead used to reduce mathematics. The only existence claims logic is committed to are those required for the possibility of having complex logical thoughts.

So logic does have ontological commmitments, but it commits one to nothing more than what is required for the possibility of formulating and interpreting complex predications. In particular, unapplied object theory has ontological commitments (see the model in the Appendix), but the unapplied theory is not committed to anything more that what is required for the possibility of reifying structural relations among relations, i.e., what is required to make sense of the abstract relations that emerge from patterns of exemplification predication that are available in first- and second-order logic.

As we've mentioned, our understanding of logic and logicality has consequences for certain controversies concerning existence claims. Logicians have faced the following issue: what should one say about the fact that standard first-order logic entails existence claims (such as $\exists x\, x = x$). This has traditionally been seen as an uncomfortable conclusion since it was thought, on the one hand, that logic should be free of existential commitment, whereas on the other hand, that first-order logic should be the background system for the assertion of any non-logical existence principles. However, we can comfortably accept the fact that logical axioms imply existence claims because a non-empty classical domain should come pre-stocked with a minimum group of abstract objects.

Moreover, some philosophers have argued that $\exists x \exists y (x \neq y)$ can't be a logical truth, since it asserts the existence of more than a non-empty domain. Our arguments in Sections 6.1 and 6.2 show that this formula is a logical truth.[53] We argued that some formulas are in fact logically true in the sense that they

---

[53]Note that $\exists x \exists y (\neg A!x \,\&\, \neg A!y \,\&\, x \neq y)$ is not a necessary truth – this isn't required for logically complex thought. But $\exists x \exists y (A!x \,\&\, A!y \,\&\, x \neq y)$ is and, thus, so is $\exists x \exists y (x \neq y)$, since the latter is implied by the former.

are true in all models that make it possible to have logically complex thoughts. Moreover, once we import mathematics into our system in the form of analytic truths, then we can derive the existence of new logical objects.

7.3 Epistemology Redux

We claim, finally, that the epistemological benefits of logicism now accrue. By showing that mathematical statements are analytic, it follows that by knowing the meanings of (the terms in) these statements, we are equipped with all the tools we need to determine whether they are true. We can know mathematical theorems by deriving them solely from logically true statements and analytic statements. Thus, no special cognitive faculty for knowledge of mathematical truths is needed other than the faculty of understanding, which is a faculty we, like Benacerraf (1981), take to be explainable in naturalistic terms.

So we don't have to posit a causal information pathway, like the causal theory of reference, to explain how we come to understand the terms of mathematical statements. Our comprehension principles already constitute the paths by which we apprehend abstract objects: they just are the means by which we cognitively grasp the objects denoted by the terms of mathematical theory T.

Note that the claim that mathematical truths are logical or analytic truths does not entail that for each mathematical claim it would be easy to determine that it is a mathematical truth. Moreover, what we have *not* done in the present paper is to say anything about the *a priori* justification of the logic underlying object theory. To be clear, though, the epistemological situation is very different from that surrounding the foundational system of *Principia Mathematica*. Whitehead & Russell couldn't very well argue that the axioms of reducibility and infinity are *logical*, even if they had tried to use the grounds we provided above: it is hard to see how such axioms are required for the possibility of (abstract) thought. Our logical framework, by contrast, requires no such axioms and the axioms it does assert are required for such thought. But in the present case, even if one accepts that our axioms are logical, there is still the question of whether they are justified. We have not addressed that latter question. We could argue that a logic such as ours is justified because it is *presupposed* somehow, or because the logic, through a process of reflective equilibrium, offers a rational reconstruction of the data (i.e., logical consequences we accept pretheoretically) that is better than other logical systems. But we have to leave this argument for another paper.

comments on our argument. We're grateful to the anonymous referees, who offered a number of important suggestions for improvement.

References

Anderson, D., and E. Zalta, 2004, 'Frege, Boolos, and Logical Objects', *Journal of Philosophical Logic*, 33(1): 1–26.

Azzouni, J., 2004, *Deflating Existential Consequence: A Case for Nominalism*, New York: Oxford University Press.

Benacerraf, P., 1981, 'Frege: The Last Logicist', in *Midwest Studies in Philosophy: VI*, P. French *et al.* (eds.), Minneapolis: U. Minnesota Press; page reference is to the reprint in Demopoulos [1995].

Blanchette, P., 2012, *Frege's Conception of Logic*, New York: Oxford University Press.

Boolos, G., 1986/87, 'Saving Frege From Contradiction', *Proceedings of the Aristotelian Society*, 87(1): 137–151; reprinted in Boolos [1998], 171–182.

———, 1990, 'The Standard of Equality of Numbers,' in *Meaning and Method: Essays in Honor of Hilary Putnam*, G. Boolos (ed.), Cambridge: Cambridge University Press, pp. 261–277; page reference is to the reprint in Boolos [1998], pp. 202–19.

———, 1998, *Logic, Logic, and Logic*, Cambridge MA: Harvard University Press.

Brandom, R., 1998, *Making it Explicit*, Cambridge, MA: Harvard University Press.

Bueno, O., C. Menzel, and E. Zalta, 2014, 'Worlds and Propositions Set Free,' *Erkenntnis*, 79: 797–820.

Burgess, J., 2003, 'Review of Kit Fine, *The Limits of Abstraction*', *Notre Dame Journal of Formal Logic*, 44(4): 227–251.

———, 2005, *Fixing Frege*, Princeton: Princeton University Press.

Carnap, R., 1931, 'Die logizistische Grundlegung der Mathematik', *Erkenntnis*, 2: 91–105.

———, 1947, *Meaning and Necessity*, Chicago: University of Chicago Press.

Church, A., 1951, 'A Formulation of the Logic of Sense and Denotation', in *Structure, Method and Meaning: Essays in Honor of Henry M. Sheffer*, P. Henle *et al.* (eds.), New York: Liberal Arts Press, 3–24.

Cook, R., 2003, 'Iteration One More Time', *Notre Dame Journal of Formal Logic*, 44(2): 63–92.

———, 2012, 'Conservativeness, Stability, and Abstraction', *The British Journal for the Philosophy of Science*, 63(3): 673–696.

Cook, R., and Ø. Linnebo, 2018, 'Cardinality and Acceptable Abstraction', *Notre Dame Journal of Formal Logic*, 59(1): 61–74.

Dedekind, R., 1888, *Was sind und was sollen die Zahlen*, Braunschweig: Vieweg und Sohn; 2nd edition, 1893; 7th edition, 1939.

Dummett, M., 1991, *Frege: Philosophy of Mathematics*, Cambridge, MA: Harvard University Press.

Field, H., 1989, *Realism, Mathematics, and Modality*, Oxford: Blackwell.

Fine, K., 2002, *The Limits of Abstraction*, Oxford: Clarendon Press.

Francez, N., and R. Dyckhoff, 2010, 'Proof-theoretic Semantics for a Natural Language Fragment', *Linguistics and Philosophy*, 33: 447–477.

Friedman, M., 1994 [1999], 'Geometry, Convention, and the Relativized A Priori: Reichenbach, Schlick, and Carnap', in W. Salmon and G. Wolters (eds.), *Logic, Language, and the Structure of Scientific Theories*, Pittsburgh: University of Pittsburgh Press, pp. 21–34; reprinted in M. Friedman, *Reconsidering Logical Positivism*, Cambridge: Cambridge University Press, 1999, pp. 59–70. [Page reference is to the original.]

Frege, G., 1893/1903, *Grundgesetze der Arithmetik*, 2 volumes, Band I (1893), Band II (1903), Jena: Verlag Hermann Pohle.

Göedel, K., 1964, 'What is Cantor's Continuum Problem,' reprinted in S. Feferman et al., *Kurt Göedel: Collected Works* (Volume II: Publications 1938–1974), Oxford: Oxford University Press, 1990, 254–270.

Goldfarb, W., 2001, 'Frege's Conception of Logic,' in J. Floyd and S. Shieh (eds.), *Future Pasts: The Analytic Tradition in Twentieth Century Philosophy*, Oxford: Oxford University Press, 25–41.

Hale, B., 1987, *Abstract Objects*, Oxford: Blackwell.

Heck, R., 1992, 'On the Consistency of Second-order Contextual Definitions,' *Noûs*, 26: 491–94.

Hintikka, J. 1959, 'Towards a Theory of Definite Descriptions,' *Analysis*, 19(4): 79–85.

Hodes, H., 1984, 'Logicism and the Ontological Commitments of Arithmetic,' *The Journal of Philosophy*, 81(3): 123–149.

———, 1991, 'Where Do Sets Come From?', *Journal of Symbolic Logic*, 56(1): 151–175.

Klev, A.M., 2017, 'Dedekind's Logicism', *Philosophia Mathematica*, 25(3): 341–36.

Leitgeb, H., and Ladyman, J., 2008, 'Criteria of Identity and Structuralist Ontology', *Philosophia Mathematica*, 16(3): 388–396.

Linnebo, Ø., 2003, 'Frege's conception of logic: From Kant to Grundgesetze,' *Manuscrito - Rev. Int. Fil. Campinas*, 26(2): 235–252.

———, 2009, 'Introduction', in Linnebo (ed.) 2009, pp. 321–329.

———, 2018, *Thin Objects: An Abstractionist Account*, Oxford: Oxford University Press.

Linnebo, Ø. (ed.), 2009, *The Bad Company Problem*, special issue, *Synthese*, 70/3 (October).

