THE REVIEW OF SYMBOLIC LOGIC
Volume 6, Number 3, September 2013

WHAT CAN A CATEGORICITY THEOREM TELL US?

TOBY MEADOWS
School of Mathematics, University of Bristol

Abstract. The purpose of this paper is to investigate categoricity arguments conducted in second order logic and the philosophical conclusions that can be drawn from them. We provide a way of seeing this result, so to speak, through a first order lens divested of its second order garb. Our purpose is to draw into sharper relief exactly what is involved in this kind of categoricity proof and to highlight the fact that we should be reserved before drawing powerful philosophical conclusions from it.

§1. Introduction. A theory is *categorical* if any two structures which satisfy it are isomorphic. It is a property of theories and it is clearly desirable. With a categorical theory, there is a sense in which we can treat any of the structures satisfying it as equivalent to any other. Any structural property we care to consider, so long as it depends only on the relations, functions and constants of a structure's underlying language, is preserved by isomorphism; and anything else is a mere artefact of the frame in which that structure is set. Thus we say that any two isomorphic structures are identical up to isomorphism and it is in this sense categoricity gives us a kind of uniqueness result. It tells us that for all intensive purposes, our theory picks out a unique structure.[1]

This property is most important in those cases where we have an antecedent belief or reason to believe that a theory is about some particular structure and no other. This is not the situation in abstract algebra. For example, we do not expect that all the structures satisfying the theory of groups are isomorphic. Indeed, the strength of this kind of theory is its generality: its ability to be instantiated by a multitude of structures. On the other hand, we might attempt to characterise a particular group, say the group consisting of the first four numbers closed under addition modulo four. In this case, it is possible for us to pick out just those structures which are isomorphic to our intended structure. We can write out a simple first order theory which does this and in proving that this theory is categorical, we show that our axiomatic enterprise has been successful: we have isolated our intended structure.

Classical model theory has shown us that examples like this are as far as we can go. For any finite structure, its complete theory will only be satisfied by structures which are all isomorphic to each other. However, the Löwenheim-Skolem theorem shows us that any theory satisfied in an infinite model will have models of every cardinality larger than the theory itself. Thus categoricity is lost. Classical model theory side-steps this problem by investigating theories which are categorical in particular cardinalities. For example, Morley's celebrated categoricity theorem shows that if a theory is categorical in one

Received: February 25, 2013.

[1] Indeed for the remainder of the paper, we shall remain somewhat relaxed about the distinction between uniqueness *simpliciter* and uniqueness up to isomorphism.

doi:10.1017/S1755020313000178

uncountable cardinal, then it is categorical in every uncountable cardinal (Chang & Keisler, 1973). This result forms the foundation for the fields of stability theory and classification theory.

However, these results are apt to appear disappointing given the kinds of theories we really expect to have a unique referent. Theories like the theory of pure identity and the theory of dense linear order are categorical when instantiated by countable structures. This is congenial. However, it can be shown that the theory of true arithmetic will be satisfied by continuum many pairwise nonisomorphic countable structures (Kaye, 1991). This is particularly disappointing since there appears to be an almost universal belief amongst mathematicians and philosophers that the language and practice of arithmetic does refer to a unique structure. Moreover, similar failures of categoricity occur with respect to analysis and set theory. A common response to this is that the model-theoretic results are merely a reflection of the inherent weakness of first order logic. But inherent weakness in comparison to what? Here the playing field widens significantly, but the main two threads are higher order logic and infinitary logic. In this paper, we shall focus on the former and pay particular attention to second order logic.

