

PROBLEMS WITH FOUNDATIONALISM IN THE PHILOSOPHY OF MATHEMATICS

1. INTRODUCTION

All of mathematics can be reduced to set theory, or so we are told. But this claim is rarely supported with a real argument. Normally it is treated as obvious, as immediately following from the fact that all mathematics can be interpreted in set theory, or as so well-established that it can be asserted without either argument or citation. Proponents of alternative foundations for mathematics make similar moves in presenting their foundationalist programmes. In this paper, I will consider one way of arguing for this claim: interpretability is logically sufficient for being reducible. In particular, I will argue that interpretability is not logically sufficient for being reducible. Foundationalists who use this line of reasoning are either employing a suppressed premiss or their line of reasoning is logically invalid.

The argument for this claim is complex. My plan is to proceed as follows: In §2 I present the definitions – both mathematical and philosophical – necessary to state the claim which I am criticising. In §3 I show that the inferential leap corresponding to the claim appears in Quine’s foundationalist philosophy of mathematics. I then turn to my actual argument for three sections. The argument is broken into two claims. The first claim is further broken into two parts; these parts are considered in §§4-5. The second claim is a proof in model theory; this proof is carried out in §6. I close by reflecting on the philosophical implications of my arguments here.

2. DEFINITIONS

We must first give a long series of definitions. There are four primary definitions, all relations between two theories: O is interpretable in T , O is relatively interpretable in T , O is representable in T , and O is foundationally reducible in

Date: January 14, 2008.

I’d like to thank Michael Detlefsen and Sean Walsh for detailed comments on earlier versions of this paper.

principle to T . Each primary definition will require several preliminary definitions. My notation throughout this paper is mostly standard for model theorists. In my experience, philosophers are sometimes unfamiliar with a common model-theoretic shorthand: an arbitrary sequence of variables $\langle x_1, x_2, \dots, x_n \rangle$ is abbreviated as \bar{x} for purposes of both quantification and predication. For example,

$$\forall x_1 \forall x_2 \forall x_3 \cdots \forall x_n$$

is abbreviated as $\forall \bar{x}$. As a heuristic point, I will always use O for the ‘object theory’ – the theory whose terms are the definienda – and T for the ‘target theory’ – the theory whose terms are the definiens. In reductions of PA to ZFC , for example, PA is the object theory and ZFC is the target theory. With the exception of a very brief discussion of Frege, I assume throughout this paper that all theories are consistent, satisfiable, and first-order, and that the logic, both syntax and semantics, is standard.

2.1. Interpretable. While the basic idea of interpretability goes back at least to Descartes’ invention of Cartesian co-ordinates – thereby interpreting Euclidean geometry in what we would call \mathbb{R}^2 – our most rigorous definition of interpretability comes from Tarski.¹ Let O and T be two theories of languages \mathcal{L}_O and \mathcal{L}_T , respectively. Without loss of generality, \mathcal{L}_O and \mathcal{L}_T have no common non-logical vocabulary; if this is not the case, replace O with a theory O' in a language \mathcal{L}'_O whose non-logical symbols are ‘dummy symbols’. Let \mathcal{L} be the language whose non-logical symbols are precisely the non-logical symbols of \mathcal{L}_O and \mathcal{L}_T .² We first need to define the notion of possible definition.

Possible definition:

- Let R be a relation symbol of \mathcal{L}_O of arity k . A possible definition of R in \mathcal{L}_T is any sentence of the form

$$\forall \bar{x} (R(\bar{x}) \leftrightarrow \Phi(\bar{x})),$$

where Φ is a formula of \mathcal{L}_T with k free variables \bar{x} .

¹Tarski, Mostowski, & Robinson, 1953, 20ff

²Hence, every non-logical symbol in \mathcal{L} is a non-logical symbol of either \mathcal{L}_O or \mathcal{L}_T .

- Let c be a constant symbol of \mathcal{L}_O . A possible definition of c in \mathcal{L}_T is any sentence of the form

$$\forall x (x = c \leftrightarrow \Phi(x)),$$

where Φ is a formula of \mathcal{L}_T with x free.

- Let f be a function symbol of \mathcal{L}_O of arity k . A possible definition of f in \mathcal{L}_T is any sentence of the form

$$\forall \bar{x} \forall y (f(\bar{x}) = y \leftrightarrow \Phi(\bar{x}, y)),$$

where Φ is a formula of \mathcal{L}_T with $k + 1$ free variables \bar{x}, y .

Interpretable: O is interpretable in T if there is a theory T' in \mathcal{L} and a subset D of T' such that all of the following hold.

- (1) For every sentence φ of \mathcal{L}_O , if $O \vdash \varphi$ then $T' \vdash \varphi$. For every sentence ψ of \mathcal{L}_T , if $T \vdash \psi$ then $T' \vdash \psi$. That is, T' is an extension of both O and T .
- (2) D is a recursive set of possible definitions of the non-logical symbols of \mathcal{L}_O in \mathcal{L}_T .
- (3) Each non-logical symbol of \mathcal{L}_O occurs in exactly one sentence in D .
- (4) For every sentence φ of \mathcal{L} such that $T' \vdash \varphi$, $T \cup D \vdash \varphi$.

The classical examples of interpretation are the interpretations of Euclidean geometry and complex analysis as pairs of real numbers. As these interpretations are actually rather difficult to carry out with this definition of interpretable, we shall instead show that Hilbert's geometry³ is interpretable in complex analysis. Actually, as even this would take more space than I wish to spend here, I shall simply show how to begin constructing an interpretation.

Let Can be the theory of complex analysis. The non-logical symbols of Can are two constant symbols $0, 1$; two binary function symbols $+, \cdot$; and one unary function symbol $\bar{}$. Can has a countable axiomatisation. The axioms can be divided into several groups: (i) field axioms; (ii) axioms defining $\bar{}$ as complex conjugation; (iii) countably many algebraic closure axioms, one for each polynomial, asserting that it

³Hilbert, 1899/1950

has a root; (iv) countably many topological closure axioms, one for each definable set σ , asserting that, if σ is bounded, then it has a least upper bound. Note that, while we have not included a partial order here, one can be defined using the fact that the complex norm and positive real numbers can be defined in this language:

$$x \leq y =_{df} \exists u \exists v (u \cdot u = x \cdot \bar{x} \wedge v \cdot v = y \cdot \bar{y} \wedge \\ \exists w (\exists z w = z \cdot \bar{z} \wedge u + w = v))$$

Let H be the following fragment of Hilbert's geometry. H has two non-logical symbols: $L(x, y, z)$, which is true on the standard interpretation if and only if x, y , and z are colinear; and $B(x, y, z)$, which is true on the standard interpretation if and only if x, y , and z are colinear and x is between y and z . The axioms of this fragment will be formalisations of Hilbert's axioms I7 (the existence of three non-colinear points) and II1-5 (axioms of order).

Now, to give an interpretation of H in Can , we must give a set D of possible definitions for L and B . First, two abbreviations will be helpful:

$$R(z) =_{df} z = \bar{z} \\ P(z) =_{df} \exists w z = w \cdot \bar{w}$$

Informally, $R(z)$ is true if and only if z is a real number, and $P(z)$ is true if and only if z is a positive real number. Next, the key insight for our interpretation is that one-dimensional lines in the complex plane are sets of the following form: for u, v fixed complex numbers,

$$\{z | z = x \cdot u + v, x \text{ real}\}.$$

To assert that x, y and z are colinear, then, is to assert that there are u and v , parameters defining such a line-set, such that all three of x, y and z are elements of

this line-set. Our set D then contains these two possible definitions:

$$\begin{aligned} \forall x \forall y \forall z (L(x, y, z) \leftrightarrow \exists u \exists v (\exists w (R(w) \wedge x = w \cdot u + v) \wedge \\ \exists w (R(w) \wedge y = w \cdot u + v) \wedge \\ \exists w (R(w) \wedge z = w \cdot u + v))) \\ \forall x \forall y \forall z (B(x, y, z) \leftrightarrow \exists u \exists v \exists p \exists q \exists r \exists s \exists t (R(p) \wedge R(q) \wedge R(r) \wedge P(s) \wedge P(t) \wedge \\ x = q \cdot u + v \wedge y = p \cdot u + v \wedge z = r \cdot u + v \wedge \\ q = p + s \wedge r = q + t)) \end{aligned}$$

The final step would be to show that, using these two possible definitions and the axioms of Can , the axioms of H follow as theorems. For the sake of space, this is left as an exercise for the reader.

The set D is called an *interpretation* of O in T . This set gives us a scheme for **interpretation** translating any sentence of \mathcal{L}_O to a sentence of \mathcal{L}_T . Let R be an arbitrary non-logical symbol in \mathcal{L}_O . By (3) there is exactly one sentence in D containing the symbol R . This sentence contains a formula Φ_R in \mathcal{L}_T . Call Φ_R the *translation* or *interpretation* **translation** of R in D .⁴

We can further define the translation or interpretation of whole formulae in D . We illustrate this with an example. Let \mathcal{L}_R be the language of rings and R be the theory of rings in \mathcal{L}_R , with the non-logical symbols $0, 1, +, \times$. 0 and 1 are constant symbols and $+$ and \times are function symbols of arity 2. Given some target language \mathcal{L}_T and interpretation D of R in \mathcal{L}_T , let the translations in D be $\Phi_0, \Phi_1, \Phi_+, \Phi_\times$. Let φ be the sentence $1 + 0 = 0 + 1$. We translate this sentence from the language

⁴Since there may be more than one set D witnessing the fact that O is interpretable in T , and these interpretations may not be consistent (think of the two standard interpretations of the natural numbers in set theory), it is more proper to speak of an interpretation or translation of a sentence or term or symbol with respect to a given interpretation D than with respect to T itself. Still, in later sections, when the particular D does not matter, I will speak of interpretations and translations in T for the sake of convenience.

of rings to the target language in the following steps (which are not unique):

$$\begin{aligned}
& 1 + 0 = 0 + 1 \\
& \Phi_+(1, 0, 0 + 1) \\
& \exists x (\Phi_+(0, 1, x) \wedge \Phi_+(1, 0, x)) \\
& \exists y \exists z \exists x (y = 0 \wedge z = 1 \wedge \Phi_+(y, z, x) \wedge \Phi_+(z, y, x)) \\
& \exists y \exists z \exists x (\Phi_0(y) \wedge \Phi_1(z) \wedge \Phi_+(y, z, x) \wedge \Phi_+(z, y, x))
\end{aligned}$$

Denote this final sentence by φ' , and call φ' a translation of φ in D .⁵ By (1), since $R \vdash \varphi$, $T' \vdash \varphi$. Furthermore, $T' \vdash \varphi \leftrightarrow \varphi'$, so $T' \vdash \varphi'$. Indeed, if φ'' is another translation of φ in D , then $T' \vdash \varphi' \leftrightarrow \varphi''$. (Note that D is fixed here.) Also, since the proof of Proposition 2, below, also shows that T' is a conservative extension of T , $T \vdash \varphi' \leftrightarrow \varphi''$, so that translations of a given sentence in the object language are provably equivalent in the target language.