Linsky, B., and Zalta, E., 2006, 'What is Neologicism?', *Bulletin of Symbolic Logic*, 12/1: 60–99.

MacBride, F., 2003, 'Speaking with Shadows: A Study of Neo-Logicism,' *British Journal for the Philosophy of Science*, 54: 103–163.

MacFarlane, J., 2002, 'Frege, Kant, and the Logic in Logicism', *The Philosophical Review*, 111(1): 25–65.

Nodelman, U. and Zalta, E., 2014, 'Foundations for Mathematical Structuralism', *Mind*, 123/489: 39–78.

Pelletier, F.J., and Zalta, E., 2000, 'How to Say Goodbye to the Third Man', *Noûs*, 34(2): 165–202.

Prawitz, D., 2006, 'Meaning Approached via Proofs', *Synthese*, 148: 507–524.

Priest, G., 2005 [2016], *Towards Non-Being: The Logic and Metaphysics of Intentionality*, Oxford: Clarendon; revised 2nd edition, 2016.

Rayo, A., 2005, 'Logicism Reconsidered', in Stewart Shapiro (ed.), *The Oxford Handbook of Philosophy of Mathematics and Logic*, New York: Oxford University Press, 203–235; doi:10.1093/0195148770.003.0007

Roeper, P., 2015, 'A Vindication of Logicism', *Philosophia Mathematica*, 24(3): 360–378.

Schroeder-Heister, P., 2006, 'Validity Concepts in Proof-Theoretic Semantics', *Synthese*, 148: 525–571.

Sellars, W., 1980, 'Inference and Meaning', in *Pure Pragmatics and Possible Worlds*, J. Sicha (ed.), Atascadero: Ridgeview, pp. 261–313.

Tarski, A., 1936 [1983], 'Über den Begriff der logischen Folgerung', in *Actes du Congrès International de Philosophie Scientifique*, fasc. 7 (Actualitès Scientifiques et Industrielles, vol. 394), Paris: Hermann et Cie, pp. 1–11; translated as 'On The Concept of Logical Consequence', in *Logic, Semantics, Metamathematics*, second edition, J. H. Woodger (trans.), John Corcoran (ed.), Indianapolis: Hackett, pp. 409–420.

Tennant, N., 1987, *Anti-Realism and Logic: Truth as Eternal*, Oxford: Clarendon.

———, 2004, 'A General Theory of Abstraction Operations', *The Philosophical Quarterly*, 54(214): 105–133.

———-, 2022, *The Logic of Number*, Oxford: Oxford University Press.

Warren, J., 2020, *Shadows of Syntax*, New York: Oxford University Press.

Weir, A., 2003, 'Neo-Fregeanism: An Embarrassment of Riches', *Notre Dame Journal of Formal Logic*, 44(1): 13–48.

Whitehead, A.N., and B. Russell, 1910–1913, *Principia Mathematica*, 3 volumes, Volume 1 (1910), Volume 2 (1912), Volume 3 (1913), Cambridge: Cambridge University Press.

Wittgenstein, L., 1953, *Philosophical Investigations*, G.E.M. Anscombe and R. Rhees (eds.), G.E.M. Anscombe (trans.), Oxford: Blackwell.

Wright, C., 1983, *Frege's Conception of Numbers as Objects*, Scots Philosophical Monographs, vol. 2, Aberdeen: Aberdeen University Press.

Zalta, E., 1983, *Abstract Objects: An Introduction to Axiomatic Metaphysics*, Dodrecht: D. Reidel.

———, 1988, *Intensional Logic and the Metaphysics of Intentionality*, Cambridge, MA: Bradford/MIT Press.

———, 1999, 'Natural Numbers and Natural Cardinals as Abstract Objects: A Partial Reconstruction of Frege's *Grundgesetze* in Object Theory', *Journal of Philosophical Logic*, 28(6): 619–660.

———, 2000, 'Neo-Logicism? An Ontological Reduction of Mathematics to Metaphysics', *Erkenntnis*, 53(1–2): 219–265.

———, 2006, 'Essence and Modality', *Mind*, 115/459: 659–693.

———, 2024, 'Mathematical Pluralism', *Noûs*, 58(2): 306–332. doi:10.1111/nous.12451

Appendix: A Minimal Model of the Logical Framework

The logic in the foregoing is consistent, as can be demonstrated by the construction of the smallest model. This model happens to be an extensional one: ordinary properties and ordinary relations are not distinguished from their exemplification extensions, and ordinary propositions are not distinguished from the two truth values The True and The False. Of course, this extensional model is not the intended one.[54] We emphasize, however, that our model doesn't require "full higher-order semantics"; we don't require, of any higher-order domain, that it be the full power set of the lower order domain.

A Bounded Language

Our analysis of mathematics does not require the full unbounded language defined in the paper in Section 3. So, as we develop a model of typed object theory, we restrict our attention to the fragment we need. We shall therefore define the bounded language $\mathcal{L}_{n,m}$, where n and m are bounds that set, respectively, the *width* and *height* of the types for the terms of the language. We begin by defining the functions h and w for the height and width, respectively, of a given type.

The *width* of type t, written $w(t)$, is defined as:

- $w(i) = 1$

- $w(\langle \, \rangle) = 1$

- $w(\langle t_1, \ldots, t_k \rangle) = \sum_1^k w(t_k)$

The *height* of type t, written $h(t)$, is defined as:

- $h(i) = 0$

- $h(\langle \, \rangle) = 1$

- $h(\langle t_1, \ldots, t_k \rangle) = 1 + \max\{h(t_1), \ldots, h(t_k)\}$

Then we define $\mathcal{L}_{n,m}$ as the language that includes any well-formed expression of $\mathcal{L}$ that can be formulated only with terms τ of type t such that $w(t) \leq n$ and $h(t) \leq m$.

Before we define a model for the bounded language $\mathcal{L}_{n,m}$, a few observations are in order. Intuitively, we want to choose bounds that will yield the smallest language and model needed for our analysis. The following two considerations play a role in setting the bounds on $\mathcal{L}_{n,m}$:

---

[54] An intended model would distinguish properties and relations from their exemplification extensions, and would distinguish propositions from their truth values. Thus, an intended model would be intensional, and if modality were added, would be hyperintensional.

- In Section 5.1, we analyze the property of being a ZF-set as an abstract property having type $\langle i \rangle$. Abstract properties of this type encode properties of type $\langle\langle i \rangle\rangle$. And properties of this latter type can be both ordinary and abstract, though the mathematical properties will encode only abstract properties of type $\langle\langle i \rangle\rangle$. But these abstract properties must, in turn, encode properties of type $\langle\langle\langle i \rangle\rangle\rangle$. Note this requires the bound on the height of the types in the language to be at least 3 and the bound on the width to be at least 1.

- In Section 5.1, we analyze the membership relation of ZF as an abstract relation having type $\langle i, i \rangle$. Abstract relations of this type encode properties of type $\langle\langle i, i \rangle\rangle$. And properties of this latter type can be both ordinary and abstract, though the membership relation will encode only abstract properties of type $\langle\langle i, i \rangle\rangle$. But these abstract properties must, in turn, encode properties of type $\langle\langle\langle i, i \rangle\rangle\rangle$. Note this requires the bound on the height of the types in the language to be at least 3 and the bound on the width to be at least 2.

Given these facts, it should be clear that the minimal fragment our analysis of mathematics requires is the bounded language $\mathcal{L}_{2,3}$. In line with what we said above, $\mathcal{L}_{2,3}$ includes any well-formed expression of $\mathcal{L}$ that can be formulated only with terms τ of type t such that $w(t) \leq 2$ and $h(t) \leq 3$. So in specifying a general model for $\mathcal{L}_{n,m}$, we shall occasionally focus on the model for the language $\mathcal{L}_{2,3}$.

The Smallest, Extensional Model for $\mathcal{L}_{n,m}$

We construct our model in two basic stages: first we construct the *structural* domains of the model, and second, we specify the *domains of quantification* and a proxy function (that assigns to each element in a domain of quantification to an element of a structural domain). The construction of the *structural* domain occurs in two stages: (1) the kernel of each type and (2) the abstract objects of each type.

Structural Domains: Kernel

(1) We define the *kernel* $\mathbf{K}_t$ of objects of type t, by induction, as follows:

- Where $t = i$, the kernel $\mathbf{K}_i$ of individuals is the union of two subdomains: the ordinary individuals $\mathbf{O}_i$ and the special individuals $\mathbf{S}_i$. For the purposes of building a specific minimal model, we stipulate that $\mathbf{O}_i$ is empty and $\mathbf{S}_i$ contains a single special individual $\boldsymbol{s}_i$, which we henceforth label as $\boldsymbol{b}$.