In the nineteenth century, Dedekind (1963) showed, using what we would now call second order logic, that the theory of (second order) arithmetic is indeed categorical, and similarly for the theory of analysis. We shall look more closely at this result in the next sections, but for the moment we make some preliminary remarks. First, the result seems to be exactly what we want. We appear to have shown that our ordinary practice does indeed line up with a unique structure up to isomorphism. Moreover, the second order theory of arithmetic appears to give us the means to do this. A succinct statement of the usual understanding of this result is provided by Mayberry (2000):

> Why has the standard definition of real number become standard? There is a straightforward logical reason; we can write down a set of axioms, the axioms for a complete ordered field, that, on the one hand, strike us as truisms about the real numbers, and, on the other, are *categorical* in the sense that all structures satisfying the axioms are mutually isomorphic. This means that for any proposition in the standard theory of real numbers, either it or its negation is a logical consequence of the axioms, that is to say, the axioms are *complete*. Of course, these axioms are of second order, so this logical completeness doesn't translate into a complete system of formal proofs which would allow us to prove all truths and refute all falsehoods. But the categoricity does guarantee that the basic notions of real analysis are well defined and its propositions have definite meanings. Real analysis pursued in this standard way is thus a definite, and determinate, particular theory.

In this paper, I would like to examine what is at stake in these claims and what philosophical fruit can be gathered from them. I suggest that our philosophical goals with regard to categoricity proofs are often imprecisely formulated and that this frequently leads us to overstate the value of these proofs in the philosophy of mathematics. Furthermore, this overstatement may serve to obscure a more restrained and accurate understanding of the significance of these results.

The paper is divided into the following sections:

- What do we want from categoricity proofs?
- How do categoricity proofs work?

- How do the proofs match up with our goals?
- Response to an objection and Conclusion.

§2. What do we want from categoricity arguments? We now introduce three basic goals we might have for the categoricity theorem. I do not want to assert that this list is exhaustive, but it should provide something of a foundation from which we may assess the value of these theorems. The goals are as follows:

(1) to demonstrate that there is a unique structure which corresponds to some mathematical intuition or practice;
(2) to demonstrate that a theory picks out a unique structure; and
(3) to classify different types of theory.

We now provide some preliminary exposition of what is involved in each goal. However, we shall be a better position to examine these issues in subsequent sections.

2.1. Demonstrate that there is a unique structure corresponding to some intuition or practice. Clearly, our mathematical practice is littered with a multitude of different structures. But when we talk about arithmetic, analysis or simple type theory we have an expectation that we are talking about one thing rather than a multitude of them. Of course, our philosophical and metaphysical commitments will vary from mathematician to mathematician and consequently this idea could mean many different things. Nonetheless, most of us would want to know that, at least, we are referring to a unique structure up to isomorphism. At first blush, a categoricity theorem gives us a means of achieving this. For example, when we produce a categoricity proof for arithmetic using second order logic, it seems that we have indeed shown that every model of arithmetic is isomorphic to every other.

We might see the argument for the second order decidability of the continuum hypothesis of Kreisel (1969) as working along these lines. Appropriating the precis from Weston (1976), we might see Kreisel as using the proof of the categoricity of analysis, as conducted in second order logic, as establishing that there is only one cardinality of the continuum. Given this, it is then argued that the continuum hypothesis has a definite answer. Perhaps the most recent and prominent employer of this kind of argument is Martin (2001).

2.2. Demonstrate that some theory picks out a unique structure. On the other hand we might not want to know that such a unique structure is out there corresponding to our practice. Perhaps we already have other reasons to believe this and as such, we may be more interested in knowing whether some theory is able to pick out this structure. For example, suppose we have an antecedent belief that there is only one correct model for arithmetic. Then the proof that second order Peano arithmetic is categorical is valuable, since it demonstrates that our axiomatic enterprise has been successful. We might see the categoricity theorems of Dedekind (1963) as proceeding in this fashion; thus showing that his axiomatisation of arithmetic successfully picks out the structure he was aiming to describe.

We also consider a more modest position along these lines. Perhaps our philosophical goal is not to find unique structures corresponding to our mathematical practice, but rather to show that our mathematical practice is coherent in the sense that every sentence used in, say, the practice of arithmetic has a determinate truth value. Such a position is sometimes known as semantic, as opposed to ontological realism (Shapiro, 1997). The second order categoricity proof would be helpful here in that by establishing isomorphism between models, we trivially get elementary equivalence and thus a complete theory. A contemporary example of such a project can be found in McGee (1997).