Next, by (4), there is a subset S of $T \cup D$ such that $S \vdash \varphi'$. This is the sense in which T , supplemented with the interpretation D , proves φ .

(1) says that T' is an extension of both O and T . Note that interpretability does *not* require T' to be a conservative extension of O . T' can prove things in \mathcal{L}_O that O does not prove. One way to put this intuitively is to say that T' can solve problems that O cannot. For example, perhaps Golbach's Conjecture is independent of PA , but is proved by the interpretation of PA in ZFC . On the other hand, as mentioned two paragraphs above, T' is a conservative extension of T .

If the set of non-logical symbols of \mathcal{L}_O is finite, then D will also be finite, and hence will be recursive. (2) does require the set of non-logical symbols of \mathcal{L}_O be countable. Since non-recursive languages are, to my knowledge, at best extremely uncommon in ordinary mathematical practice, (2) is not a deep restriction.

Combining (1), (2), and (4) with the assumption that T is axiomatisable gives the following: if φ is a sentence of \mathcal{L}_O such that $O \vdash \varphi$, then there is a set S of axioms for T and possible definitions for the relevant non-logical symbols in \mathcal{L}_T such that

⁵Since my terminology here has confused some readers, let me emphasise that Φ_R is a translation in D of the *non-logical symbol* R , while φ' is a translation in D of the *sentence* φ . If R is used in φ , then Φ_R will be used in φ' .

$S \vdash \varphi'$, where φ' is the translation of φ in the relevant D . So, in our example of rings, the axiomatisation of T supplemented with the translations of the non-logical symbols $0, 1$ and $+$ in D is sufficient to prove φ in the above sense.

Let us return to our more general context of O and T . Let f be a function symbol in \mathcal{L}_O . Then

$$O \vdash \forall \bar{x} \exists y \forall z (f(\bar{x}) = y \wedge (f(\bar{x}) = z \rightarrow z = y))$$

by the standard logic of function symbols. Call this sentence ψ_f . Then $T' \vdash \psi_f$ by (1). Let Φ_f be the translation of f in D and ψ'_f be the sentence

$$\forall \bar{x} \exists y \forall z (\Phi_f(\bar{x}, y) \wedge (\Phi_f(\bar{x}, z) \rightarrow z = y)).$$

(ψ'_f is the translation of ψ_f in D .) Then $T' \vdash \psi_f \leftrightarrow \psi'_f$, and so $T' \vdash \psi'_f$. That is, T proves that every function of O is well-defined, in the sense of proof given above. While Tarski requires this for Φ_f to be a possible definition for the function symbol f ,⁶ this was not necessary – the definition of interpretability already implies it.

2.2. Relatively interpretable. Interpretability is too strong of a condition for some purposes. For example, PA is not actually interpretable in ZF . We need a slightly weaker notion: relative interpretability.

We first define and comment on relativisation. Let T be a theory of a language \mathcal{L} and P a predicate symbol (of arity 1) not among the non-logical symbols of \mathcal{L} . When $\bar{x} = \langle x_1, \dots, x_n \rangle$, we write $P(\bar{x})$ for $(P(x_1) \wedge P(x_2) \wedge \dots \wedge P(x_n))$. Also, for the sake of simplicity of presentation, assume that all sentences are in some prenex normal form.

Relativisation:

- Let $\varphi(\bar{x})$ be a quantifier-free formula of \mathcal{L} . Then the relativisation $\varphi^P(\bar{x})$ of $\varphi(\bar{x})$ with respect to P is $\varphi(\bar{x})$.
- Let $\forall \bar{x} \varphi(\bar{x}, \bar{y})$ be a formula of \mathcal{L} , where the outermost quantifier of φ (if φ is not quantifier-free) is \exists .⁷ The relativisation of $\forall \bar{x} \varphi(\bar{x}, \bar{y})$ with

⁶*Op. cit.*

⁷The relativisation of a given φ is still well-defined without this qualification, by DeMorgan's Laws. But this presentation is simpler.

respect to P is the formula

$$\forall \bar{x} (P(\bar{x}) \rightarrow \varphi^P(\bar{x}, \bar{y})),$$

where $\varphi^P(\bar{x}, \bar{y})$ is the relativisation of $\varphi(\bar{x}, \bar{y})$ with respect to P .

- Let $\exists \bar{x} \varphi(\bar{x}, \bar{y})$ be a formula of \mathcal{L} , where the outermost quantifier of φ is \forall (if φ is not quantifier-free). The relativisation of $\exists \bar{x} \varphi(\bar{x}, \bar{y})$ with respect to P is the formula

$$\exists \bar{x} (P(\bar{x}) \wedge \varphi^P(\bar{x}, \bar{y})),$$

where $\varphi^P(\bar{x}, \bar{y})$ is the relativisation of $\varphi(\bar{x}, \bar{y})$ with respect to P .

- The relativisation T^P of any set of sentences T with respect to P is the set of all relativisations of the sentences in T with respect to P .

Tarski proves several convenient facts about the relationship between the original theory and its relativisation.

Proposition 1.

- (1) T^P is axiomatisable if and only if T is axiomatisable; and
- (2) if the set of non-logical symbols of \mathcal{L} is finite, T^P is finitely axiomatisable if and only if T is finitely axiomatisable.

Proof. (Tarski et al., 1953, 25ff). □

Also, T^P is interpretable in T : define the non-logical symbols of \mathcal{L}_T as themselves⁸ and add

$$\forall x (P(x) \leftrightarrow x = x).$$

Since T is also trivially interpretable in T^P , it follows that, if φ is any sentence of \mathcal{L}_T and φ^P is its relativisation to P , then

$$T \vdash \varphi \Leftrightarrow T^P \vdash \varphi^P.$$

⁸More precisely, since we replaced the non-logical symbols of \mathcal{L}_T with ‘dummy symbols’, define each in terms of its original. Cf. Tarski et al., 1953, 27-8.

Note that this does *not* mean that T^P is a conservative extension of T or vice-versa, ie, that

$$T \vdash \varphi \Leftrightarrow T^P \vdash \varphi.$$

Now we can define relative interpretability.

Relatively interpretable: O is relatively interpretable in T if, for a predicate symbol P not among the non-logical symbols of \mathcal{L}_O , O^P is interpretable in T .

Informally, the addition of the predicate P allows T to pick out a subset of its domain, and then prove the theorems of O about the items in that subset, and I therefore call P the *domain predicate*. Note that this definition requires a translation of P in T ; P cannot be taken as a new non-logical symbol. Hence, the subset of the domain of the model of T that is the domain for the model of O is definable.

**domain
predicate**

The classic post-Fregean constructions of various number systems out of set theory are paradigm examples of relative interpretability. Constructing the natural numbers out of the finite ordinals, for example, involves defining $P(x)$ as ‘ x is a finite ordinal’, defining $x = 0$ as ‘ x is the empty set’, and then defining the operations of succession, addition, and multiplication in terms of being one of the ordered pairs or triples in a set (the graph of the respective set-theoretic relation) with a long and somewhat tedious but not terribly complex definition.

When engaged in a foundationalist project, philosophers are typically interested in relative interpretability, not interpretability proper. For the sake of simplicity of presentation, however, I will generally speak of interpretability, and argue with reference to interpretability, bringing up relative interpretability only when the distinction might be relevant.

2.3. Representable. Where interpretable was a syntactic notion (albeit one with semantic consequences; see Proposition 2 below), representable is a semantic notion. We first need the notions of a satisfaction set and embedding one model in another.

Satisfaction set:

- Let R be a relation symbol of arity k in a language \mathcal{L} and \mathfrak{M} be an interpretation of \mathcal{L} with domain M . (Note that every model of a given

theory is an interpretation of the corresponding language.) Then the satisfaction set $R^{\mathfrak{M}}$ of R in \mathfrak{M} is the set

$$\{\bar{a} \in M^k : \mathfrak{M} \models R(\bar{a})\}$$

of k -tuples in M satisfying R according to \mathfrak{M} .

- Let c be a constant symbol in a language \mathcal{L} and \mathfrak{M} be an interpretation of \mathcal{L} with domain M . Then the satisfaction set $c^{\mathfrak{M}}$ of c in \mathfrak{M} is the singleton set

$$\{a \in M : \mathfrak{M} \models a = c\}.$$

- Let f be a function symbol of arity k in a language \mathcal{L} and \mathfrak{M} be an interpretation of \mathcal{L} with domain M . Then the satisfaction set $f^{\mathfrak{M}}$ of f in \mathfrak{M} is the set

$$\{(\bar{a}, b) \in M^{k+1} : \mathfrak{M} \models f(\bar{a}) = b\}$$

of $k + 1$ -tuples in M which is the graph of f in \mathfrak{M} .

- Let φ be a formula of arity $k > 0$ in a language \mathcal{L} and \mathfrak{M} be an interpretation of \mathcal{L} with domain M . Then the satisfaction set $\varphi^{\mathfrak{M}}$ of φ in \mathfrak{M} is the set

$$\{\bar{a} \in M^k : \mathfrak{M} \models \varphi(\bar{a})\}$$

of k -tuples in M satisfying φ according to \mathfrak{M} .

Embedding: An embedding of $\mathfrak{N} \models O$ in $\mathfrak{M} \models T$ is a sequence of \mathcal{L}_T -formulae

$$\langle \varphi, \varphi_1, \varphi_2, \dots, \varphi_n \rangle$$

such that

$$\mathfrak{N} \cong_{\mathcal{L}_O} \langle \varphi^{\mathfrak{M}}, \varphi_1^{\mathfrak{M}}, \dots, \varphi_n^{\mathfrak{M}} \rangle.$$

That is, an embedding is a sequence of formulae such that the satisfaction sets of those formulae in \mathfrak{M} form an interpretation of \mathcal{L}_O isomorphic to \mathfrak{N} . The satisfaction set of φ gives the domain of the new model, and the

satisfaction sets of the φ_i give the satisfaction sets of the non-logical symbols of \mathcal{L}_O .⁹

Representable: O is representable in T if, for every $\mathfrak{N} \models O$ there is $\mathfrak{M} \models T$ such that there is an embedding of \mathfrak{N} in \mathfrak{M} .