- Where $t = \langle \, \rangle$, the kernel $\mathbf{K}_{\langle \, \rangle}$ of propositions is the union of two subdomains: the ordinary propositions, $\mathbf{O}_{\langle \, \rangle}$ and the special propositions $\mathbf{S}_{\langle \, \rangle}$. For the purposes of building a specific minimal model, we stipulate that
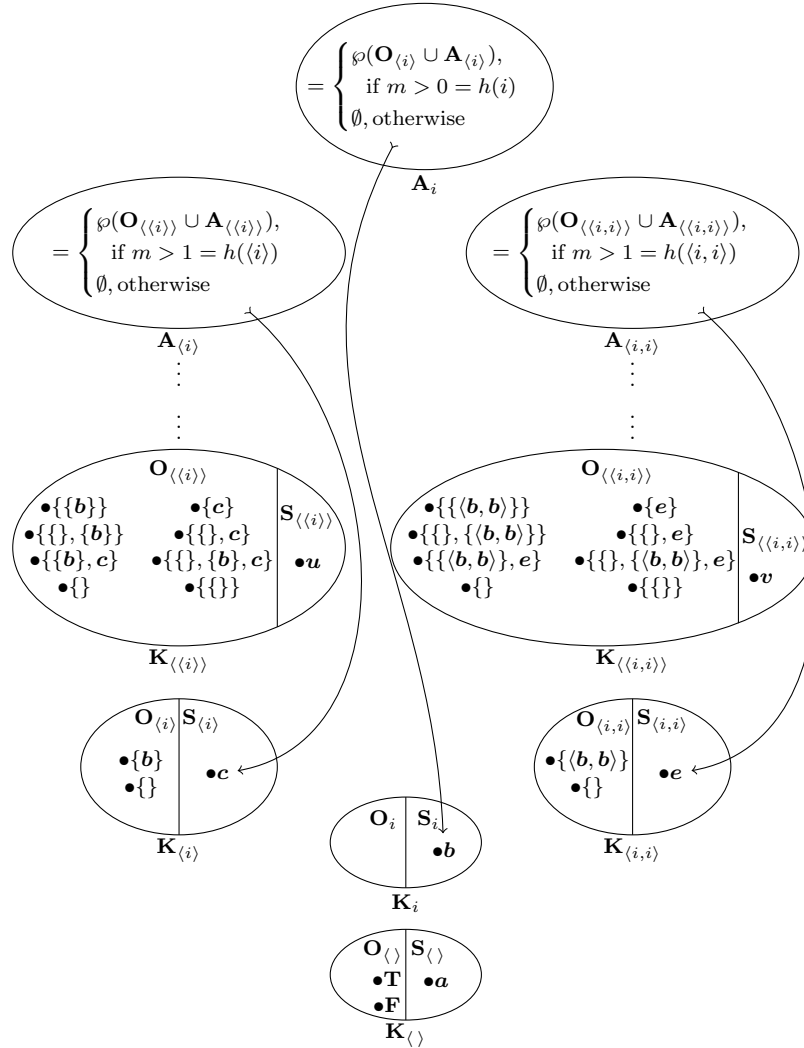
Figure 1: A fragment of the minimal model with unrestricted typed comprehension for abstracta. The domains of ordinary objects, from the bottom up, are: the *kernel of propositions* $\mathbf{K}_{\langle\rangle}$ $(= \mathbf{O}_{\langle\rangle} \cup \mathbf{S}_{\langle\rangle})$; the *kernel of individuals* $\mathbf{K}_i$ $(= \mathbf{O}_i \cup \mathbf{S}_i)$; the *kernel of properties of individuals* $\mathbf{K}_{\langle i\rangle}$ $(= \mathbf{O}_{\langle i\rangle} \cup \mathbf{S}_{\langle i\rangle})$; the *kernel of binary relations among individuals* $\mathbf{K}_{\langle i,i\rangle}$ $(= \mathbf{O}_{\langle i,i\rangle} \cup \mathbf{S}_{\langle i,i\rangle})$; the *kernel of properties of properties of individuals* $\mathbf{K}_{\langle\langle i\rangle\rangle}$; the *kernel of properties of relations among individuals* $\mathbf{K}_{\langle\langle i,i\rangle\rangle}$; and so on. The domains of abstract objects, from the top down, are: the *abstract individuals* $\mathbf{A}_i$ $(=$ the power set of $\mathbf{O}_{\langle i\rangle} \cup \mathbf{A}_{\langle i\rangle})$; the *abstract properties of individuals* $\mathbf{A}_{\langle i\rangle}$ $(=$ the power set of $\mathbf{O}_{\langle\langle i\rangle\rangle} \cup \mathbf{A}_{\langle\langle i\rangle\rangle})$; and the *abstract relations among individuals* $\mathbf{A}_{\langle i,i\rangle}$ $(=$ the power set of $\mathbf{O}_{\langle\langle i,i\rangle\rangle} \cup \mathbf{A}_{\langle\langle i,i\rangle\rangle})$; and so on.

$\mathbf{O}_{\langle\,\rangle}$ contains two propositions, labeled $\mathbf{T}$ and $\mathbf{F}$, and $\mathbf{S}_{\langle\,\rangle}$ contains a single special proposition $\boldsymbol{s}_{\langle\,\rangle}$, which we henceforth label as $\boldsymbol{a}$.

- Where $t = \langle t_1, \ldots, t_n \rangle$ $(n \geq 1)$, for any types $t_1, \ldots, t_n$, the kernel $\mathbf{K}_{\langle t_1, \ldots, t_n \rangle}$ of relations among objects having types $t_1, \ldots, t_n$, respectively, is the union of two subdomains, $\mathbf{O}_{\langle t_1, \ldots, t_n \rangle}$ and $\mathbf{S}_{\langle t_1, \ldots, t_n \rangle}$, where $\mathbf{O}_{\langle t_1, \ldots, t_n \rangle} = \wp(\mathbf{K}_{t_1} \times \ldots \times \mathbf{K}_{t_n})$ and $\mathbf{S}_{\langle t_1, \ldots, t_n \rangle}$ contains at least one special object $\boldsymbol{s}_{\langle t_1, \ldots, t_n \rangle}$. For the purposes of building a specific minimal model, we label:

$$\mathbf{s}_{\langle i \rangle} \text{ as } \boldsymbol{c}$$
$$\boldsymbol{s}_{\langle i,i \rangle} \text{ as } \boldsymbol{e}$$
$$\boldsymbol{s}_{\langle\langle i \rangle\rangle} \text{ as } \boldsymbol{u}$$
$$\boldsymbol{s}_{\langle\langle i,i \rangle\rangle} \text{ as } \boldsymbol{v}.$$

Given these stipulations, we have the following consequences:

- $\mathbf{K}_{\langle i \rangle}$ (i.e., the kernel of objects of type $\langle i \rangle$) $= \wp(\mathbf{K}_i) \cup \{\boldsymbol{c}\}$, and so $\mathbf{K}_{\langle i \rangle} = \{\{\boldsymbol{b}\}, \{\,\}, \boldsymbol{c}\}$. (This is pictured in the graphic.)

- $\mathbf{K}_{\langle i,i \rangle}$ (i.e., the kernel of objects of type $\langle i, i \rangle$) $= \wp(\mathbf{K}_i \times \mathbf{K}_i) \cup \{\boldsymbol{e}\}$, and so $\mathbf{K}_{\langle i,i \rangle} = \{\{\langle \boldsymbol{b}, \boldsymbol{b}\rangle\}, \{\,\}, \boldsymbol{e}\}$. (This is pictured in the graphic.)

- $\mathbf{K}_{\langle i,\langle\,\rangle\rangle}$ (i.e., the kernel of relations between individuals and propositions) $= \wp(\mathbf{K}_i \times \mathbf{K}_{\langle\,\rangle}) \cup \{\boldsymbol{s}_{\langle i,\langle\,\rangle\rangle}\}$, and so $\mathbf{K}_{\langle i,\langle\,\rangle\rangle} = \{\{\langle \boldsymbol{b}, \mathbf{T}\rangle\}, \{\langle \boldsymbol{b}, \mathbf{F}\rangle\}, \{\langle \boldsymbol{b}, \boldsymbol{a}\rangle\}, \{\langle \boldsymbol{b}, \mathbf{T}\rangle, \langle \boldsymbol{b}, \mathbf{F}\rangle\}, \{\langle \boldsymbol{b}, \mathbf{T}\rangle, \langle \boldsymbol{b}, \boldsymbol{a}\rangle\}, \{\langle \boldsymbol{b}, \mathbf{F}\rangle, \langle \boldsymbol{b}, \boldsymbol{a}\rangle\}, \{\langle \boldsymbol{b}, \mathbf{T}\rangle, \langle \boldsymbol{b}, \mathbf{F}\rangle, \langle \boldsymbol{b}, \boldsymbol{a}\rangle\}, \{\,\}, \boldsymbol{s}_{\langle i,\langle\,\rangle\rangle}\}$. (This is not pictured in the graphic.)

Etc.