2.3. Classifying different types of theory. Finally, consider a distinction between theories which we intend to have unique models (up to isomorphism) and those which we do not. For example, we do not expect a theory of rings to pick out a unique structure, but we do expect a theory of arithmetic to do this. Shapiro (1997) has cast this as a distinction between algebraic and nonalgebraic theories and we shall find this terminology useful. A similar distinction is also made in Feferman (1999). A categoricity proof might then be used as a tool to distinguish between these types of theories. By providing an appropriate second order axiomatisation of a theory we might, for example, be able to classify algebraic and nonalgebraic theories on the basis of whether one can produce a successful categoricity proof for them.

§3. How do categoricity arguments work? Our goal in this section is to investigate what is required to mount a categoricity proof. For the purposes of simplicity of notation and theory we shall use arithmetic as the main target of investigation, although the results can be easily generalised to other nonalgebraic theories. While it is obvious that the theorems can be conducted in ZFC, the following discussion is intended demonstrate that something is lost in making this move too quickly. The goal of this paper is to highlight the categoricity theorem itself and I would like to suggest that moving to ZFC actually distracts from this purpose. Consequently, I have tried to avoid, or at least minimise the impact of, issues pertaining to set theory here.[2] I would also like to stress that the results and subsequent discussion should not be construed as an argument for skepticism as to the existence of a unique model of arithmetic. For example, an argument along the lines of Halbach & Horsten's (2005) could be used to assuage these doubts. Rather, our effort here is to draw into question the epistemic value, or leverage, that can be drawn from the categoricity proof as conducted in second order logic. We propose to attempt some measure of this on the basis of the ontological and theoretical outlay required for the proof.

The section is broken into two parts. First we shall investigate a way of conducting the categoricity proof in first order logic rather than second order logic. This is not the usual approach. However, in making the shift to this perspective, we shall be able to see more clearly what is actually required. We shall then show how to run the proof using second order logic and the comparison between the approaches will fuel the discussion for the remainder of the paper.

3.1. A first order perspective. The theorem is going to be about the first order theory of Peano arithmetic (PA) and we shall call this the *target theory.* To be as clear as possible we are going to make our metatheory explicit. We shall use the first order theory ACA_0. We shall adopt a multi-sorted articulation of this theory as developed in Simpson (1999). It is an expansion of both the language and theory of PA. We expand the language by adding variables $X, Y, \ldots$ for the second sort of object, and by adding the binary relation $\in$ which can only stand between objects of the first and second sort. Thus $x \in X$ is well

[2] Of course, this simplifying assumption comes at a cost in that many of the issues raised here strike to the heart of the philosophical-mathematical concepts of structure and set. I aim to deal more thoroughly with these points in a sequel to this paper.

https://doi.org/10.1017/S1755020313000178 Published online by Cambridge University Press

formed while $x \in x$ and $X \in X$ are not. We then expand the theory with the addition of the following axiom (and scheme):

- (Comprehension) $\forall Y \forall y \exists X \forall x(x \in X \leftrightarrow \varphi(x, Y, y))$, where φ is an arithmetic formula (i.e. one not involving quantification over the second sort of variable).
- (Induction) $\forall X(0 \in X \wedge \forall x(x \in X \rightarrow sx \in X) \rightarrow \forall y(y \in X))$.

We shall call this the *frame theory*. Clearly, there is a relationship between this theory and second order logic. One natural interpretation of the second sort of variable ($X, Y, \ldots$) is as collections of natural numbers. Although this gives the theory a second order flavour, it should not distract us from the fact that ACA_0 is still a first order theory and thus subject to the ensuing model theoretic peculiarities.

Heuristically, if we were thinking in terms of a graph theoretic representation of a model of this theory, then we should think of both sorts of objects as being represented by vertices. Then the second sort of objects could have arrows pointing to collections of objects of the first sort. More specifically, if for some a, A in a structure, we have $a \in A$, then there would be an arrow pointing from A to a. Of course, nothing really hangs on this explanation. I just want to emphasise that ACA_0 is a first order theory and collections present only one way of interpreting the second sort of objects. Having made this point, we shall now refer to the variables of the first sort as *number variables* and variables of the second sort as *class variables*.