Equivalently, O is representable in T if and only if a representative of every isomorphism class of the models of O can be embedded within some model of T .

Recall, from our discussion above of complex analysis, that the real numbers are definable in the complex plane using complex conjugation:

$$R(x) =_{df} x = \bar{x}.$$

Let Ran be the theory of real analysis. The non-logical symbols of the language of Ran are two constant symbols $0, 1$ and two binary functions $+, \cdot$. The order $<$ is definable in this theory using the fact that a real number is positive if and only if square:

$$x < y =_{df} \exists z x + z \cdot z = y.$$

Ran has a countable axiomatisation, which consists of the field axioms with a countable set of axioms, one for every definable set σ , asserting that, if σ is bounded, then it has a least upper bound. Let $\mathfrak{R} \models Ran$. Define a new field \mathfrak{C} whose elements are ordered pairs $\langle x, y \rangle$ from \mathfrak{R} with the operations

$$\begin{aligned} \langle x, y \rangle + \langle w, z \rangle &= \langle x + w, y + z \rangle \\ \langle x, y \rangle \cdot \langle w, z \rangle &= \langle x \cdot w - y \cdot z, x \cdot w + y \cdot z \rangle \\ \overline{\langle x, y \rangle} &= \langle x, -y \rangle. \end{aligned}$$

Then $\mathfrak{C} \models Can$ and \mathfrak{R} has an embedding in \mathfrak{C} , using the formula $R(x)$ for the definition of the domain predicate. Hence Ran is representable in Can .

The definition of representable is a strong one: it requires not just that the satisfaction sets for the non-logical symbols be definable, but also that the domain of the resulting model be definable. This *definability condition* is crucial for the

**definability
condition**

⁹This definition could be extended to allow definitions using parameters from a subset A of M . However, this seems to amount to nothing more than moving to a new language \mathcal{L}'_T with some additional constants. For my purposes, the extension does not seem to be important.

pair of theorems that constitute the central argument of this paper. In §5, I will consider whether a weaker definability condition would be more appropriate. For now, note that the notion of embedding is significantly stronger than the standard model-theoretic notion of embedding, which does not require that the domain of the embedded model be definable, but also does not apply to structures of different languages.

Finally, we note a semantic consequence of interpretability. First, note that $\mathfrak{M} \models T'$ witnesses the fact that O is interpretable in T if and only if $\mathfrak{M} \upharpoonright_{\mathcal{L}_T} \models T$ and $\mathfrak{M} \upharpoonright_{\mathcal{L}_O} \models O$.

Proposition 2. *If O is (relatively) interpretable in T , then for all $\mathfrak{M} \models T$ there is $\mathfrak{N} \models O$ such that there is an embedding of \mathfrak{N} in \mathfrak{M} .*

Proof. Let $\mathfrak{M} \models T$. For each R a non-logical symbol of \mathcal{L}_O , set $R^{\mathfrak{M}} =_{df} \Phi_R^{\mathfrak{M}}$, where Φ_R is the interpretation of R in some fixed D an interpretation of O in T . Set

$$\mathfrak{M}' =_{df} \langle M, P_1^{\mathfrak{M}}, P_2^{\mathfrak{M}}, \dots, P_j^{\mathfrak{M}}, R_1^{\mathfrak{M}}, \dots, R_k^{\mathfrak{M}} \rangle,$$

where the P_i are the non-logical symbols of \mathcal{L}_T . Then $\mathfrak{M}' \models T'$ and so

$$\mathfrak{N} =_{df} \langle M, R_1^{\mathfrak{M}}, \dots, R_k^{\mathfrak{M}} \rangle \models O.$$

(If O is only relatively interpretable in T , replace M in the definition of \mathfrak{N} with the satisfaction set $P^{\mathfrak{M}}$ of the domain predicate P .) □

Note that this does not guarantee either that all these embedded models of O are elementarily equivalent¹⁰ or that O is representable in T .

2.4. Foundationally reducible in principle. This notion is not a mathematical one; rather, it is philosophical. It is an attempt to state one very weak and very broad motivation for any foundationalist programme.

Foundationally reducible in principle: O is foundationally reducible in principle to T if a complete account of the ontology and epistemology of O can be

¹⁰Two \mathcal{L} -structures \mathfrak{M} and \mathfrak{N} are elementarily equivalent if, for all sentences φ of \mathcal{L} , $\mathfrak{M} \models \varphi$ if and only if $\mathfrak{N} \models \varphi$. Elementary equivalence is weaker than isomorphism, even between elementarily equivalent models of the same cardinality.

based, at least in principle, on an account of the ontology and epistemology of T .

By an account of the ontology of a theory, I mean a philosophical account of the ontological or existential commitments of that theory. Thus, an account of the ontology of set theory provides us with a way of answering the questions ‘Do sets exist?’ and ‘Which sets exist?’ By an account of the epistemology of a theory, I mean a philosophical account of the warranted assertability of the claims of the theory. Thus, an account of the epistemology of set theory provides us with a way of answering the question ‘Do we know that no set can be a member of itself?’

We will first examine several features of this definition, and then turn to some more specific ways of understanding being reducible in the context of foundationalist philosophy of mathematics.

First, this is meant to be a definition of being *foundationally* reducible. Unlike other types of reductions, a foundational reduction aims at satisfying what I will call the *full coverage condition*. The full coverage condition is expressed by the use of ‘complete’ in the definition above: *all* ontological and epistemological issues regarding the object theory can be resolved by grounding it on the target theory. The foundational reduction of O to T must be, in some sense, complete and without exception. By contrast, a physicalist reduction of human minds to material entities may allow for the existence of immaterial minds – it just asserts that *our* minds are not immaterial. While this goal of full coverage is certainly related to certain notions of reduction in the philosophy of science and the philosophy of mind, it is by no means common to all notions of reduction. For the sake of presentation, I will use ‘reducible’ and ‘foundationally reducible’ as synonyms in the remainder of this paper. The full coverage condition will be scrutinised more carefully in §4.

**full coverage
condition**

Second, the definition speaks only of being reducible *in principle*. One can say that all mathematics is reducible in principle to set theory without saying that all mathematics should be or has been so reduced. One could also say that, for example, all mathematics is reducible in principle to graph theory – because all mathematics is reducible in principle to set theory, and set theory is reducible in principle to

graph theory¹¹ – without thereby saying that graph theory should be a foundations for mathematics. Furthermore, one could say that two theories, such as arithmetic and set theory, are both foundationally reducible in principle to each other. This qualification is an attempt to stress the difference between the indicative ‘is reduced’ and the hypothetical ‘can be reduced’.

So, third, ‘in principle’ carries some modal force. I will not give a deep analysis of this modality here. I will say, however, that it allows additional necessary conditions before a proposed reduction or foundations can be declared legitimate or acceptable. For example, perhaps a foundations must be ‘natural’ or ‘elegant’ or ‘beautiful’. A proposed reduction of set theory to graph theory, for example, may be ‘unnatural’, while a proposed reduction of arithmetic to set theory is ‘natural’.¹² Certain epistemological or psychological conditions might also be required, such as requiring that the target theory be somehow ‘easier’ to work with than the object theory. One might argue against the proposed reductions of elementary arithmetic to set theory by claiming that an account of our knowledge of the transfinite set-theoretic hierarchy is more difficult to give than an account of our knowledge of finite arithmetic.

The reason foundationalists attempt to give a foundations for mathematics is to answer the philosophical questions – questions about epistemology and metaphysics – we have about mathematics, or at least narrow down the range of such questions that must be answered directly. This means, fourth, that *all normal mathematics being foundationally reducible in principle to a given theory T* is a necessary, but not

¹¹Define a certain directed graph G as follows: Let the set N of nodes be the members of a given model of ZFC , and, for all $x, y \in N$, put a directed edge from x to y if and only if $x \in y$ in the model of ZFC . Then G is a directed graph that is isomorphic to the given set-theoretic model. The argument in the present paper could be easily modified to show that this construction does not actually imply that set theory is reducible to graph theory, even in principle, but it is similar to interpretation-like constructions which are often thought to show that a given theory is reducible to set theory.

¹²The debate over category-theoretic foundations is an excellent example of these sorts of considerations: both sides seem to admit some sort of in-principle reduction to category-theoretic foundations, and then argue over whether or not these reductions have other desirable – and presumably necessary – features. See Feferman, 1977, Hellman, 2003, and McLarty, 2004. Advocates of category-theoretic foundations also seem to make what I call the foundationalist’s argument. See McLarty, 1993.

sufficient, condition for T to be a foundations for all normal¹³ mathematics. Call this *universal reducibility*. A foundationalist must show that her proposed foundations has universal reducibility before her proposal can be accepted. A foundationalist programme that does not or has not yet shown that its proposed foundations has universal reducibility is in a dangerous position. Hence, since I will show that O being interpretable in T is not sufficient for O being reducible to T , it follows that universal interpretability is not sufficient for universal reducibility, and foundationalists must do more than simply appeal to universal interpretability to show that their programme satisfies this necessary condition for being a foundations for all normal mathematics.

**universal
reducibility**

It is easiest, I think, to understand what an account of the epistemology and ontology of a given theory is supposed to do in terms of the sorts of questions that motivate the search for these accounts: what are the ontological commitments of mathematics, what is the nature of mathematical entities, and how do we know about these things? We could ask such questions about all branches of mathematics: Do numbers exist? Do we have knowledge of infinite-dimensional vector spaces, and if so, how? What sort of entities are differentiable manifolds? The foundationalist project attempts to answer such questions by, first, giving a foundations for all mathematics, and then answering these questions for the foundations. A set-theoretic foundationalist, for example, claims that we can answer the above questions once we have answered the following sorts of questions: Do sets exist? Do we have knowledge of sets satisfying certain definable predicates, and if so, how? More generally, do we have knowledge of sets, and if so, how? What sort of entities are sets?