Structural Domains: Abstract Objects

(2) Given m as the maximum height, we recursively define the domain $\mathbf{A}_t$ as follows:

$$\mathbf{A}_t = \begin{cases} \wp(\mathbf{O}_{\langle t \rangle} \cup \mathbf{A}_{\langle t \rangle}) & \text{if } h(t) < m \\ \emptyset & \text{otherwise.} \end{cases}$$

The identity conditions for elements of $\mathbf{A}_t$ depend on the higher-type elements of $\mathbf{O}_{\langle t \rangle}$. Given these stipulations, we have the following consequences for $\mathbf{A}_{\langle i \rangle}$ and $\mathbf{A}_i$ when $m = 1$:

$$\mathbf{A}_{\langle i \rangle} = \emptyset$$

$$\begin{aligned} \mathbf{A}_i &= \wp(\mathbf{O}_{\langle i \rangle} \cup \mathbf{A}_{\langle i \rangle}) \\ &= \wp(\mathbf{O}_{\langle i \rangle} \cup \emptyset)) \\ &= \wp(\{\{b\}, \{\}\}) \\ &= \{\{\}, \{\{b\}, \{\}\}, \{\{b\}\}, \{\{\}\}\} \end{aligned}$$

When $m = 2$:

$$\mathbf{A}_{\langle\langle i\rangle\rangle} \,=\, \emptyset$$

$$
\begin{aligned}
\mathbf{A}_{\langle i\rangle} \,&=\, \wp(\mathbf{O}_{\langle\langle i\rangle\rangle} \cup \mathbf{A}_{\langle\langle i\rangle\rangle}) \\
&=\, \wp(\wp(\mathbf{K}_{\langle i\rangle}) \cup \emptyset) \\
&=\, \wp(\{\ldots \text{all 8 subsets of } \mathbf{K}_{\langle i\rangle} \ldots\}) \\
&=\, \{\ldots \text{all 256 subsets of } \wp(\mathbf{K}_{\langle i\rangle}) \ldots\}
\end{aligned}
$$

$$
\begin{aligned}
\mathbf{A}_i \,&=\, \wp(\mathbf{O}_{\langle i\rangle} \cup \mathbf{A}_{\langle i\rangle}) \\
&=\, \wp(\mathbf{O}_{\langle i\rangle} \cup \{\ldots \text{all 256 elements of } \mathbf{A}_{\langle i\rangle} \ldots\}) \\
&=\, \wp(\{\,\{b\},\ \{\}\,\} \cup \{\ldots \text{all 256 elements of } \mathbf{A}_{\langle i\rangle} \ldots\}) \\
&=\, \wp(\{\{b\},\ldots \text{all 256 elements of } \mathbf{A}_{\langle i\rangle} \ldots \}) \\
&=\, \wp(\{\ldots \text{all 257 elements of } \mathbf{O}_{\langle i\rangle} \cup \mathbf{A}_{\langle i\rangle} \ldots\}) \\
&=\, \{\ldots \text{all } 2^{257} \text{ subsets of } \mathbf{O}_{\langle i\rangle} \cup \mathbf{A}_{\langle i\rangle} \ldots\}
\end{aligned}
$$

But as noted earlier, for the purposes of building a specific model for our analysis of mathematics, we will set the bound of m to 3.

Domains of Quantification and the Proxy Function

The domains over which the variables of our language range are now defined simply as follows. Where $\mathbf{D}_t$ is the domain of quantification for type t:

$$\mathbf{D}_t = \mathbf{A}_t \cup \mathbf{O}_t$$

We next define an extended proxy function $\|\cdot\|$ in two steps. In the first step, we define a *proxy function* $|\cdot|$ so that it maps abstract individuals, abstract properties, and abstract relations to special individuals, special properties, and special relations, respectively. In the second step, we extend this function to the *extended proxy function* $\|\cdot\|$ which is defined on all the domains of quantification: it preserves what $|\cdot|$ assigns to the abstract entities but also assigns each ordinary element in each domain of quantification to itself as proxy.

The function $|\cdot|$ is defined generally as, for each type t:

$$|\cdot| : \mathbf{A}_t \to \mathbf{S}_t$$

In the minimal model, there is only one special object in each domain $\mathbf{S}_t$, and so all the abstract objects in $\mathbf{A}_t$ get mapped to the same proxy.[55] For example, in the minimal model:

- where $\mathbf{a}_i$ is an abstract individual in $\mathbf{A}_i$ (i.e., where $\mathbf{a}_i$ is a set of 1st level properties), then $|\mathbf{a}_i| = \boldsymbol{b}$,

- where $\mathbf{a}_{\langle i\rangle}$ is a 1st level abstract property in $\mathbf{A}_{\langle i\rangle}$ (i.e., where $\mathbf{a}_{\langle i\rangle}$ is a set of 2nd level properties of properties), then let $|\mathbf{a}_{\langle i\rangle}| = \boldsymbol{c}$, and

---

[55] This is not just an artifact of the extensional model construction. Since abstract objects are individuated by bundles of properties, there must be more abstract objects than properties. That means that there must be distinct abstract objects that are indiscernible—i.e., they have the same pattern of exemplifications. Since the proxy determines the exemplification pattern, such indiscernible but distinct abstract objects must share the same proxy.

- where $\mathbf{a}_{\langle i,i \rangle}$ is a 1st level abstract relation in $\mathbf{A}_{\langle i,i \rangle}$ (i.e., where $\mathbf{a}_{\langle i,i \rangle}$ is a set of 2nd level properties of relations), then let $|\mathbf{a}_{\langle i,i \rangle}| = \boldsymbol{e}$.

We then extend $|\cdot|$ to the *extended proxy function* $\|\cdot\|$ as follows. Where t is any type and $\mathbf{D}_t$ is the domain of type t (as defined below), and $\boldsymbol{o}$ is a variable ranging over the entities of domain $\mathbf{D}_t$, then for all $\boldsymbol{o} \in \mathbf{D}_t$:

$$\|\boldsymbol{o}\| = \begin{cases} |\boldsymbol{o}| & \text{when } |\boldsymbol{o}| \text{ is defined} \\ \boldsymbol{o} & \text{otherwise} \end{cases}$$

The Model

We now introduce a *bounded domain*, $\mathbf{D}^{n,m}$, as follows:

$$\mathbf{D}^{n,m} =_{df} \{\mathbf{D}_t \mid w(t) \leq n \,\&\, h(t) \leq m\}$$

In other words, a bounded domain collects all the domains of the types t within the width and height bounds n and m.

In order to preserve the information about whether objects in the domain are ordinary or abstract, we define two *indicator* functions $\boldsymbol{A}$ and $\boldsymbol{O}$ defined as follows. Where $\boldsymbol{o}$ is again a variable ranging over the entities of domain $\mathbf{D}_t$, then for all $\boldsymbol{o} \in \mathbf{D}_t$:

$$\boldsymbol{A}(\boldsymbol{o}) = \begin{cases} \mathbf{T} & \text{if } \boldsymbol{o} \in \mathbf{A}_t, \text{ where } \boldsymbol{o} \text{ is of type } t \\ \mathbf{F} & \text{otherwise} \end{cases}$$

$$\boldsymbol{O}(\boldsymbol{o}) = \begin{cases} \mathbf{T} & \text{if } \boldsymbol{o} \in \mathbf{O}_t, \text{ where } \boldsymbol{o} \text{ is of type } t \\ \mathbf{F} & \text{otherwise} \end{cases}$$

We next define an *interpretation* $\mathbf{V}$ that assigns to each constant κ of the language an element of an appropriate domain of quantification, i.e.,

If κ is a constant of type t, $\mathbf{V}(\kappa) \in \mathbf{D}_t$

$\mathbf{V}$ also assigns a special entity to each predicate $A!$ of the language:

$$\mathbf{V}(A!^{\langle t \rangle}) = \mathbf{S}_t$$

That is, the interpretation of the predicate constant $A!^{\langle t \rangle}$ is the set $\mathbf{S}_t$ of proxy elements of type t. Using the definitions above, we then define a model $\mathbf{M}$ as a structure of the form:

$$\mathbf{M} = \langle \mathbf{D}^{n,m}, \|\cdot\|, \boldsymbol{A}, \boldsymbol{O}, \mathbf{V} \rangle$$

where $\mathbf{D}^{n,m}$ is a bounded domain, $\|\cdot\|$ is an extended proxy function, $\boldsymbol{A}$ and $\boldsymbol{O}$ are the indicator functions that identify which elements of the bounded domain are abstract and ordinary, respectively, and $\mathbf{V}$ is an interpretation function. These elements have all been defined as above.

As noted earlier, for analysis of mathematics developed in this paper, we need models with bounded domain $\mathbf{D}^{2,3}$. We leave the complete list of types included within this bound to a footnote.[56] Note that many of these types don't play a role in our analysis of mathematics and aren't represented in Figure 1.

---

[56] We list the types for each width and height:

Simultaneous Definition of Denotation and Truth

Assignments to the Variables

In the usual way, an assignment $\boldsymbol{f}$ to the variables is a function that takes each variable in the language to an element of the domain over which the variable ranges. Strictly speaking, $\boldsymbol{f}$ should be relativized to the model $\mathbf{M}$, but we now always suppress the index to $\mathbf{M}$. More specifically:

- If α^t is a variable of type t, $\boldsymbol{f}(\alpha^t) \in \mathbf{D}_t$.

Moreover, where α^t is a variable of type t and $\boldsymbol{o}$ is an object in the domain $\boldsymbol{D}_t$, we use the notation $\boldsymbol{f}[\alpha/\boldsymbol{o}]$ to refer to the assignment function just like $\boldsymbol{f}$ except that it assigns the object $\boldsymbol{o}$ to the variable α. And where $\alpha_1, \ldots, \alpha_n$ are variables of type $t_1, \ldots, t_n$, respectively, and $\boldsymbol{o}_1, \ldots, \boldsymbol{o}_n$ are objects in the domains $\boldsymbol{D}_{t_1}, \ldots, \boldsymbol{D}_{t_n}$, respectively, we use the notation $\boldsymbol{f}[\alpha_i/\boldsymbol{o}_i]_{i=1}^n$ to refer to the assignment just like $\boldsymbol{f}$ except that it assigns $\boldsymbol{o}_1, \ldots, \boldsymbol{o}_n$ to $\alpha_1, \ldots, \alpha_n$, respectively.

Denotation and Satisfaction

Relative to the model $\mathbf{M}$ and variable assignment $\boldsymbol{f}$, we next define, by simultaneous recursion, (a) the *denotation* $\boldsymbol{d_f}(\tau)$ of term τ, and (b) $\boldsymbol{f}$ *satisfies* φ. Strictly speaking, $\boldsymbol{d_f}(\tau)$ should also be indexed to the model $\mathbf{M}$, but we now always suppress the index to $\mathbf{M}$:

D1 Where κ is any constant of type t, $\boldsymbol{d_f}(\kappa) = \mathbf{V}(\kappa)$.