3.1.1. Satisfaction. In order to articulate the categoricity theorem, we need to be able to talk about theories and in particular say when a theory is true in a model. For this, we need to be able to represent models and to define a satisfaction relation. Using techniques similar to Simpson (1999), we can code models of Peano arithmetic using classes.[3]

First we assume that we have a recursive function which codes up finite sequences: $\langle \cdot \rangle : \omega^{<\omega} \rightarrow \omega$ such that $n_i < \langle n_0, \ldots, n_m \rangle$ for all $i \leq m$; and we assume that we have a recursive coding function for sentences: $\ulcorner \cdot \urcorner : Sent \rightarrow \omega$.

A class $X = \mathcal{A}$ will represent a structure of of the form $(A, 0^{\mathcal{A}}, s^{\mathcal{A}}, +^{\mathcal{A}}, \times^{\mathcal{A}})$ where A is the domain of the model and the components form the signature of the structure (see Marker, 2002). We code the different parts of the structure by tagging the elements of the class with flags. For example, we might define the domain A as follows:

$$n \in A \leftrightarrow \exists m \in \mathcal{A}\ (m = \langle 5, n \rangle).$$

Thus elements of the domain are tagged with the flag '5'. We shall denote those classes which represent structures by $\mathcal{A}, \mathcal{B}, \ldots$ and we call these *class models*.

We now define a satisfaction relation $\models \subseteq \mathcal{P}(\omega) \times \omega$ between models (as classes) and sentences (as code numbers) using our multi-sorted ACA_0.

DEFINITION 3.1. *A class X is* s-closed for $\mathcal{A}$ *if for every n for all φ, ψ*

- $Atom(n) \wedge T(n, \mathcal{A}) \rightarrow n \in X$;
- $n = \ulcorner \neg\varphi \urcorner \rightarrow (\ulcorner \varphi \urcorner \in X \leftrightarrow n \notin X)$;
- $n = \ulcorner \varphi \wedge \psi \urcorner \rightarrow (\ulcorner \varphi \urcorner \in X \wedge \ulcorner \psi \urcorner \in X \leftrightarrow n \in X)$; *and*
- $n = \ulcorner \forall x \varphi(x) \urcorner \rightarrow (\forall m \in A \ulcorner \varphi(\underline{m}) \urcorner \in X \leftrightarrow n \in X)$.

[3] Our goal here is not so much to provide a self-contained set of definitions but rather: to convince the reader that this is easily achievable; and to refer to appropriate sources for further information.

https://doi.org/10.1017/S1755020313000178 Published online by Cambridge University Press

$Atom(\cdot)$ is a recursive predicate which says of some n whether or not it is the code of a sentence. $T(\cdot,\cdot)$ is a recursive predicate which says of some number n whether or not it is is the code of a true atomic sentence.

DEFINITION 3.2. *φ is true in model $\mathcal{A}$, abbreviated $\mathcal{A} \models \varphi$ if for every X such that X is s-closed for $\mathcal{A}$, $\ulcorner \varphi \urcorner \in X$.*

In order to state the theorem, we also need the following property for models of PA.

DEFINITION 3.3. *$\mathcal{A}$ is* well founded *iff for any X such that $X \cap A \neq \emptyset$ there exists an $<_{\mathcal{A}}$ −least element of $X \cap A$.*

The reader will note that this condition on a model has the same effect as saying that the second order induction axiom is satisfied in a standard model. We shall return to this point in our discussion of the second order perspective.[4]

3.1.2. Categoricity. We are now ready to state and prove the theorem.

THEOREM 3.4. *(ACA_0) For any two well founded models $\mathcal{A}$, $\mathcal{B}$ of PA, a function f exists such that $f : \mathcal{A} \cong \mathcal{B}$.*