Fifth, the definition does not specify *how* the foundationalist must answer the questions concerning either the foundational or non-foundational theories. The definition is simply an attempt to capture the general idea of a reduction as that idea

¹³One of Feferman's generalisations of G2 says that, for T any consistent extension of PA , $T + \text{Con}(T)$ is not interpretable in T . Hence, assuming (perhaps unreasonably!) that it is reasonable to expect the foundations T to be an extension of PA , it would then be unreasonable to expect T to provide a foundations for its own metamathematics. The locution 'normal mathematics' is intended to grant foundationalists this exemption. In what follows, when I speak of mathematics, I typically mean only normal mathematics in this sense.

is used by philosophers of mathematics. Let us spend a few pages considering a few different types of reduction that a foundationalist might propose.

One type of reduction is an elimination: the entities of the object theory O are eliminated in favour of entities of the target theory T . There are no entities of the type described by O ; there are rather just entities of the type described by T standing in some O -like relations. For example, perhaps there are no graphs proper; there are rather just ordered pairs of the form $\langle N, S \rangle$, where S is a set of pairs on N . Hence our questions about the metaphysical status of graphs and our knowledge of graphs are literally rewritten as questions about the metaphysical status of certain kinds of sets and our knowledge of them; or, the account of the epistemology and ontology of graphs is nothing more than the account of the epistemology and ontology of sets. Reduction-as-elimination is called *eliminativism* in philosophy of mathematics as in philosophy of mind. Quine's approach to the foundations of mathematics is eliminativist. We will see this in more detail in §3.

Eliminativism is not the only type of reduction. A less ambitious type is what I will call *representationism*. On this view, entities of the type described by O are not eliminated completely, but instead represented 'paradigmatically' by entities of the type described by T . For example, one might think that the theory of differentiable manifolds DM is representable (in the sense of §2.3) by ZFC , but also that physical space is both a differentiable manifold and not a set. On this view, differentiable manifolds that are not sets do exist, and hence differentiable manifolds have not been eliminated in favour of sets. But the representability, one might think, means that one can give an account of our mathematical knowledge of differentiable manifolds in terms of our knowledge of sets. We first know about all differentiable manifolds (that is, every isomorphism class of the models of DM) by knowing about their individual representations in the set-theoretic universe V , and then conduct experiments determining, first, that physical space satisfies the axioms of DM , and, second, to which set-theoretic model it is isomorphic. Similarly, there only seem to be two sorts of models of DM : physical space, and set-theoretic models. By restricting our attention to the 'pure' or 'abstract' set-theoretic models – by ignoring the physical model – we can give an account of the ontological status of

eliminativism

**representatio-
nism**

differentiable manifolds that is based on our account of the ontological status of sets. The metaphysical status of space is a problem for metaphysicians and philosophers of physics, not the philosophers of mathematics. Not all differentiable manifolds are sets, but by appealing to representability, the philosopher's concerns about pure differentiable manifolds are addressed by addressing concerns about pure sets.

The characteristic feature of representationism is the demarcation of a subclass of the models of a theory as the 'paradigm' models of that theory – in the example above, the subclass of the abstract or set-theoretic models of DM . A representationist need not stick to this physical/abstract way of demarcating the subclass of paradigm models, however. I will allow her to make that distinction any way she likes, so long as every isomorphism class has a representative among the paradigm models¹⁴. Shapiro may be a representationist: 'It is not that one thinks of the [set-theoretic] iterative hierarchy as literally containing all structures, or all categories. Rather, we think of the iterative hierarchy as containing an isomorphism type for each structure'¹⁵.

Yet a third, still weaker, type of reduction is *relative satisfiability* or *relative consistency*.¹⁶ On this view, O is reduced to T if the satisfiability (respectively, consistency) of T is sufficient for the satisfiability (respectively, consistency) of O . This also may be a view held by Shapiro: **relative satisfiability**

The algebraic structuralist does not construct the structures of mathematics within his or her favored set theory. Set theory does not supply the ultimate subject matter for any branch of mathematics. Rather, we use set theory to establish that a given theory is coherent [roughly, satisfiable].¹⁷

While this is a reduction in some sense, I think that it is too weak to capture what the foundationalist is after. This sort of reduction would only allow us to answer a few, albeit very important, philosophical questions about O once we have answered the parallel questions about T . There are still a great many questions about O that

¹⁴This qualification is needed to satisfy the full coverage condition. See §4.

¹⁵Shapiro, 1997, 73

¹⁶I'd like to thank Chris Porter for pointing out this possibility.

¹⁷*Op. cit.*

cannot be answered even once we know that O is satisfiable. T will not give us full coverage of O .

There is a fourth type of reduction, and one which is absolutely critical for me to consider here. One can take interpretability to be itself a kind of reduction. That is, O is foundationally reducible in principle to T if and only if O is interpretable in T .¹⁸ I have two reactions to this proposal.

First, interpretability is a notion of proof theory, and being foundationally reducible is a notion of philosophy. To define one in terms of the other is, I think, to make a gross category error, and to confuse mathematical logic with metaphysics and epistemology. More moderately, if one thinks that interpretability is just a necessary and sufficient condition for being foundationally reducible, not a definition, then one appears to be making the inferential leap this paper shows is invalid. This connection needs an argument.

Second, if one simply wants to stipulate that interpretability is being reducible, thereby circumventing the need for arguments, then this paper can be taken to show that some ‘reductions’ have some deeply counter-intuitive features. They will be ‘reductions’ which do not do the sort of things foundational reductions ought to do – they do not give full coverage of the ‘reduced’ theory.

Generally speaking, we do not need to be particular about the sense in which the foundationalist is offering a reduction of mathematics to the proposed foundations. The considerations I present in §4 provide restrictions, but, as we shall see, these restrictions are motivated by the idea that a reduction should be the first step to finding a foundations, and hence be able to give complete accounts for the epistemology and ontology of mathematics. Most of the remainder of this paper is concerned with the relation between the philosophical notion of being reducible and the mathematical notions introduced above.

2.5. The foundationalist’s argument. Finally we can present the line of thought I will be criticising, and my criticism of it. We are concerned with the way a (set-theoretic) foundationalist reasons from the interpretability of any theory in ZFC to

¹⁸An equivalent proposal is to say that O is foundationally reducible in principle to T if and only if all proofs in O can be translated as proofs in T .

any theory being reducible, at least in principle, to *ZFC*. As we will see in §3, no intermediate steps are presented between these two claims. Some sort of inferential leap is being made.¹⁹ One way of making that leap is to say that interpretability logically implies being reducible in principle. Reading the foundationalist charitably, we can replace her inferential leap with an argument, where the second premiss says that interpretability implies being reducible in principle. That is, we read the foundationalist as appealing to a hidden premiss concerning the relationship between interpretability and being reducible in principle. I will call this argument *the foundationalist's argument*. For *ZFC*, it runs as follows:

**the founda-
tionalist's
argument**

- (1) Any theory O is interpretable in *ZFC*.
- (2) For any theories O, T , if O is interpretable in T then O is foundationally reducible in principle to T .
- (3) Hence any theory O is foundationally reducible in principle to *ZFC*.
- (4) Hence *ZFC* satisfies a necessary condition for being a foundations for all mathematics.

Similar arguments could be given by simply substituting one's proposed foundations T for *ZFC*.

With the definitions given above, the content of steps (1) and (3) should be clear. The move from (3) to (4) is the claim that universal reducibility is necessary for being a foundations, as in §2.4. I take the content of the conclusion to also be clear, and reiterate that being philosophical reducible in principle is only a necessary condition for a foundations, and a weak one at that. All together, the structure of the argument is to show that satisfying a certain syntactical or proof-theoretic relation is a sufficient condition for satisfying a necessary condition for being a foundations.

Since (2) is my interpolation, I focus on it. The reader may be sceptical – and this scepticism is appropriate, as my central argument in this paper is that this premiss of the argument is false. But, as we shall see in §3, at least one prominent foundationalist philosopher of mathematics has made the direct inferential leap from (1) to (3). Unless the foundationalist is appealing to some other hidden premiss, if inferring (3) from (1) is logically valid, it must be because (2) is true.

¹⁹I'd like to thank Kristin Shrader-Frechette for suggesting this way of putting the point.

My argument that (2) is false – that the direct inference from interpretability to being reducible in principle is invalid – takes the form of two claims.

Claim 1: If O is foundationally reducible in principle to T , then O is representable in T .

Claim 2: O being interpretable in T does not imply that O is representable in T .

Claim 1 asserts that representability is a necessary condition for being philosophical reducible in principle. Given Claim 1, if (2) were true, the interpretability of O in T would imply that O is representable in T . Claim 2 says that the interpretability O in T does not imply that O is representable in T . Hence (2) cannot be true. Note that by ‘imply’ I mean the ordinary relation of philosophical implication. Again, the foundationalist may be appealing to a true hidden premiss, or may fix the problem I identify by adding a true second premiss. At the very least, I argue that she cannot take interpretability *by itself* to be sufficient for being foundationally reducible, even in principle.

I will proceed with the rest of the paper as follows. I will first review Quine’s approach to the foundations of mathematics, to assure the reader that I am not tilting at windmills in attacking the foundationalist’s argument. I will then argue for my Claim 1 in two sections. I will then take one section to prove Claim 2 from Claim 1.

3. QUINE ON THE FOUNDATIONS OF MATHEMATICS

Quine scholarship is notoriously difficult. Despite his prominence and influence as an Analytic philosopher, as a writer Quine had a definite preference for interesting and amusing turns of phrase over clarity and precision, and many of his arguments are built on evocative but ultimately vague examples. In addition, Quine’s articles are often written in a dialectical (albeit not dialogical) format, rather than as a straightforward argument – Quine offers a thesis, considers a few examples, refines the thesis in light of those examples, and repeats, only stating his true view an indeterminate few pages or paragraphs before the end. It is therefore doubly difficult for the interpreter to decide conclusively whether a given sentence is Quine’s actual

view in a given article. Finally, the scope and systematicity of Quine's thought means that important remarks on, for example, the foundations of mathematics will often be found in articles that are primarily, and according to title, on, for example, philosophy of language, epistemology, or modal metaphysics.

Therefore, rather than attempting to accurately capture Quine's mature views on the foundations of mathematics by an exhaustive survey, I will focus here on two pieces Quine dates to 1964²⁰. These pieces are relatively self-contained, relatively clear, and, while missing such critical later developments as Quine's constructivism²¹, they present the notions and methods that form the backbone of Quine's foundationalism throughout his philosophical career. Examining them thereby circumvents the difficulties presented in the last paragraph without excessive distortion of Quine's views.