D2 Where α is an variable of type t, $\boldsymbol{d_f}(\alpha) = \boldsymbol{f}(\alpha)$.

S1 If Π is a term of type $\langle t_1, \ldots, t_n \rangle$ ($n \geq 1$), and $\tau_1, \ldots, \tau_n$ any terms of types $t_1, \ldots, t_n$, respectively, then $\boldsymbol{f}$ satisfies $\Pi \tau_1 \ldots \tau_n$ iff (a) $\boldsymbol{d_f}(\tau_1), \ldots,$ $\boldsymbol{d_f}(\tau_n)$ are all defined, (b) $\boldsymbol{O}(\boldsymbol{d_f}(\Pi)) = \mathbf{T}$ (which implies $\boldsymbol{d_f}(\Pi)$ is defined as well), and (c) $\langle \|\boldsymbol{d_f}(\tau_1)\|, \ldots, \|\boldsymbol{d_f}(\tau_n)\| \rangle \in \boldsymbol{d_f}(\Pi)$.

S2 If Π is any constant or variable of type $\langle \rangle$, then $\boldsymbol{f}$ satisfies Π iff $\boldsymbol{d_f}(\Pi) = \mathbf{T}$.

- $w = 1, h = 0$: $\mathbf{D}_i$
- $w = 1, h = 1$: $\mathbf{D}_{\langle \rangle}, \mathbf{D}_{\langle i \rangle}$
- $w = 1, h = 2$: $\mathbf{D}_{\langle \langle \rangle \rangle}, \mathbf{D}_{\langle \langle i \rangle \rangle}$
- $w = 1, h = 3$: $\mathbf{D}_{\langle \langle \langle \rangle \rangle \rangle}, \mathbf{D}_{\langle \langle \langle i \rangle \rangle \rangle}$
- $w = 2, h = 0$: (empty)
- $w = 2, h = 1$: $\mathbf{D}_{\langle i, i \rangle}$
- $w = 2, h = 2$: $\mathbf{D}_{\langle i, \langle \rangle \rangle}, \mathbf{D}_{\langle i, \langle i \rangle \rangle}, \mathbf{D}_{\langle \langle \rangle, i \rangle}, \mathbf{D}_{\langle \langle i \rangle, i \rangle}, \mathbf{D}_{\langle \langle \rangle, \langle \rangle \rangle}, \mathbf{D}_{\langle \langle \rangle, \langle i \rangle \rangle}, \mathbf{D}_{\langle \langle i \rangle, \langle \rangle \rangle}, \mathbf{D}_{\langle \langle i \rangle, \langle i \rangle \rangle}$
- $w = 2, h = 3$: $\mathbf{D}_{\langle \langle i, \langle \rangle \rangle \rangle}$, $\mathbf{D}_{\langle \langle i, \langle i \rangle \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle \rangle, i \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle i \rangle, i \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle \rangle, \langle \rangle \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle \rangle, \langle i \rangle \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle i \rangle, \langle \rangle \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle i \rangle, \langle i \rangle \rangle \rangle}$, $\mathbf{D}_{\langle i, \langle \langle \rangle \rangle \rangle}$, $\mathbf{D}_{\langle i, \langle \langle i \rangle \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle \rangle \rangle, i \rangle}$, $\mathbf{D}_{\langle \langle \langle i \rangle \rangle, i \rangle}$, $\mathbf{D}_{\langle \langle \rangle, \langle \langle \rangle \rangle \rangle}$, $\mathbf{D}_{\langle \langle \rangle, \langle \langle i \rangle \rangle \rangle}$, $\mathbf{D}_{\langle \langle i \rangle, \langle \langle \rangle \rangle \rangle}$, $\mathbf{D}_{\langle \langle i \rangle, \langle \langle i \rangle \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle \rangle \rangle, \langle \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle i \rangle \rangle, \langle \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle \rangle \rangle, \langle i \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle i \rangle \rangle, \langle i \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle \rangle \rangle, \langle \langle \rangle \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle \rangle \rangle, \langle \langle i \rangle \rangle \rangle}$, $\mathbf{D}_{\langle \langle \langle i \rangle \rangle, \langle \langle \rangle \rangle \rangle}$

S3 If Π is any term of type $\langle t \rangle$ and τ is any term of type t, then $\boldsymbol{f}$ satisfies $\tau\Pi$ iff (a) $\boldsymbol{d_f}(\tau)$ is defined, (b) $\boldsymbol{A}(\boldsymbol{d_f}(\tau)) = \mathbf{T}$, and (c) $\boldsymbol{d_f}(\Pi) \in \boldsymbol{d_f}(\tau)$.

S4 And so on for the clauses for negation, conditionals, universal quantification, etc. E.g., $\boldsymbol{f}$ satisfies $\forall\alpha\varphi$ iff $\forall\boldsymbol{o}(f[\alpha/\boldsymbol{o}]$ satisfies $\varphi)$.

D3 Where $[\lambda\alpha_1 \ldots \alpha_n\, \varphi]$ is any λ-expression $(n \geq 1)$, $\alpha_1, \ldots, \alpha_n$ are variables of type $t_1, \ldots, t_n$, respectively, and $\boldsymbol{o}_1, \ldots, \boldsymbol{o}_n$ are objects of type $t_1, \ldots, t_n$, respectively, then

$$\boldsymbol{d_f}([\lambda\alpha_1 \ldots \alpha_n\, \varphi]) = \{\langle \|\boldsymbol{o}_1\|, \ldots, \|\boldsymbol{o}_n\|\rangle \mid \boldsymbol{f}[\alpha_i/\boldsymbol{o}_i]_{i=1}^n \text{ satisfies } \varphi\},$$
where this set of n-tuples is an element of $\mathbf{O}_{\langle t_1, \ldots, t_n \rangle}$.

In the special case where $[\lambda\alpha_1 \ldots \alpha_n\, \varphi]$ is *elementary*, i.e., has the form $[\lambda\alpha_1 \ldots \alpha_n\, \Pi\alpha_1 \ldots \alpha_n]$, then the above definition has the consequence that

- if $\boldsymbol{O}(\boldsymbol{d_f}(\Pi)) = \mathbf{T}$, then $\boldsymbol{d_f}([\lambda\alpha_1 \ldots \alpha_n\, \varphi]) = \boldsymbol{d_f}(\Pi)$, and
- if $\boldsymbol{A}(\boldsymbol{d_f}(\Pi)) = \mathbf{T}$, then $\boldsymbol{d_f}([\lambda\alpha_1 \ldots \alpha_n\, \varphi]) = \{\}$ in $\mathbf{O}_{\langle t_1, \ldots, t_n \rangle}$[57]

D4 Where α is any variable of type t and $\boldsymbol{o}$ is any object of type t, then

$$\boldsymbol{d_f}(\iota\alpha\varphi) = \begin{cases} \boldsymbol{o} \in \mathbf{D}_t, \text{if } \boldsymbol{f}[\alpha/\boldsymbol{o}] \text{ satisfies } \varphi\ \& \\ \quad \forall\boldsymbol{o}'(\boldsymbol{f}[\alpha/\boldsymbol{o}'] \text{ satisfies } \varphi \rightarrow \boldsymbol{o}' = \boldsymbol{o}) \\ \text{undefined, otherwise} \end{cases}$$

D5 And so on for the other cases where φ is a complex term of type $\langle\,\rangle$, i.e., where φ is any complex propositional formula. E.g., $\boldsymbol{d_f}(\neg\varphi) = \mathbf{T}$ iff $\boldsymbol{d_f}(\varphi) = \mathbf{F}$.

Truth

In the usual manner we say that φ is true just in case every assignment $\boldsymbol{f}$ satisfies φ.

Axioms

Since we've assumed classical logic in the description of our model, it remains to show that the axiom groups of Section 4.1 and 4.2 are true in the above model. It is easy to see that the following lemma holds:

Substitution Lemma. If $\boldsymbol{d_f}(\tau)$ is defined and φ is any formula, then $\boldsymbol{f}$ satisfies φ_α^τ if and only if $\boldsymbol{f}[\alpha/\boldsymbol{d_f}(\tau)]$ satisfies φ.

In other words, $\boldsymbol{f}$ satisfies φ_α^τ whenever the assignment just like $\boldsymbol{f}$ except that it assigns $\boldsymbol{d_f}(\tau)$ to α satisfies φ (assuming $\boldsymbol{d_f}(\tau)$ is defined). This Lemma holds because, according to our semantics, the truth value of an atomic formula is

---

[57] In other words, if the head relation term in an elementary λ-expression denotes an abstract relation, then the λ-expression denotes an *ordinary* relation (of type $\langle t_1, \ldots, t_n \rangle$ that is never exemplified.

calculated in terms of the denotations of its terms, assuming they all have such, and this feature is inherited by all the molecular and quantified formulas built out of such formulas.

In what follows, we omit the proofs that the axioms for classical logic hold in our model. We also omit proofs for the Hintikka (1959) axiom for descriptions (AXIOM 1) and AXIOMS 4 – 7. These are obviously true in the model (since the model was constructed in part to make these axioms true).[58] It remains to show only that the three distinctive principles, the AXIOM for the substitution of identicals, and AXIOMS 2 and 3, are true in the model.