Proof. (Sketch) Suppose we have two models $\mathcal{A}$ and $\mathcal{B}$ such that $\mathcal{A} \models PA$ and $\mathcal{B} \models PA$. Using a β-function[5] B, we define the graph F of a function $f : A \to B$ as follows:

$$\begin{aligned} \langle a, b\rangle \in F \quad \leftrightarrow \quad & \exists k(B(0^{\mathcal{A}}, 0^{\mathcal{B}}, k) \wedge \\ & \forall m < k \forall n < k(B(m,n,k) \leftrightarrow B(s^{\mathcal{A}}m, s^{\mathcal{B}}n, k)) \wedge \\ & B(a,b,k)) \end{aligned}$$

By arithmetic comprehension, this relation exists as a class of pairs. We claim F is an isomorphism between $\mathcal{A}$ and $\mathcal{B}$. It suffices to show that:

[4] We should note that the use of class models is a little eccentric. Traditional model theory makes use of the full resources of set theory and thus admits models of arbitrary cardinality. However, on any natural understanding, a class model must be countable. Thus we are limited to providing a proof of countable categoricity, which is a weaker result. For example, as noted above, the first order theory of dense linear order is countably categorical but also has models of every larger cardinality. Our main reason for taking this approach is methodological: in order to get some idea of the value of the proof, we want to minimise the ontological outlay that goes into demonstrating it. Regardless of our comfort with traditional resources, we claim that it is still interesting to have some idea of the requirements for the proof. However, if from the beginning, we avail ourselves of quantification over every model of any cardinality, then this particular measure will not work. In a nutshell, if we want to take into account weak ontological requirements, then we can only use a weaker version of the theorem.
With regard to the weakness of countable categoricity, we make two further remarks. First, countable categoricity is sufficient for the purposes of this paper. While full categoricity is the ultimate goal, countable categoricity is a necessary part of that. Thus, issues in obtaining countable categoricity those issues will persist into the quest for full categoricity. Second, we observe that if we were to take up a more expressive frame theory, which was capable of talking about larger – perhaps arbitrarily large – models, then full categoricity with respect to that frame will simply follow. Our goal here, however, is to merely use arithmetic as a case study for this technique. We want to provide an illustration of the method with a view to providing a clear analysis of its requirements.

[5] $B(x,y,z)$ is a β function if for any finite sequence $S = (a_1, b_1), \ldots, (a_n, b_n)$ there is some $k \geq a_1, \ldots, a_n, b_1, \ldots, b_n$ such that $(a,b) \in S$ iff $B(a,b,k)$ (Kaye, 1991).

https://doi.org/10.1017/S1755020313000178 Published online by Cambridge University Press

(1) $dom(F) = A$;
(2) $ran(F) = B$;
(3) $\forall x(\exists y\langle x, y\rangle \in F \rightarrow \forall z(\langle x, z\rangle \in F \rightarrow z = y))$; and
(4) $\forall x(\exists y\langle y, x\rangle \in F \rightarrow \forall z(\langle z, x\rangle \in F \rightarrow z = y))$.
(5) $+$ and $\times$ are preserved by F.

(1) Suppose the domain of f is a (nonempty) proper subset of A. By arithmetic comprehension, the set:

$$H = \{x | \neg\exists y \ \langle x, y\rangle \in F\}.$$

exists and is nonempty. Since $\mathcal{A}$ is well founded H has a least element h such that $h = s^{\mathcal{A}}h^*$ for some h^*. Since $s^{\mathcal{A}}h^*$ is the least such element there must be some $b \in B$ such that $\langle h^*, b\rangle \in F$. But since in our definition of f, B is a β function there will be some k which represents the sequence ending in ...(h^*, b), $(h, s^{\mathcal{B}}b)$: contradiction. The other cases are similar. □

REMARK. *In Shapiro's (1991) reconstruction of Dedekind's argument, the bijection is defined by means of a union, or closure. Thus the argument of the existence of the bijection uses* Π^1_1 *comprehension. If we had adopted this approach, then we would not have been able to prove the theorem in* ACA_0*. We also note that variations on the categoricity theorem here can be proven in even weaker theories, although the proof is more technical (Simpson & Yokoyama, 2012).*

The upshot of all this is that we have shown in the multi-sorted theory ACA_0 that any two well founded models of arithmetic are isomorphic. We seem to have proved that there is a unique model of arithmetic up to isomorphism and further that we have an axiomatic device which achieves this.