The first piece, 'Foundations of mathematics'²², Quine tells us was commissioned and published by *Scientific American*. He begins by motivating and illustrating the problem 'characteristic' of the search for a foundations for mathematics by talking about infinitesimals in analysis: the foundationalist wants 'to *get rid* of the infinitesimal and make do with clearer ideas while still saving the useful superstructure'²³. This elimination is accomplished by definition – in the case of infinitesimals, by the Bolzano-Cauchy definition of limits: 'This complicated fact about short but not infinitesimal times and distances can be used as a *definition* of what it means to be going a mile a minute at a given instant. The differential calculus can be reconstructed *on this basis*, and the *objectionable foundation dispensed with*'²⁴. Clearly definition, elimination – that is, eliminativism – and foundations are intertwined for Quine.

He next tells a similar story about the complex numbers. In doing so, we see more clearly the connection between definition and elimination.

²⁰Quine, 1964a/1976 and Quine, 1964b/1976

²¹Quine argued for the adoption of the 'axiom of constructibility', $V = L$, in set theory on metaphysical grounds; cf. Quine, 1992, 94. While his logic was, vehemently, classical first-order, and he was certainly no intuitionist, there is still a definite sense in which Quine was a constructivist.

²²Quine, 1964a/1976

²³*Ibid.*, 23, my emphasis

²⁴*Op. cit.*

Define the complex numbers as mere ordered pairs of real numbers, and then extend the usual algebraic operations of plus, times, and power, by definition, so as to make sense of these operations when they are applied to these ordered pairs. The definitions can be so devised as to provide us in the end with an algebra of ordered pairs of real numbers that is *formally indistinguishable* from the classical algebra of complex numbers. One tends to say not that *the complex numbers have been eliminated* in favour of ordered pairs, but that *they have been explained* as ordered pairs. *One may say either; the difference is only verbal.*²⁵

In particular, elimination is explication, is accomplished by definition, and one secures the foundations of mathematics by eliminating/explaining away suspicious, dubious, or unclear mathematical entities. Indeed, Quine almost explicitly defines the foundationalist project in this way: ‘the foundations of mathematics . . . is a process . . . of *reducing* some notions to others, and so diminishing the inventory of basic mathematical concepts’²⁶.

It is clear that Quine’s definitions are at least on the way to interpretability: if Quine defines being P as satisfying Φ , then, in the language in which we speak of both P s and things satisfying Φ , the definition is written

$$\forall x (P(x) \leftrightarrow \Phi(x)),$$

so Quine’s definitions are Tarski’s possible definitions. Since Quine is primarily or only concerned with finite languages, the requirement that the set D of possible definitions be recursive is trivially satisfied. What is missing in this article is, first, any explicit discussion of what the various theories entail or prove (Tarski’s (1) and (4)), and second, the requirement that each of the terms to be eliminated is defined only once (Tarski’s (3)).

The second seems to be taken care of by Quine’s structuralism. It is not so explicitly stated in this piece, but it is still present. ‘Any version of [natural] number

²⁵*Ibid.*, 24-5, my emphasis

²⁶*Ibid.*, 28, my emphasis

will suffice that causes the numbers to consist of a first together with the total yield of such an [successor] operator'²⁷. In particular, we can choose either Frege's or von Neumann's reduction of the natural number to sets. Neither is any better than the other for the foundationalist's purposes²⁸, and Quine does not explain how to choose one over the other. Presumably we just make a choice however we like and move on: 'Whether we take numbers in either of these ways or in some other, the next step is to define the arithmetical operations'²⁹. Relevant to my purposes here, Quine does not take both definitions at once. It's therefore reasonable to assume that Quine would accept this part of Tarski's definition.

The first missing feature, the requirement of extensionality and conservativeness, appears to be completely absent from Quine's discussion. One might reasonably assume that Quine wants truth or a certain structure to be preserved in this process of eliminativist reduction, and then argue that, since Quine's logic is classical first-order, these correspond to Tarskian model-theoretic entailment, which in turn corresponds via Completeness to proof-theoretic implication, as in Tarski's definition of interpretation. One might also point out that there isn't any other obvious alternative way of choosing between possible definitions. But these lines are more suggestive and speculative than rigorously interpretive. Still, it is at least consistent to assume Quine's reductions are captured by interpretability.

Now, does Quine accept the two premisses of the foundationalist's argument? That is, does he make the inferential leap from interpretability to being reducible? (1) is stated as clearly as one could want: 'Every sentence expressible in the notation of pure classical mathematics, whether in arithmetic or the calculus of elsewhere, can be paraphrased into this thumbnail vocabulary'³⁰ of alternative denial, universal quantification, and set membership, ie, the language of *ZFC*. (2) is the inferential leap from (1) to (3): '*Since* all mathematics can be so paraphrased, all mathematical truth can be seen as truth of set theory. Every mathematical problem can be transformed into a problem of set theory'³¹. *Since* all mathematics is interpretable

²⁷*Ibid.*, 26

²⁸*Ibid.*, 27

²⁹*Op. cit.*

³⁰*Ibid.*, 30

³¹*Ibid.*, 31, my emphasis

in set theory, *it follows that* all mathematics is foundationally reducible, at least in principle, to set theory. Interpretability implies being reducible.

So this article is almost as nice an example of the foundationalist argument as a critic could ask for. Indeed, the first version of the thesis that is the primary conclusion of this paper occurred to me while re-reading this piece. But it may not be fair to attribute to Quine a view stated in a piece meant for a general audience. Indeed, the other piece we will consider in this paper, dated to the same year, seems to reject this inference from interpretation to reduction.

In the introduction to *The ways of paradox and other essays*, Quine classifies the concerns of ‘Ontological reduction and the world of numbers’³² as ontological. More particularly, he is concerned with the idea of a reduction of one class of entities, of any sort, to another class; it just happens, conveniently for my purposes, that his paradigm examples of reductions and proposed reductions all involve mathematical entities. He presents three proposed standards of reduction, arguing that the third is the only one appropriate for metaphysical and foundationalist purposes. These are as follows:

- (a) [A reduction is given by] any effective [mapping of closed sentences on closed sentences [that] . . . preserves truth³³;
- (b) [We have a reduction if] each of the erstwhile primitive predicates of θ carry over into a predicate or open sentence about the new objects³⁴; and
- (c) We specify a function, not necessarily in the notation of θ or θ' , which admits as arguments all objects in the universe of θ and takes values in the universe of θ . This is the proxy function. Then to each n -place primitive predicate of θ , for each n , we effectively associate an open sentence of θ' in n free variables, in such a way that the predicate is fulfilled by an n -tuple of arguments of the proxy function always and only when the open sentence is fulfilled by the corresponding n -tuple of values.³⁵

³²Quine, 1964b/1976

³³*Ibid.*, 215

³⁴*Ibid.*, 216

³⁵*Ibid.*, 218

Quine quickly rejects (a) as making reduction far too cheap. Indeed, with this standard, using the Downward Löwenheim-Skolem theorem, any theory can be ‘reduced’ to the natural numbers, ‘but does this entitle us to say that it is once and for all *reducible* to that domain, in a sense that would allow us thenceforward to repudiate the old objects for all purposes and recognize just the new ones, the natural numbers?’³⁶ Quine’s answer to this rhetorical question is ‘no’.

It is clear that (b) is the method of ‘Foundations of mathematics’. Quine says that this standard has ‘narrowed . . . appreciably’ the ‘standard of ontological reduction’ compared to (a), but it is ‘not . . . very narrow’ and ‘still too liberal’³⁷. However, he never says in what way this standard has gotten, does get, or will get things wrong.

Instead, he simply moves on to (c), which strengthens (b) by requiring a ‘proxy function’ between the ‘universes’ of the theories. Clearly Quine has in mind a mapping between the domains of two models, one each of the given theories θ and θ' , corresponding to my O and T , respectively. That is, he assumes that we are given $\mathfrak{N} \models O$ and $\mathfrak{M} \models T$. Then the standard of reduction (c) is the existence of a function $f : N \rightarrow M$ and, for each n -place non-logical relation symbol R in \mathcal{L}_O , a n -place formula Φ_R in \mathcal{L}_T (suppressing the requirements of effectiveness and ignoring constant and function symbols for the sake of simplicity of presentation) such that, for all $\bar{a} \in N^n$,

$$\mathfrak{N} \models R(\bar{a}) \Leftrightarrow \mathfrak{M} \models \Phi_R(\overline{f(\bar{a})}).$$

This is clearly equivalent to having an embedding of \mathfrak{N} in \mathfrak{M} that is compatible with the interpretation D . Relative to a given interpretation D of O in T and models \mathfrak{N} and \mathfrak{M} , call this requirement *having a compatible proxy function*.

But which \mathfrak{N} and \mathfrak{M} ? Quine treats θ and θ' as though they were, not just sets of sentences, but sets of sentences that already have specific models attached. They are therefore *non-algebraic theories*, theories with a preferred or intended model. And Quine’s recurrent examples here concern reductions of and to the natural numbers, one of the standard examples of a non-algebraic theory. Quine is

³⁶*Ibid.*, 215, his emphasis

³⁷*Ibid.*, 217

apparently concerned about reductions of or to only the intended or standard model of Peano Arithmetic (up to isomorphism), not the non-standard models.³⁸

In the case where O is a non-algebraic theory, having a compatible proxy function is strictly stronger than just having an interpretation. In Proposition 2 at the end of §2.3 I showed that for every model $\mathfrak{M} \models T$ there is some model $\mathfrak{N} \models O$ such that there is an embedding of \mathfrak{N} in \mathfrak{M} in a way compatible with the interpretation. However, unless O is categorical in the cardinality of \mathfrak{M} , there is no way of guaranteeing that \mathfrak{N} is (isomorphic to) the intended model. Requiring a compatible proxy function makes the specification of the models \mathfrak{M} and \mathfrak{N} *prior* to the interpretation.

In this case, then, it appears that Quine rejects (2) of the foundationalist's argument. Though he does not explain why, it is clear that he thinks interpretability is not sufficient for being reducible.