Axiom: Substitution of Identicals

The proof that substitution of identicals is true in the model is by cases. The cases are:

$$x = y$$
$$F = G$$
$$R = S$$
$$p = q$$

Consider the first case:

> Assume some assignment, say $\boldsymbol{f}$, satisfies $x = y$. We want to show $\boldsymbol{f}$ satisfies φ if and only if $\boldsymbol{f}$ satisfies φ', where φ' is the result of replacing 0 or more free occurrences of x by y in φ. Now our assumption implies, by definition of $x = y$, that $\boldsymbol{f}$ satisfies:
>
> $(O!x \ \& \ O!y \ \& \ \forall F(Fx \equiv Fy)) \lor (A!x \ \& \ A!y \ \& \ \forall F(xF \equiv yF))$
>
> But since there are no ordinary individuals in the model (the domain $\mathbf{O}_i$ is empty), it follows that $\boldsymbol{f}$ satisfies:
>
> $A!x \ \& \ A!y \ \& \ \forall F(xF \equiv yF)$
>
> So by S4, $\boldsymbol{f}$ satisfies $\forall F(xF \equiv yF)$. Now by S3, we know:
>
> $\boldsymbol{f}$ satisfies 'xF' iff (a) $\boldsymbol{d_f}(x)$ is defined, (b) $\boldsymbol{A}(\boldsymbol{d_f}(x)) = \mathbf{T}$, and (c) $\boldsymbol{d_f}(F) \in \boldsymbol{d_f}(x)$.
>
> And we know something analogous for $\boldsymbol{f}$ satisfies yF. Now suppose for reductio that $\boldsymbol{d_f}(x) \neq \boldsymbol{d_f}(y)$. Since both objects are in $\mathbf{A}_i$, they are both sets of type $\langle i \rangle$ properties, so there must be a property that is an element of one, say, $\boldsymbol{d_f}(x)$, and not the other, that is, $\boldsymbol{d_f}(y)$. Let κ be such a property, so that we know $\kappa \in \boldsymbol{d_f}(x)$ and not $\kappa \in \boldsymbol{d_f}(y)$. Now

---

[58]For example, AXIOM 4 asserts that objects encoding properties are abstract. AXIOM 5 asserts that if a relation is exemplified, it is not abstract. AXIOM 6 asserts that λ-expressions don't denote abstract relations. And, given the definition of an *elementary* λ-expression in D3 above, AXIOM 7 asserts that η-Conversion holds for elementary λ-expressions in which the head relation is ordinary. It should be relatively straightforward to see that these are all true in the model.

consider the variable assignment $\boldsymbol{f}[F/\kappa]$. Since $\kappa \in \boldsymbol{d_f}(x)$, it follows by S3 that $\boldsymbol{f}[F/\kappa]$ satisfies xF. And since $\kappa \notin \boldsymbol{d_f}(y)$, it follows by S3 that $\boldsymbol{f}[F/\kappa]$ doesn't satisfy yF. So by the biconditional clause of S4, $\boldsymbol{f}[F/\kappa]$ doesn't satisfy $xF \equiv yF$. Then, by the universal quantifier clause of S4, $\boldsymbol{f}$ doesn't satisfy $\forall F(xF \equiv yF)$, This contradicts the assumed fact that $\boldsymbol{f}$ does satisfy $\forall F(xF \equiv yF)$. Hence, $\boldsymbol{d_f}(x) = \boldsymbol{d_f}(y)$. So by reasoning from the Substitution Lemma, we can argue as follows: $\boldsymbol{f}$ satisfies φ if and only if $\boldsymbol{f}[x/\boldsymbol{d_f}(x)]$ satisfies φ iff $\boldsymbol{f}[x/\boldsymbol{d_f}(y)]$ satisfies φ iff $\boldsymbol{f}$ satisfies φ'.

Given the definitions for identity in Section 3, the proof of the remaining cases, i.e., $F = G$, $R = S$, and $p = q$, are similar to the above.[59]

AXIOM 2: λ-Conversion

To see AXIOM 2 holds, note that the axiom is trivially true in the case where $[\lambda\alpha_1 \ldots \alpha_n\, \varphi]$ is an elementary λ-expression. For in that case, the λ-expression has the form:

$$[\lambda\alpha_1 \ldots \alpha_n\, \Pi\alpha_1 \ldots \alpha_n]$$

where $\alpha_1, \ldots, \alpha_n$ have types $\langle t_1, \ldots, t_n \rangle$, respectively, and Π is any simple n-place relation term (i.e., constant or variable) of type $\langle t_1, \ldots, t_n \rangle$. So, by the consequence noted at the end of D3, the following holds:

$$[\lambda\alpha_1 \ldots \alpha_n\, \Pi\alpha_1 \ldots \alpha_n]\beta_1 \ldots \beta_n \equiv \Pi\beta_1 \ldots \beta_n$$

For there are two cases: if Π denotes an ordinary relation, then the denotation of the λ-expression is just the denotation of Π itself, and if Π denotes an abstract relation, then Π denotes the empty set and both sides of the biconditional $[\lambda\alpha_1 \ldots \alpha_n\, \Pi\alpha_1 \ldots \alpha_n]\beta_1 \ldots \beta_n \equiv \Pi\beta_1 \ldots \beta_n$ are false (given that the contrapositive of AXIOM 5 tells us that abstract relations aren't exemplified).

When $[\lambda\alpha_1 \ldots \alpha_n\, \varphi]$ is non-elementary, then φ is any propositional formula that is free of descriptions (by the clause for Complex terms in the definition of our language). Our semantic definition of truth requires us to show that every assignment function $\boldsymbol{f}$ satisfies:

$[\lambda\alpha_1 \ldots \alpha_n\, \varphi]\beta_1 \ldots \beta_n \equiv \varphi^{\beta_1 \ldots \beta_n}_{\alpha_1, \ldots, \alpha_n}$, provided β_i is substitutable for α_i in φ $(1 \leq i \leq n)$

For simplicity and ease of readability, we prove this only for the 1-place case, so that where α, β are variables of some type t, our task is to show that every assignment function $\boldsymbol{f}$ satisfies:

$[\lambda\alpha\, \varphi]\beta \equiv \varphi^{\beta}_{\alpha}$, provided β is substitutable for α in φ

---

[59]For example, if $\boldsymbol{f}$ satisfies $F = G$, it satisfies $(O!F \,\&\, O!G \,\&\, \forall x(xF \equiv xG)) \lor (A!F \,\&\, A!G \,\&\, \forall\mathcal{H}(F\mathcal{H} \equiv G\mathcal{H}))$. If the values of F and G are both in $\mathbf{O}_{\langle i \rangle}$ and the same individuals encode them (which in the model means they are members of the same members of $\mathbf{A}_i$), then they are identical. And if the values of F and G are both in $\mathbf{A}_{\langle i \rangle}$, and they encode the same properties of properties (which in the model means they have the same members), they are identical.

Moreover, if we use $x, y, z, \ldots$ as arbitrarily chosen variables of type t, we simply have to show:

$[\lambda x \, \varphi]y \equiv \varphi^y_x$, provided y is substitutable for x in φ

So suppose $\boldsymbol{f}$ is an arbitrary assignment function. Then, by the clauses in S4, we have to show $\boldsymbol{f}$ satisfies $[\lambda x \, \varphi]y$ if and only if $\boldsymbol{f}$ satisfies φ^y_x:

Note that by clause S1, $\boldsymbol{f}$ satisfies $[\lambda x \, \varphi]y$ if and only if:

(a) $\boldsymbol{d}_f(y)$ is defined

(b) $\boldsymbol{O}(\boldsymbol{d}_f([\lambda x \, \varphi])) = \mathbf{T}$, i.e., $\boldsymbol{d}_f([\lambda x \, \varphi]) \in \boldsymbol{O}_{\langle t \rangle}$, by definition of $\boldsymbol{O}$

(c) $\|\boldsymbol{d}_f(y)\| \in \boldsymbol{d}_f([\lambda x \, \varphi])$

But (a), (b) and (c) hold if and only if:

$\|\boldsymbol{d}_f(y)\| \in \{\|\boldsymbol{o}\| \mid \boldsymbol{f}[x/\boldsymbol{o}]$ satisfies $\varphi\}$,
where this latter set is an element of $\boldsymbol{O}_{\langle t \rangle}$

by clause D3.[60] So by set-abstraction, the fact that φ is propositional and description-free, and the **Lemma on Proxies** (see below), the above holds if and only if:

$\boldsymbol{f}[x/\boldsymbol{d}_f(y)]$ satisfies φ

i.e., by the Substitution Lemma, if and only if:

$\boldsymbol{f}$ satisfies φ^y_x

For this conclusion to hold, it remains only to show:

Lemma on Proxies: Let x be a variable of type t. Then if φ is a description-free propositional formula and and $\boldsymbol{o}$ and $\boldsymbol{o}'$ have the same proxy, then:

L1 for any term τ in φ, $\|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau)\| = \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\tau)\|$, and

L2 $f[x/\boldsymbol{o}]$ satisfies φ iff $f[x/\boldsymbol{o}']$ satisfies φ

Proof. The proof proceeds by induction on the λ-rank of propositional formulas φ, i.e., how deeply nested is the deepest λ expressions in φ. But in what follows, the notion of λ-rank applies to any formula or term.[61] (We ignore the trivial case where x doesn't occur free in φ.) For L2, without loss of generality, we

---

[60]The left-to-right direction, i.e., from (a), (b) and (c) to this conclusion, is immediate from D3 and (c). Now for the right-to-left direction, assume that $\|\boldsymbol{d}_f(y)\| \in \{\|\boldsymbol{o}\| \mid \boldsymbol{f}[x/\boldsymbol{o}]$ satisfies $\varphi\}$. But since y is a variable, $\|\boldsymbol{d}_f(y)\|$ is defined and so (a) holds. And since $\|\boldsymbol{d}_f(y)\|$ is an element of $\{\|\boldsymbol{o}\| \mid \boldsymbol{f}[x/\boldsymbol{o}]$ satisfies $\varphi\}$, the latter can't be empty, and so by D3, $\boldsymbol{d}_f([\lambda x \, \varphi])$ is an element of $\boldsymbol{O}_{\langle t \rangle}$. Thus (b) holds. And by D3, (c) holds.