3.1.3. The problem. But of course there is a well-known catch. The proof was conducted using a first order theory and as such, ACA_0 has nonstandard models. For example, it could contain infinite numbers or it might have a merely denumerable collection of classes. Thus, who is to say whether our metatheory picks out a particular structure or interpretation. At this point, it is natural to ask why this result is interesting. Why should we care that according to some nonstandard model of arithmetic every pair of inner models $\mathcal{A}$ and $\mathcal{B}$ is isomorphic? It might seem that our success is merely apparent. We return to this issue in the next section.

3.1.4. An unpalatable solution. It is instructive to investigate a pathological response to the problem. First, observe that the lack of guarantee of a standard model for our frame theory could draw into question the value of categoricity proofs conducted there. But if we had some means of showing that our frame theory was categorical, then we would have a solution. To this end, we might make our multi-sorted theory ACA_0 the target theory and search for a suitable frame theory. A natural enough candidate is the theory of multi-sorted order arithmetic in which we have a third sort of variable $\mathfrak{X}, \mathfrak{Y}$, ... to represent families of classes in addition to classes of numbers and numbers. We shall call this new sort of variable a *family variable*. With regard to the $\in$ relation, we shall only allow formulae of the form $x \in X$, $X \in \mathfrak{X}$. We then add the following axioms:

- (Comprehension) $\forall\mathfrak{Y}\forall Y\forall y\exists\mathfrak{X}\forall Z(Z \in \mathfrak{X} \leftrightarrow \varphi(Z, \mathfrak{Y}, Y, y))$ where φ involves no quantification over *family* variables.
- (Extensionality) $\forall Y\forall Z(\forall x(x \in Y \leftrightarrow x \in Z) \rightarrow Y = Z)$.

https://doi.org/10.1017/S1755020313000178 Published online by Cambridge University Press

For want of a name, we shall call this theory $\Delta^2_0 - CA_0$. Using analogous techniques to the previous case, we then use families to code models of ACA_0. Thus a model will be a family $\mathfrak{X} = \mathcal{M}$ be a tuple of the form $(M^1, M^2, 0^{\mathcal{M}}, s^{\mathcal{M}}, +^{\mathcal{M}}, \times^{\mathcal{M}}, \in^{\mathcal{M}})$ where:

- M^1 is the *number* domain;
- M^2 is the *class* domain;

We shall denote models of this kind by $\mathcal{M}, \mathcal{N}$ and call them *family models*.

Note that we need to move into third order arithmetic to be able to naturally represent these models. By Cantor's theorem there must be nondenumerably many classes. This means there is no way of representing all of the classes using a single class modified with labeling tricks. Of course, whatever theory we adopt, there will be a countable model and thus we could get away with using a class (of naturals) but this would be in poor faith with our goals. Analogously to the previous case, we define a satisfaction predicate for the language of ACA_0. And finally we make the following definition, which is needed to secure the result. We make this definition from the point of view of the frame theory $\Delta^2_0 - CA_0$.

DEFINITION 3.5. *($\Delta^2_0 - CA_0$) A family model $\mathcal{M}$ (of the target theory ACA_0) is* full *if for every subclass X of the number domain of M^1, there is some C in M^2 such that*

$$\forall z(z \in X \leftrightarrow z \in C).$$

In other words, the family model $\mathcal{M}$ is closed under subsets. We are ensuring that from the perspective of our frame theory $\mathcal{M}$ does not leave any classes out.