However, suppose we now consider the case where O is an algebraic theory. In this case, O has more than one distinct and 'intended' model (up to isomorphism). The standard examples of algebraic theories are the theories of rings, groups, and graphs. Every ring, group, and graph – at least one for each cardinal, and often more than one for each cardinal – is a model of the respective theory, and these models are usually only distinguished up to isomorphism in the respective language. What are we to make of the requirement of having a proxy function in this case? In particular, from which models should there be compatible proxy functions? There are two possibilities: either all of them (that is, one representative of each isomorphism class of the models of O), or none of them.

If the latter, then the standard of being reducible, at least for algebraic theories, is just (b). If the former, then consider what is required of a reduction of group theory to set theory. For each of the distinct models of the axioms of group theory (up to isomorphism) – at least one for each cardinal, by the Upward Löwenheim-Skolem theorem – there must be a proxy function from the model to the domain V of the standard model of set theory. Then there must be an interpretation of the axioms

³⁸There is good reason for a naturalist about mathematics to think this concern is myopic: the study of non-standard models of Peano Arithmetic has generated powerful tools for proving theorems about the 'real' natural numbers. For an excellent introduction to work in this area, see Kaye, 1991.

of group theory in set theory that is compatible with this proxy function. Note that Quine requires each proxy function f be expressible (which I read as definable) ‘in the metatheory where we are explaining and justifying the discontinuance of θ in favour of θ' ’³⁹. However, if the language of the metatheory has only countably many non-logical symbols at its disposal, it cannot define all these proxy functions.⁴⁰ While Quine does not say just what the metatheory is, it is a rather implausible to think that it has uncountably many non-logical symbols – it would be no language that humans could actually learn and use. Quine has set the bar far too high: no algebraic theory with infinite models can satisfy (c). Hence, in the absence of an alternative standard, I conclude that the Quinean standard of reduction for an algebraic theory must be (b).

Yet this still means that Quine does not make the move I criticise with respect to non-algebraic theories. This may be thought a rather strong concession: perhaps my argument no longer applies to foundationalist programmes for non-algebraic theories. This is not the case. The foundationalist inferential leap that I am considering in this paper is not from interpretability to being reducible *when it comes to non-algebraic theories*. Rather, it is from interpretability *tout court* to being reducible *tout court*. Perhaps Quine is making his inferential leap using the hidden premiss that we are only interested in non-algebraic theories. Perhaps interpretability qualified in this way does imply being reducible, but this is incompatible with giving a foundations for *all* of mathematics. That is, just because the foundationalist does not make an invalid inference in the case of non-algebraic theories does not mean that her line of thought can be extended or generalised to give a more widely applicable valid argument. In particular, it appears that Quine still makes this inference in the case of algebraic theories. While he does not make the foundationalist’s argument for non-algebraic theories, his foundationalism is still subject to my criticism.

Quine is not the only prominent foundationalist who appears to make the foundationalist’s argument. The inferential leap from interpretability to being reducible

³⁹*Ibid.*, 218

⁴⁰The proxy function f is definable in the metalanguage \mathcal{L}^* if and only if there is a formula Φ_f of \mathcal{L}^* with two free variables x, y such that, for all $x \in N$ and $y \in M$, $f(x) = y$ if and only if $\Phi_f(x, y)$. If \mathcal{L}^* has only countably many non-logical symbols, then its formulae can be enumerated using a Gödel numbering. That is, there are only countably many such functions f definable in \mathcal{L}^* .

shows up in the debates over category-theoretic foundations⁴¹ and in David Lewis' attempts to give a foundations for mathematics in mereology⁴². It is also made in discussions of foundationalism by non-foundationalists.⁴³ For the sake of space, however, I will leave a detailed examination of these other instances of the argument for another paper. Quine's foundationalism is sufficiently prominent and influential among foundationalist projects in the philosophy of mathematics of the last fifty years that I take it to be a suitable target in itself.

4. FULL COVERAGE

I turn now to my argument. Recall that this argument was based on two claims. This section and the next are devoted to arguing for Claim 1, viz,

Claim 1: If O is foundationally reducible in principle to T , then O is representable in T .

I shall argue for Claim 1 by considering the following two points:

Full coverage: O is foundationally reducible in principle to T only if *every* model of O has an embedding in some model of T .

Definability: A model \mathfrak{N} of O has an embedding in some model \mathfrak{M} of T only if *all* the relevant sets in M are definable in \mathcal{L}_T .

Both of these points have appeared before in this paper. The first is the full coverage condition, first pointed out in the discussion of being reducible in §2.4. The second is the definability condition, which was first identified in the discussion of representability in §2.3.⁴⁴ I will address the second issue – the issue of the definability condition – in §5. In this section I consider the first issue – the importance of the full coverage condition to programmes in the philosophy of mathematics that aim at giving a foundationalist reduction of all normal mathematics.

First, a brief note on the logic of the argument for Claim 1. The full coverage condition takes us from being reducible to some notion of representability. The definability condition specifies this notion of representability as the one I gave above.

⁴¹Cf. the citations given in n. 12, p. 14.

⁴²Lewis, 1991, Lewis, 1993

⁴³See, for instance, Shapiro, 1991, 250 and Shapiro, 1997, 54.

⁴⁴I will take these formulations to be canonical for both conditions.

Hence the two together take us from being reducible to representable, as in Claim 1. Full coverage is therefore the linchpin in my criticism of the foundationalist's argument, as it connects the philosophical notion of being reducible to the model-theoretic notion about which theorems can be proven.

As I claimed above when I introduced the term, the full coverage condition is intended to capture a particular feature of foundational reductions. Unlike other reductions – such as the physicalist's reduction of only some minds to matter – a foundationalist's reduction of O to T aims to reduce 'everything about' O to T . In this section, I will simultaneously explicate this notion of full coverage while arguing for it by appealing to independent epistemic and ontic considerations of the purposes or aims of foundationalism in the philosophy of mathematics.

4.1. The epistemic aims of foundationalism. Of course, a complete treatment of the epistemic aims of foundationalism would require far more space than I have here. Instead, I will focus on the following *general epistemic claim*: In any epistemic foundationalist programme, when O is reduced to T , warrant for knowledge claims concerning O is logically dependent on the warrant for knowledge claims concerning T . The adjective 'epistemic' allows for the possibility of foundationalist programmes which take no interest in epistemology – a purely metaphysical programme, for example, might reject the consequent of the general claim or find it irrelevant.

**general
epistemic
claim**

Consider Frege's logicism. One of Frege's primary goals to the rejection of the Kantian claim that warrant for arithmetic knowledge depends on an extra-rational faculty of (probably temporal) intuition.⁴⁵ Hence, Frege wishes to claim instead that warrant for knowledge claims concerning arithmetic is logically dependent on the warrant for knowledge claims concerning some theory T of pure logic. Indeed, he wants the stronger claim that knowledge of arithmetic depends *only* on knowledge of T , and hence not intuition, but this certainly implies the particular instance of the general epistemic claim. A Kantian opponent could reject Frege's programme by claiming that arithmetic knowledge depends on *both* intuition and the pure logic of T , or by claiming that arithmetic knowledge does *not* depend on the pure logic of T for at least some arithmetic claims.

⁴⁵See, for example, Frege, 1884/1950, §§1-17.

Not all logicians are interested in an epistemic foundationalist programme. Russell's logicism, for example, is probably metaphysical, not epistemic, though I will not defend this claim here. Frege's, at least on standard readings, is epistemic, and is interested in defending particular instances of the general epistemic claim.

However, there is a respect in which Frege's foundationalism is incompatible with the picture I am painting of foundationalists in this paper, and I would be remiss in appealing to Frege here to support the full coverage condition and ignoring this potential difficulty. The problem is that Frege's foundationalism does not aim at universal reducibility. Frege believed that geometry could not be given a foundations in pure logic, and instead must make some essential use of spatial intuition.⁴⁶ Hence, Frege does not require that all normal mathematics be foundationally reducible in principle to his pure logic T . On the other hand, in the discussion of being reducible above, the fourth point I made was that universal reducibility to T was a necessary but not sufficient condition for T to be a foundations for all mathematics.

I do not believe this is a serious problem, either for the argument of this subsection or for my paper as a whole. First, in this subsection, I am only interested in one aspect of the full coverage condition, which is independent of considerations of universal reducibility as I have formally defined these terms. More precisely, the acceptance or rejection of the full coverage condition will change the meaning of being foundationally reducible, and hence, for example, whether a given theory actually has universal reducibility, but the formal definition of universal reducibility does not involve the full coverage condition. In this section, I am only interested in the relation of being foundationally reducible between arbitrary but particular O and T , rather than a property attributed by T by attaching a universal quantifier to this relation. In short, when Frege does make claims that some O is foundationally reducible to some T , he does make the general epistemic claim I am connecting to full coverage; it makes no difference to my argument that he does not do this for all O .

Second, I claimed above that universal reducibility is necessary but not sufficient for a given T to be a foundations for *all* normal mathematics. The fact that Frege's

⁴⁶Cf. *Ibid.*, §§13-14.

pure logic T does not have universal reducibility means that it cannot be a foundations for all mathematics. But this does not mean that it cannot be a foundations for arithmetic [*Arithmetik*], and this is all Frege is claiming. He neither wants nor expects his pure logic T to serve as a foundations for geometry, and hence does not care about universal reducibility. In the context of the foundationalist's argument, Frege rejects premiss (1), but might still accept premiss (2), and it is only the latter that is the target of this paper.

Let us now leave Frege for more general foundationalist considerations. If the foundationalist rejects the general epistemic claim, then she says that, first, her foundationalist programme aims at an account of mathematical knowledge (at least for a given O) in terms of the proposed foundations, and yet, second, denies that the proposed foundations are actually doing the work of providing warrant for that knowledge. This is absurd. If the foundations do not provide warrant for the epistemic superstructure, then what is the epistemological point to having the foundations at all? What work could the foundations possibly be doing? It is as though, in mainstream epistemology, a foundationalist were to simultaneously insist on the epistemological importance of basic beliefs and deny that they play any significant rôle in providing warrant for non-basic beliefs. On what else is the warrant for non-basic beliefs to be built?

So the foundationalist must accept the general epistemic claim. Suppose now that she rejects the full coverage condition. This would mean that knowledge of all models of O is logically dependent on knowledge of T , and yet that there are models of O which cannot be embedded in any model of T . Suppose \mathfrak{N} is one such model. Knowledge of \mathfrak{N} is supposed to depend on knowledge of T , but without a relation between \mathfrak{N} and the models of T , there is no logical relation between \mathfrak{N} and T , and hence no way for warrant for knowledge claims concerning \mathfrak{N} to be logically dependent on the warrant for knowledge claims concerning T . There is no way that the particular instance of the general epistemic claim can be maintained in this case. Hence, in the context of epistemic foundationalist programmes, O is foundationally reducible in principle to T only if *every* model of O has an embedding in some model of T .