[61]If φ (or τ) contains no λ-expressions, it has a λ-rank of 0. If no λ-expression in φ (τ) contains a λ-expression, then its λ-rank is 1. If φ (τ) contains a λ-expression whose matrix has λ-rank n and no λ-expression in φ (τ) has a λ-rank greater than n, then φ (τ) has a λ-rank of $n + 1$.

need only prove the left to right direction, i.e., that if $f[x/\boldsymbol{o}]$ satisfies φ, then $f[x/\boldsymbol{o}']$ satisfies φ. So assume $f[x/\boldsymbol{o}]$ satisfies φ.

The base case is where φ is a formula of λ-rank 0, i.e., without λ-expressions. We first consider atomic formulas φ of the form $\Pi^n \tau_1 \ldots \tau_n$ ($n \geq 0$) where none of $\Pi^n, \tau_1, \ldots, \tau_n$ are λ-expressions. Note that to prove L1, we must show that (1) $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\Pi) = \boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\Pi)$ and (2) for each τ_i, $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau_i) = \boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\tau_i)$. We will prove (1) and (2) in the course of proving L2. There are two subcases for L2: $n \geq 1$ or $n = 0$. When $n \geq 1$, then each of $\Pi^n, \tau_1, \ldots, \tau_n$ is either a constant or a variable. And when $n = 0$, then φ has the form Π where Π is a constant or a variable of the empty type. We cover these two subcases in turn.

In the first subcase ($n \geq 1$), φ is governed by S1. So φ has the form $\Pi^n \tau_1 \ldots \tau_n$ and contains no λ-expressions. So we know[62] (i) x is either Π^n or one of the τ_i, (ii) $\tau_1, \ldots, \tau_n$ have any types and (iii) Π is of an appropriate type to relate $\tau_1, \ldots, \tau_n$. Then, since $f[x/\boldsymbol{o}]$ satisfies φ, it follows that:

(a) $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau_1), \ldots, \boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau_n)$ are all defined

(b) $\boldsymbol{O}(\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\Pi)) = \mathbf{T}$

(c) $\langle \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau_1)\|, \ldots, \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau_n)\| \rangle \in \boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\Pi)$.

We're trying to show $\boldsymbol{f}[x/\boldsymbol{o}']$ satisfies φ, i.e., all of (d) – (f) have to hold:

(d) $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\tau_1), \ldots, \boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\tau_n)$ are all defined.

(e) $\boldsymbol{O}(\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\Pi)) = \mathbf{T}$.

(f) $\langle \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\tau_1)\|, \ldots, \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\tau_n)\| \rangle \in \boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\Pi)$.

Proof of (d). This follows from (a) because for any term τ, if $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau)$ is defined, then $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\tau)$ is defined. (If $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\tau)$ were undefined, then τ would have to be a description. But φ is description-free.)

Proof of (e). This follows from (b) by cases. (i) If Π is a constant or a variable other than x, then $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\Pi) = \boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\Pi)$, because $\boldsymbol{f}[x/\boldsymbol{o}]$ and $\boldsymbol{f}[x/\boldsymbol{o}']$ differ only by their assignment to the variable x which is different than Π. So if $\boldsymbol{O}(\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\Pi)) = \mathbf{T}$, then $\boldsymbol{O}(\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\Pi)) = \mathbf{T}$. (ii) If Π is x, then by (b), $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\Pi)) = \boldsymbol{o}$. Since $\boldsymbol{o}$ and $\boldsymbol{o}'$ have the same proxy and $\boldsymbol{O}(\boldsymbol{o}) = \mathbf{T}$ (i.e., $\boldsymbol{o}$ is ordinary), $\boldsymbol{o} = \boldsymbol{o}'$. So $\boldsymbol{O}(\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\Pi)) = \mathbf{T}$. Note that we have now proved part (1) of L1 where φ falls under the first subcase.

Proof of (f). There are 2 cases to consider: one or more of the τ_i is x or Π is x. Suppose one or more of the τ_i is x. Then we note that $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\Pi) = \boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\Pi)$ because x does not occur in Π (because Π isn't a λ-expression and x has to have a different type from Π and so can't be Π). Moreover, for each τ_i, $\|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau_i)\| = \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}']}(\tau_i)\|$. If τ_i is x in which case this follows from the fact that $\boldsymbol{o}$ and $\boldsymbol{o}'$ have the same proxy. Otherwise, τ_i is a constant or or a variable

---

[62]We need not consider the case where Π contains x free as proper subterm because the only way for that to happen is if Π were a λ-expression (which is ruled out in the base case) or if Π were a description (which is ruled out because φ is description-free).

other than x and so the denotation of τ_i under $\boldsymbol{f}[x/\boldsymbol{o'}]$ is the same as that under $\boldsymbol{f}[x/\boldsymbol{o}]$. The claim then follows from (c). Alternatively, suppose then that Π is x. Then $\|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau_i)\| = \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\tau_i)\|$ (since in this case, x isn't one of the τ_i). Since, by (b), Π is ordinary, then by the argument in (e), $\boldsymbol{o} = \boldsymbol{o'}$, and so the claim is trivially true by (c). Note that we have now proved part (2) of L1 where φ falls under the first subcase.

In the second subcase ($n = 0$), φ has the form Π where Π is of the empty type. So Π can only be a constant or a variable. If φ is a constant or variable of type $\langle\,\rangle$, S2 applies. Then either Π is either a constant, or a variable other than x, or x itself. If Π is a constant or a variable other than x, then $\boldsymbol{f}[x/\boldsymbol{o}]$ satisfies Π iff $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\Pi) = \mathbf{T}$ (by S2) iff $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\Pi) = \mathbf{T}$ (since Π doesn't contain a free occurrence of x) iff $\boldsymbol{f}[x/\boldsymbol{o'}]$ satisfies Π (by S2). If Π is x, then $\boldsymbol{f}[x/\boldsymbol{o}]$ satisfies φ iff $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(x) = \mathbf{T}$ (by S2) iff $\boldsymbol{o} = \mathbf{T}$ iff $\boldsymbol{o'} = \mathbf{T}$ (see below) iff $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(x) = \mathbf{T}$ iff $\boldsymbol{f}[x/\boldsymbol{o'}]$ satisfies φ. To see that $\boldsymbol{o} = \mathbf{T}$ iff $\boldsymbol{o'} = \mathbf{T}$, recall that $\|\boldsymbol{o'}\| = \|\boldsymbol{o}\|$ (by hypothesis) and since $\mathbf{T} \notin \mathbf{S}_{\langle\,\rangle}$, the only object with $\mathbf{T}$ at its proxy is $\mathbf{T}$ itself. So $\|\boldsymbol{o'}\| = \|\boldsymbol{o}\| = \mathbf{T}$ iff $\boldsymbol{o'} = \boldsymbol{o} = \mathbf{T}$. Note that this proves part (1) of L1 where φ falls under the second subcase. There is no part (2) of L1 for this subcase.

We now have that L1 and L2 hold for atomic formulas φ of λ-rank 0. We conclude this base case by noting that L1 and L2 hold for complex formulas φ of the form $\neg\psi$, $\psi \to \chi$, and $\forall\alpha\psi$. By S4 and D5, the truth of L1 and L2 for these complex formulas is grounded in the truth of L1 and L2 for the atomic formulas with no λ-expressions given that complex formulas of rank 0 contain no λ-expressions.

Inductive cases: IH: The lemma holds for ψ with λ-rank of n or less, i.e., for ψ with λ-rank of n or less we may assume:

> IH-L1: $\|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau)\| = \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\tau)\|$, for any term τ in ψ, and

> IH-L2: $f[x/\boldsymbol{o}]$ satisfies ψ if and only if $f[x/\boldsymbol{o'}]$ satisfies ψ.

We need to show that it holds for φ with λ-rank of $n + 1$.

To show L1, we need to show that $\|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau)\| = \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\tau)\|$, for any term τ in φ. We first consider the case where φ is atomic. Fix an arbitrary such τ and consider its denotation:

- D1 and D2 only apply if τ has a λ-rank of 0, so the result follows immediately from IH-L1.