THEOREM 3.6. *Suppose $\mathcal{M}$ and $\mathcal{N}$ are full, well-founded models of ACA_0, then $\mathcal{M} \cong \mathcal{N}$.*

Proof. (Sketch only) Assume that we have used the argument of the previous proof to obtain a bijection on the number domains. It suffices to show that there is a bijection between the class domains and that it is structure preserving for the $\in$ relations. We let $\mathfrak{G}$ be relation between the family models by the family $\mathfrak{G}$ be such that:

$$\text{for all } C \in M^2 \text{ and } D \in M^2, (C, D) \in \mathfrak{G} \text{ if } \forall x(x \in C \leftrightarrow f(x) \in D).$$

We then use fullness of the family models to show that $\mathfrak{G}$ is a surjection and extensionality to show injectivity. □

Thus from the perspective of $\Delta^2_0 - CA_0$, any two full, well founded models of ACA_0 are isomorphic. It could seem like this meets our philosophical goals, but of course, whatever problems we had before, emerge once again. $\Delta^2_0 - CA_0$ is a first order theory and it also has nonstandard models.

Continuing in this vein, we could keep carrying out similar categoricity proofs for an n-sorted first order arithmetic and further into the theory of types and beyond.[6] Along these lines, we state the following result without proof. We shall call a ramified hierarchy Γ_α (for some ordinal α) the theory formed by forming an α-sorted theory where each successor level enjoys predicative comprehension over the previous levels and at limit stages we take the union of the previous theories.

[6] A similar regress is described informally by Shapiro (1990) as a kind of game between proponents of first order and second order logic.

https://doi.org/10.1017/S1755020313000178 Published online by Cambridge University Press

THEOREM 3.7. *Let Γ_α be the theory α^{th} ramified hierarchy over PA. Then Γ_α is categorical from the perspective of $\Gamma_{\alpha+1}$. In other words, from $\Gamma_{\alpha+1}$'s perspective it is the case that any two full and well-founded models of Γ_α are isomorphic.*

The proof is just an extension of the previous two, except that we exploit the well-foundedness of the type hierarchy in order to ensure via induction that each level of the hierarchy is isomorphic. Of course, we never reach a point at which we arrive at a theory which we can show is *absolutely* categorical. Indeed, the regress seems to have quite the opposite effect.

3.2. A second-order perspective. In this section, we shall see that second order logic presents an interesting and arguably natural way around the problem illustrated above. The version of the categoricity theorem presented below is slightly different from the canonical version (of say Shapiro's 1991) in that we once again pay particular attention to the metatheoretic requirements of the proof. This time our target theory is PA^2, which is a fully fledged second order theory articulated in second order logic. We introduce a new type of variable $X, Y, ...$ which are intended to range over arbitrary subsets of finite cartesian products of the natural numbers. Since we are using second order logic, we do not use a membership relation, but rather write Xx to indicate that, after interpretation, x is in the extension of X. We call the expanded language $\mathcal{L}^2$. The logical axioms and rules are then extended to accommodate the second order quantifiers and the following logical axiom schema is introduced:

$$\forall Y \forall y \exists X \forall z (Xz \leftrightarrow \varphi(z, y, Y))$$

where φ is an arbitrary formula of $\mathcal{L}^2$. As with ACA_0, this schema is known as comprehension and allows us to prove the existence of classes. However, in distinction from ACA_0, the comprehension schema admits formulae which quantify over sets and not just numbers. To form the theory PA^2, we add the usual Peano axioms and replace the induction schema with the following induction axiom:

$$\forall X (X0 \wedge \forall x (Xx \rightarrow Xsx) \rightarrow \forall y Xy).$$

Clearly, the resultant theory bears a marked resemblance to the theory ACA_0. Indeed, modulo some tractable difficulties around tuples, the formulae of one theory can be obtained from the other by either adding the membership relation or removing it.[7] In this sense, the distinction appears to be merely cosmetic. To motivate the difference, we might appeal to the graph theoretic illustration we used with regard to ACA_0 above. We observed there that since we were just adding a new sort of object to the domain, a natural way to represent this was by means of more vertices. Then the membership relation was represented by arrows between number vertices and class vertices. However, in this case we should not think of ourselves as adding a new sort of object. Rather, the new variables should be understood to range over all the collections of objects already given in the domain. Thus we might think of them as being represented as loops around collections of vertices instead.