4.2. **The ontic aims of foundationalism.** My discussion of metaphysical aspects of foundationalism will parallel my discussion of epistemological aspects. In particular, I will focus on the following *general ontic claim*: In any metaphysical foundationalist programme, when O is reduced to T , the existential commitments of O are a subclass of the existential commitments of T . The adjective ‘metaphysical’ allows for the possibility of foundationalist programmes which take no interest in metaphysics – a purely epistemological programme, for example, might reject the consequent of the general claim or find it irrelevant.

Here Quine, once again, is our paradigm foundationalist. On any eliminativist reduction of O to T , *prima facie* existential commitments to O are literally annihilated, and we are left with only existential commitments to T . Not only does Quine accept the general ontic claim, he can see no difference between a reduction and an elimination of existential commitments.⁴⁷

A non-eliminativist foundationalist might not make the general ontic claim quite so easily. A representationist, for example, would think that the reduction of O to T only involves the *paradigmatic* models of O , not *all* models of O . Here a modification of the general ontic claim might be more appropriate: when O is reduced to T , the *proper* existential commitments of O are a subclass of the existential commitments of T . The proper existential commitments of O come from the paradigmatic models of O . Recall that our example illustrating representationism dealt with differentiable manifolds: while the paradigmatic models of the theory DM of differentiable manifolds are pure sets, physical space is a differentiable manifold and yet not a pure set, and hence not among the paradigmatic models. Hence the representationist might claim that a reduction of DM to sets does not need to deal with the physical model, ie, physical space. Only the set-theoretic, paradigmatic models are important. However, the existential commitments of the paradigmatic models must still be a subclass of the existential commitments of set theory in order for DM to be reduced to set theory.

Speaking more generally, if the foundationalist rejects the general ontic claim, then she says that, first, her foundationalist programme aims at an account of the

⁴⁷Cf. the quotation at n. 25, above.

existence of mathematical entities (at least for a given O) in terms of the proposed foundations, and yet, second, denies that the existence of these entities is adequately captured by the proposed foundations. This is absurd. If the foundations do not provide grounds for the existential claims, then what is the ontological point to having the foundations at all? What work could the foundations possibly be doing? It is as though, in mainstream philosophy of mind, a physicalist were to simultaneously insist that human minds are brain states and that human minds have nothing to do with physical brains.

Any metaphysical foundationalist programme must therefore endorse the general ontic claim or a slight modification of it. Suppose now that the foundationalist rejects the full coverage condition. This would mean that the existential commitments of O are a subclass of the existential commitments of T , and yet that there are models of O which cannot be embedded in any model of T . Suppose \mathfrak{N} is one such model, and that \mathfrak{N} is a proper or pure model of O . As in the epistemic case, without a relation between \mathfrak{N} and the models of T , there is no logical relation between \mathfrak{N} and T , and hence no way for the existential commitments of \mathfrak{N} to be a subclass of the existential commitments of T . That is, the particular instance of the general ontic claim cannot be maintained in this case. Hence, in the context of metaphysical foundationalist programmes, O is foundationally reducible in principle to T only if *every* model of O has an embedding in some model of T .

If all foundationalist programmes in the philosophy of mathematics have either metaphysical or epistemic aspects, then I have now shown that the foundationalist must accept the full coverage condition. I can think of no other sort of foundationalist programme, either actual or plausible. Informally, my argument is quite simple: if the foundationalist does not satisfy the full coverage condition, then her reductions leave things out, making it impossible for her foundations to do the sorts of things that motivate the search for a foundations for mathematics in the first place.

5. DEFINABILITY

In the last section, I began my argument for Claim 1 by discussing the full coverage condition. In this section I turn to the second step in that argument, the definability condition. Recall that the definability condition requires that both the domain and the interpretations of the relations for an embedding be sets definable in the language of T . One might object that this is too strong. In this section, I will consider the two most obvious weakenings of this requirement. I can think of no other alternatives that are at least plausible.⁴⁸

5.1. Complete undefinability. Suppose neither N nor the interpretations of any of the non-logical symbols need be definable. Then representability is too easy.

Proposition 3. *Let T be any theory with infinite models and κ an infinite cardinal. Then there is a model \mathfrak{M} of T such that, for any theory O , every model of O of cardinality up to κ can be embedded (in the weak sense of complete undefinability) in \mathfrak{M} .*

Proof. For simplicity of presentation, suppose the only non-logical symbols of \mathcal{L}_O are relation symbols R_1, \dots, R_n . By the Upward Löwenheim-Skolem theorem, T has a model \mathfrak{M} of cardinality κ . Let \mathfrak{N} be a model of O with cardinality $\leq \kappa$. Then there is an injection f of the domain of \mathfrak{N} into the domain of \mathfrak{M} . Let N' be the image of f . Let $R_1^{\mathfrak{N}}, \dots, R_n^{\mathfrak{N}}$ be the satisfaction sets of the relation terms in \mathfrak{N} , as

⁴⁸A reader has suggested a third, but it is not actually a weakening. Say that O is n -weakly-represented in T if every model of O can be Π_n^0 -embedded in a model of T , ie, there is an isomorphic copy \mathfrak{N} of the model of O in the model \mathfrak{M} of T such that every Π_n^0 -sentence of \mathcal{L}_T satisfied in \mathfrak{M} is satisfied by \mathfrak{N} (as a \mathcal{L}_T -structure). However, let ZFC^+ be the extension of ZFC to a language \mathcal{L}' with a new constant for each definable set and axioms identifying each constant as satisfying one of the defining formulae. Note, first, that ZFC^+ is interpretable in ZFC and, second, that for each definable set α there is a Π_2^0 -sentence proved by ZFC^+ :

$$ZFC^+ \vdash \forall x \exists y y = \alpha.$$

Then, since PA is representable in ZFC , it is representable in ZFC^+ . If representability implied n -weak-representability for $n > 1$, then every model of PA would be isomorphic to a set-theoretic model whose domain contained every definable set. But this is impossible – the domain of the model itself must be definable, by the definability condition, but would also be an element of the model, violating regularity. Indeed, I suspect – though do not have a proof at this time – that, if O is axiomatisable with an upper bound on the complexity of its axioms, there will be some sufficiently high n such that n -weakly-representable implies representable. If this is the case, then n -weakly-representable will actually turn out to be a strengthening of representable, at least for these theories.

above. For $i = 1 \dots n$, set $R_i^{\mathfrak{M}'} = f[R_i^{\mathfrak{M}}]$, ie, use f to give interpretations for the R_i ‘in \mathfrak{M}' . By construction,

$$\mathfrak{N} \cong_{\mathcal{L}_O} \langle N', R_1^{\mathfrak{M}'}, \dots, R_n^{\mathfrak{M}'} \rangle. \quad \square$$

5.2. Domain undefinability. On this weakening, representability requires only that the interpretations of the relations be definable sets. The domain need not be definable. It is equivalent to a version of relative interpretability on which the domain predicate P is not defined, and is also one way of generalising the standard model-theoretic notion of an embedding⁴⁹. It is also very nearly the requirement Quine sets in 1964:

It is not required that ... a [proxy] function be expressible in the original theory θ [O] ..., much less that it be available in the final theory θ' [T] It is required ... of us, out in the metatheory where we are explaining and justify the discontinuance of θ in favor of θ' , that we have some means of expressing a proxy function⁵⁰

The problem I have with this lies in the overlap of proof theory and epistemology, and is best explained with some relatively concrete examples.

Say the foundationalist wishes to give PA a foundations in ZFC . One epistemological requirement for her foundations is that she be able to prove that all natural numbers are congruent to 1, 2, or 3 modulo 3. This is fairly easy to state and prove in PA :

$$PA \vdash \forall x \exists! y (x = SSS0 \cdot y \vee x = SSS0 \cdot y + S0 \vee x = SSS0 \cdot y + SS0)$$

Once the non-logical symbols of PA have been given definitions in ZFC , ZFC can state the set-theoretic translation of this sentence. But it cannot, however, prove this translation. There will be sets that are neither ‘congruent to 0 mod 3’, ‘congruent to 1 mod 3’, nor ‘congruent to 2 mod 3’. The closest ZFC can come is a conditional: if x is in N (the domain for the model of PA defined by the

⁴⁹Marker, 2002, 8

⁵⁰*Ibid.*

interpretation), then x can be so classified:

$$\forall x (x \in N \rightarrow \exists! y (x = SSS0 \cdot y \vee x = SSS0 \cdot y + S0 \vee x = SSS0 \cdot y + SS0)).$$

But this is a *ZFC* sentence if and only if N is definable, and it is only provable in *ZFC* if N is definable.

Similarly, suppose the foundationalist wants to use her foundations for *PA* in *ZFC* to disprove Goldbach's Conjecture. Let $\Phi(x)$ be the *PA* formula for ' x is even and greater than 2 and is the sum of two primes', and let $\Phi_{-GC}(x)$ be the translation of $\Phi(x)$ in *ZFC*. It is not sufficient for disproving Goldbach's Conjecture for

$$ZFC \vdash \exists x \Phi_{-GC}(x),$$

as this just says that some set is 'even' and 'greater than 2' and 'is the sum of two primes'. What she needs is

$$ZFC \vdash \exists x (x \in N \wedge \Phi_{-GC}(x)),$$

which says that some 'number' is 'even' and 'greater than 2' and 'is the sum of two primes'. But, again, this latter is a *ZFC* sentence if and only if N is definable.

Quine's undefinable domains are sufficient for his metaphysical purposes: the undesirables, whatever they are, have been identified with some subset of the acceptables, and are thus nothing more than acceptables in some particularly interesting relations. But this is not sufficient for epistemological purposes. It's not enough to know that numbers or their surrogates are out there, somewhere, among the sets. If we are to know anything substantial about them – to prove any substantial theorems of number theory using set theory – we need to be able to find them, or at least the model representing them.⁵¹

I have now finished my argument for Claim 1. In the next section, I give the proof of Claim 2 (Propositions 9 and 11) and, assuming Claim 1, use this to show

⁵¹After writing this section, I discovered Shapiro making a similar point in an argument for the importance of categoricity: 'In order to embed a structure D into a structure E , one must have a means of recognizing a substructure of E as isomorphic to D . Otherwise, one cannot be certain that D really is a substructure of E .' Shapiro, 1991, 124

that premiss (2) is false (Corollaries 10 and 12). The foundationalist’s argument as written is therefore unsound.