- D3 applies when τ is of the form $[\lambda\alpha_1 \ldots \alpha_n\, \psi]$. Now if τ has λ-rank n or less, then the result follows by IH-L1. So we need only be concerned with the case when τ has λ-rank $n + 1$ where ψ has λ-rank n. We have to show $\|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}([\lambda\alpha_1 \ldots \alpha_n\, \psi])\| = \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}([\lambda\alpha_1 \ldots \alpha_n\, \psi])\|$. Note that D3 defines the denotation of the λ-expression as the set of n-tuples (of proxies of objects) such that $\boldsymbol{f}[\alpha_i/\boldsymbol{o}_i]_{i=1}^n$ satisfies ψ. But ψ has λ-rank n (one less than τ), so by IH-L2, $\boldsymbol{f}[\alpha_i/\boldsymbol{o}_i]_{i=1}^n$ satisfies ψ if and only if $\boldsymbol{f}[\alpha_i/\boldsymbol{o'}_i]_{i=1}^n$ satisfies ψ, so the set of proxy-tuples must be the same and L1 holds.

- D4 would apply if τ were a description. But φ is description-free.

We now have the L1 holds for atomic formulas φ of λ-rank up to $n+1$. We can then conclude that L1 holds for complex formulas φ of the form $\neg\psi$, $\psi \to \chi$, and $\forall\alpha\psi$. By D5, the truth of L1 for these complex formulas is grounded in the truth of L1 for the atomic formulas with λ-rank up to $n+1$ given that a complex formula can not have a λ-rank higher than its component atomic formulas.

To show L2, we have to show: $f[x/\boldsymbol{o}]$ satisfies φ if and only if $f[x/\boldsymbol{o'}]$ satisfies φ, when φ has an λ-rank of $n+1$. We now make use of L1, which has been proved for all λ-ranks. So we know that for any term τ in φ, $\|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau)\| = \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\tau)\|$. We first consider the case where φ is atomic. Moreover, we need only consider formulas governed by S1, for (a) if the λ-rank of φ is some n other than 0, then φ is not a constant or variable, and so S2 doesn't apply, (b) S3 doesn't apply since that governs encoding formulas, which aren't propositional. So we are consider φ of the form $\Pi\tau_1 \ldots \tau_n$ where any of Π, $\tau_1,\ldots,\tau_n$ can have a λ-rank of $n+1$. Without loss of generality, we need only prove the left to right direction, i.e., that if $f[x/\boldsymbol{o}]$ satisfies φ, then $f[x/\boldsymbol{o'}]$ satisfies φ. So assume $f[x/\boldsymbol{o}]$ satisfies φ. By assumption, it follows that:

(a) $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau_1),\ldots,\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau_n)$ are all defined.

(b) $\boldsymbol{O}(\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\Pi)) = \mathbf{T}$.

(c) $\langle\|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau_1)\|,\ldots,\|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau_n)\|\rangle \in \boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\Pi)$.

We're trying to show $\boldsymbol{f}[x/\boldsymbol{o'}]$ satisfies φ, i.e., all of (d) – (f) have to hold:

(d) $\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\tau_1),\ldots,\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\tau_n)$ are all defined.

(e) $\boldsymbol{O}(\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\Pi)) = \mathbf{T}$.

(f) $\langle\|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\tau_1)\|,\ldots,\|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\tau_n)\|\rangle \in \boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\Pi)$.

Proof of (d). This follows by the exact reasoning for clause (d) in the base case, which did not depend on the λ-rank of τ_i.

Proof of (e). There are three cases to consider, two of which are identical to the proof of clause (e) in the base case. The three cases are (i) Π is a constant or a variable other than x, (ii) Π is x, or (iii) Π is a λ-expression. So it remains to show only (iii). Π is a λ-expression and by D3, the denotation of all λ-expressions are ordinary. Thus (e) follows since $\boldsymbol{O}(\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\Pi)) = \mathbf{T}$.

Proof of (f). By assumption (b), we know that the denotation of Π is ordinary. And since the denotation of Π is ordinary, the proxy of the denotation is just equal to the denotation itself. So:

$$
\begin{aligned}
\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\Pi) \quad &= \quad \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\Pi)\| \quad &&\text{(since Π is ordinary)} \\
&= \quad \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\Pi)\| \quad &&\text{(by L1)} \\
&= \quad \boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\Pi) \quad &&\text{(since Π is ordinary)}
\end{aligned}
$$

Moreover, we also know, by L1, that the proxies of the denotations for each τ are identical, that is for each τ_i $\|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o}]}(\tau_i)\| = \|\boldsymbol{d}_{\boldsymbol{f}[x/\boldsymbol{o'}]}(\tau_i)\|$. So the tuple in

(f) is identical to the tuple in (c). Hence it follows that (f) must hold if (c) holds.

We now have that L2 holds for atomic formulas φ of λ-rank $n + 1$. We conclude this inductive case by noting that L2 holds for complex formulas φ of the form $\neg\psi$, $\psi \to \chi$, and $\forall\alpha\psi$. By S4, truth of L2 for these complex formulas is grounded in the truth of L2 for the atomic formulas with λ-rank less than or equal to $n + 1$, given that complex formulas only have rank $n + 1$ in virtue of the ranks of their atomic components. That is, S4 doesn't increase the λ-rank.

AXIOM 3: Comprehension Principles for Abstracta

At the beginning of the appendix, we noted that our analysis of mathematics does not require the full unbounded language. In general we need only validate comprehension up to a given fixed type height h. So we fix n and m such that the bounded language $\mathcal{L}_{n,m}$ has $m > h$. Now we will show AXIOM 3 holds in the model for any arbitrary t such that $h(t) < m$. Recall that by definition (2) of $\mathbf{A}_t$ (in the Appendix subsection Structural Domains: Abstract Objects), when $h(t) < m$, we have that $\mathbf{A}_t = \wp(\mathbf{O}_{\langle t \rangle} \cup \mathbf{A}_{\langle t \rangle})$. We now show that $\exists\alpha(A!\alpha \,\&\, \forall F(\alpha F \equiv \varphi))$ holds by showing that $\{F \mid \varphi\}$ defines an abstract object of type t.

Proof. Pick an arbitrary formula φ, where α of type t doesn't occur free in φ. To show that the instance $\exists\alpha(A!\alpha \,\&\, \forall F(\alpha F \equiv \varphi))$ is true in the model, we have to show that every assignment $\boldsymbol{f}$ satisfies this instance. So pick an arbitrary assignment $\boldsymbol{f}$. By S4 (and the definition of the existential quantifier), we have to show that for some element, say $\boldsymbol{o}^*$, in the domain over which α ranges, $\boldsymbol{f}[\alpha/\boldsymbol{o}^*]$ satisfies $A!\alpha \,\&\, \forall F(\alpha F \equiv \varphi)$. So again by S4, we have to show that for some element $\boldsymbol{o}^*$ in the domain over which α ranges, (a) $\boldsymbol{f}[\alpha/\boldsymbol{o}^*]$ satisfies $A!\alpha$ and (b) $\boldsymbol{f}[\alpha/\boldsymbol{o}^*]$ satisfies $\forall F(\alpha F \equiv \varphi)$. So we show that when we choose the set $\{\boldsymbol{o} \mid \boldsymbol{f}[F/\boldsymbol{o}] \text{ satisfies } \varphi\}$ as our witnesss, both (a) and (b) hold.

(a) To show:

$$\boldsymbol{f}[\alpha/\{\boldsymbol{o} \mid \boldsymbol{f}[F/\boldsymbol{o}] \text{ satisfies } \varphi\}] \text{ satisfies } A!\alpha$$

we have to show, by S1, that:

(i) $\boldsymbol{d}_{\boldsymbol{f}[\alpha/\{\boldsymbol{o}\mid\boldsymbol{f}[F/\boldsymbol{o}]\text{ satisfies } \varphi\}]}(\alpha)$ is defined,

(ii) $\boldsymbol{O}(\boldsymbol{d}_{\boldsymbol{f}[\alpha/\{\boldsymbol{o}\mid\boldsymbol{f}[F/\boldsymbol{o}]\text{ satisfies } \varphi\}]}(A!)) = \mathbf{T}$, and

(iii) $\|\boldsymbol{d}_{\boldsymbol{f}[\alpha/\{\boldsymbol{o}\mid\boldsymbol{f}[F/\boldsymbol{o}]\text{ satisfies } \varphi\}]}(\alpha)\| \in \boldsymbol{d}_{\boldsymbol{f}[\alpha/\{\boldsymbol{o}\mid\boldsymbol{f}[F/\boldsymbol{o}]\text{ satisfies } \varphi\}]}(A!)$.

I.e., by applying definitions, we have to show:

(i) $\{\boldsymbol{o} \mid \boldsymbol{f}[F/\boldsymbol{o}] \text{ satisfies } \varphi\} \in \mathbf{A}_t$.

(ii) $A!$ is ordinary.

(iii) $\|\{\boldsymbol{o} \mid \boldsymbol{f}[F/\boldsymbol{o}] \text{ satisfies } \varphi\}\| \in \mathbf{V}(A!^{\langle t \rangle}) = \mathbf{S}_t$.

(ii) is trivial; (iii) follows directly from (i) by definition of the proxy function, and so it remains to show (i).

Note that $\{\boldsymbol{o} \mid \boldsymbol{f}[F/\boldsymbol{o}] \text{ satisfies } \varphi\}$ is a fixed set of objects of type $\langle t \rangle$. But all objects of type $\langle t \rangle$ are in $\mathbf{O}_{\langle t \rangle}$ or $\mathbf{A}_{\langle t \rangle}$. Since $h(t) < m$, it follows that $\{\boldsymbol{o} \mid \boldsymbol{f}[F/\boldsymbol{o}] \text{ satisfies } \varphi\} \in \wp(\mathbf{O}_{\langle t \rangle} \cup \mathbf{A}_{\langle t \rangle}) = \mathbf{A}_t$.