6. PROOF OF CLAIM 2

Let LO be the theory of linear orders. That is, LO is the theory in the language $\mathcal{L}_{<}$, whose sole non-logical symbol is the binary relation symbol $<$, given by the axioms

$$\begin{aligned} \forall x \neg x < x \\ \forall x \forall y \forall z ((x < y \wedge y < z) \rightarrow x < z) \\ \forall x \forall y (x < y \vee y < x \vee x = y) \end{aligned}$$

Let DLO be the theory of dense linear orders without endpoints. This is the extension of LO that adds the axioms

$$\begin{aligned} \forall x \forall y \exists z (x < y \rightarrow (x < z \wedge z < y)) \\ \forall x \exists y \exists z (y < x \wedge x < z) \end{aligned}$$

DLO has no finite models, and is countably categorical: all countable models are isomorphic to \mathbb{Q} with the standard order.⁵² By Vaught’s test, it is therefore complete.⁵³

Let LO_n be the theory of linear orders with n elements. This is the extension of LO that adds an axiom asserting the existence of exactly n elements. Let $LO_n \vee m$ be the theory of linear orders with n or m elements. This is the extension of LO that adds an axiom asserting that either exactly n elements exist or exactly m elements exist. For example, models of $LO_4 \vee 5$ all have either four or five elements.

Lemma 4. *LO is interpretable in DLO .*

Proof. Since DLO is an extension of LO , this is trivial. □

We need one basic result from model theory.⁵⁴

⁵²Marker, 2002, 48

⁵³*Ibid.*, 42

⁵⁴*Ibid.*, 23

Proposition 5. *Let \mathfrak{M} be a \mathcal{L} -structure (equivalently, an interpretation of \mathcal{L}). If $X \subset M^n$ is definable, then every \mathcal{L} -automorphism $f : M \rightarrow M$ of \mathfrak{M} fixes X setwise, that is, $f[X] = X$.*

Proof. By the definition of isomorphism, if $f : M \rightarrow N$ is a \mathcal{L} -isomorphism, for any \mathcal{L} -formula φ and $\bar{a} \in M^n$,

$$\mathfrak{M} \models \varphi(\bar{a}) \Leftrightarrow \mathfrak{N} \models \varphi(f(\bar{a})).$$

Let $\varphi(\bar{x})$ be the \mathcal{L} -formula defining X and f an automorphism of \mathfrak{M} . Then

$$\mathfrak{M} \models \varphi(\bar{a}) \Leftrightarrow \mathfrak{M} \models \varphi(f(\bar{a})),$$

and hence $\bar{a} \in X$ if and only if $f(\bar{a}) \in X$. □

Lemma 6. *LO is not representable in DLO.*

Proof. If LO were representable in DLO, then there would be an embedding of some model of LO n in \mathbb{Q} , that is, a $\mathcal{L}_{<}$ -definable subset N of \mathbb{Q} and formula Φ_n of $\mathcal{L}_{<}$ with two free variables such that

$$\mathfrak{M} = \langle N, \Phi_n^{\mathbb{Q}} \rangle$$

is a model of LO n . Set g to be the girth of N , that is,

$$g = \max\{|x - y| : x, y \in N\}.$$

Note that, since $|N| = n$, g is rational.

Now set $f(x) = x + g + 1$. $f : \mathbb{Q} \rightarrow \mathbb{Q}$ is a $\mathcal{L}_{<}$ -automorphism and $f[N] \cap N = \emptyset$. But since N is $\mathcal{L}_{<}$ -definable and f is a $\mathcal{L}_{<}$ -automorphism, by Proposition 5, $f[D] = D$, contradiction. □

Note that this same proof shows that LO n is not relatively interpretable in DLO: DLO cannot give a translation of the domain predicate P , since this subset (N) is not definable. This means we need another pair of lemmas.

Lemma 7. *LO4 \vee 5 is relatively interpretable in LO4.*

Proof. LO4 is an extension of LO4 \vee 5. □

Lemma 8. *$LO4 \vee 5$ is not representable in $LO4$.*

Proof. There is a five-element model \mathfrak{N} of $LO4 \vee 5$. But every model \mathfrak{M} of $LO4$ has exactly four elements. Hence there is no injection $f : N \rightarrow M$, and hence no embedding of \mathfrak{N} in \mathfrak{M} . □

These lemmas give us the following results.

Proposition 9. *Interpretability is not sufficient for representability.*

Proof. By Lemmas 4 and 6. □

Corollary 10. *Interpretability is not sufficient for being foundationally reducible in principle.*

Proof. By Proposition 9 and Claim 1. □

Proposition 11. *Relative interpretability is not sufficient for representability.*

Proof. By Lemmas 7 and 8. □

Corollary 12. *Relative interpretability is not sufficient for being foundationally reducible in principle.*

Proof. By Proposition 11 and Claim 1. □

7. CONCLUSION

In this paper, I have argued that one obvious way of understanding a certain inferential leap that is prominent throughout the philosophy of mathematics literature involves a false premiss. My argument has not shown that all of mathematics cannot be reduced to set theory, category theory, or whatever foundations one prefers. My point has been primarily a matter of logic, not content: that this argument from this premiss to this conclusion is unsound does not mean that every argument from this premiss to this conclusion is unsound.

There is therefore a fairly open-ended research programme, involving both mathematics and philosophy, to be undertaken here. If philosophers of mathematics are

still sufficiently interested in foundationalism to want a foundations for mathematics, we need to examine more closely the logical connections between the method we use to try and identify foundations – giving interpretations – and the actual reductions thereby produced. We also need to consider what additional premisses of arguments or features of proposed foundations could be added to develop logically valid arguments for claims that all of mathematics is reducible, at least in principle, to a proposed foundations.

Consider first-order set-theoretic foundations in particular. If we believe that set theory can provide a foundations for mathematics – say, number theory in particular – what is the basis for this belief? I have argued in this paper that it cannot *just* be that Peano arithmetic can be interpreted in set theory. Perhaps the fact that Peano arithmetic has an intended model, when combined with the fact that Peano arithmetic can be interpreted in set theory, does imply that set theory can be a foundations for number theory. This should be investigated both mathematically and philosophically. Furthermore, such considerations will not work for algebraic theories, such as the theory of groups. Other properties of set theory and the interpretations of group theory in set theory should be investigated here. Again, this investigation has both mathematical and philosophical components.

And this is just for set theory. Similar research into category-theoretic foundations, the most prominent contemporary rival to set theory, would also be appropriate.

A first step in this research programme would be to diagnose the failure of interpretability by itself to imply representability. More simply, how does interpretability fall short? I will close with a few preliminary remarks on this question.

Interpretability is built around the idea that the target theory should be able to prove everything the object theory proves, and in such a way that the proofs and disproofs in the object theory all correspond to proofs and disproofs in the target theory. However, interpretability is not sensitive to what sentences are *independent* of the object theory. The examples from §6 are all cases where the target theory is significantly stronger than the object theory, in that there are theorems of the target theory that are neither proved nor disproved by the object theory – for example,

with respect to $LO4$ and $LO4 \vee 5$, that the order has four elements rather than either four or five elements.

This is perhaps not a problem for non-algebraic models – the ‘real’ theory of elementary arithmetic, one might think, is the set of all sentences true of the intended model, which is complete, not the proper subset of this that is the set of implications of the axioms of Peano Arithmetic. But the strength and utility of algebraic theories is their ability to characterise a wide variety of mathematical entities using just a short list of axioms – for example, the way the three axioms of group theory characterise all groups *qua* groups, or the way the axioms of linear orders identify what is common to all linear orders as such. The sentences that are neither proved nor disproved by the axioms of group theory – such as a sentence asserting that the group is Abelian – are just as significant to the theory as the implications of the axioms. And interpretability is not sensitive to this fact. It is not sensitive to the significance of logical independence and having a variety of ‘intended’ or ‘standard’ models.

One avenue of research, then, is to consider ways of expressing logical independence and the class of ‘standard’ models of a given theory using the equipment of formal languages and model theory. Perhaps these considerations would lead to formal methods that in turn would lead us to representability.

REFERENCES

- Feferman, S. (1977). Categorical foundations and foundations of category theory. In R. Butts & J. Hintikka (Eds.), *Logic, foundations of mathematics, and computability* (p. 149-72). Dordrecht, Reidel.
- Frege, G. (1950). *The foundations of arithmetic* (J. Austin, Trans.). Northwestern. (Original work published 1884)
- Hellman, G. (2003). Does category theory provide a framework for mathematical structuralism? *Philosophia mathematica* (3), 11, 129-57.
- Hilbert, D. (1950). *The foundations of geometry* (E. J. Townsend, Trans.). Open Court. (Original work published 1899)
- Kaye, R. (1991). *Models of Peano arithmetic*. Oxford.

- Lewis, D. (1991). *Parts of classes*. Blackwell.
- Lewis, D. (1993). Mathematics is megethology. *Philosophia mathematica*, 1, 3-23.
- Marker, D. (2002). *Model theory: An introduction*. Springer.
- McLarty, C. (1993, December). Numbers can be just what they have to. *Noûs*, 27(4), 487-98.
- McLarty, C. (2004). Exploring categorical structuralism. *Philosophia mathematica* (3), 12, 37-53.
- Quine, W. V. O. (1976a). Foundations of mathematics. In *The ways of paradox and other essays* (Revised and enlarged ed., p. 22-32). Harvard. (Original work published 1964a)
- Quine, W. V. O. (1976b). Ontological reduction and the world of numbers. In *The ways of paradox and other essays* (Revised and enlarged ed., p. 212-20). Harvard. (Original work published 1964b)
- Quine, W. V. O. (1992). *Pursuit of truth* (Revised ed.). Harvard.
- Shapiro, S. (1991). *Foundations without foundationalism: A case for second-order logic*. Clarendon.
- Shapiro, S. (1997). *Philosophy of mathematics: Structure and ontology*. Oxford.
- Tarski, A., Mostowski, A., & Robinson, R. (1953). *Undecidable theories*. North-Holland.