7 After such a translation, we can see that the more inclusive comprehension scheme makes PA^2 a more powerful theory than ACA_0. For example, PA^2 can be used to prove the consistency of PA, but ACA_0 cannot. This, however, is not a deep source of difference between second order theories and multi-sorted (first order) theories. For example, we could easily add a comparable comprehension axiom scheme to our multi-sorted theory to get $\Pi^1_\infty - CA_0$. Proof theoretically speaking, this is just as powerful as PA^2 but is still, however, a first order theory.

But of course, this is just further illustration. The real distinction between the second and first order theories is a matter of *intention*. With PA^2, the only intended model is the one in which we quantify over all of the subsets of the domain, whereas with ACA_0, we are often happy to work with far less.

As in the previous section, we now need to decide on a metalanguage from which to articulate and then prove the theorem. Our first problem is to represent the models of our theory PA^2. Since this is a second order theory, we might suppose that we need some way of representing both the numbers and classes of a model of PA^2. But in fact this is not the case. As suggested by the illustration above, in representing a second order model, we should not think of ourselves as adding a new sort of object to the domain. The domain itself is sufficient to determine the range of the second order quantifiers: we just take the powersets of each of the finite products of the domain. This means that when we represent a model of PA^2, we only need to take care of the numbers as the classes will take care of themselves. Thus, using the coding techniques of the previous proof, we can use a collection of natural numbers to represent an arbitrary model of PA^2. Again, we shall use $\mathcal{A}$, $\mathcal{B}$ to represent these models.[8]

The next thing we need is some way of saying that some class $\mathcal{A}$ is a model of PA^2. We have a couple of options here. First, we could move into some version of third order arithmetic and define the required satisfaction predicate there. If we take this option, then we will fall into much the same type of problem as we did in the first order case. By admitting a new layer of objects for use in the definition, we open up the question of the categoricity of the new theory. On the other hand, we can also provide an axiomatisation of the satisfaction relation. Thus we expand the language with a new relation symbol $\models$ and axiomatise it in the usual way. Call the resultant theory Γ. We observe that, strictly speaking, we have moved into third order arithmetic in the sense that the new relation is between classes and numbers. However, in contrast to the attempt to define satisfaction, we have no need to quantify over representatives of a third level. As such, we shall regard this foray into the third order as ontologically innocent. We make $PA^2 + \Gamma$ our frame theory and from this position, we can state and prove the theorem.

THEOREM 3.8. ($PA^2 + \Gamma$) *Any two models $\mathcal{A}$ and $\mathcal{B}$ of PA^2 are isomorphic.*

Proof. (Sketch) The proof is carried out, as before, by constructing a bijection F between the domain of $\mathcal{A}$ and $\mathcal{B}$. The only difference is that in this case, we do not need to stipulate that the models are well-founded. The fact that the induction axiom (interpreted over every subsets of the domain) is satisfied in $\mathcal{A}$ and $\mathcal{B}$ is sufficient. □

So now we have a proof of the categoricity of PA^2 conducted from the same theory expanded with its satisfaction definition. We have avoided the need to expand our ontology with a new sort of objects: the regress has been compressed. Of course, a corresponding semantic regress has been started in that we need to define a satisfaction relation. But this should not be too surprising. We were not expecting to get around Tarski's theorem. The key difference is that our categoricity theorem has avoided the extra ontological outlay

[8] We should remark that models of PA^2 are being represented by entities with a contentious ontological status. We shall, however, do our best to gloss over this issue as the arguments of this paper stand regardless of our position on this matter.

incurred by the first order approach. So in contrast to the first order approach, we appear to be getting much closer to our initial goals.[9]

§4. A comparison between the first and second order perspectives. Before we revisit to the philosophical goals described in Section §2, we attempt some analysis of the distinction between the first and second order approaches described in the previous sections. In Section 3.2, we gave an illustration of the distinction via a representation in graph theory. We said that in the multi-sorted case, we were working in first order logic and thus the class variables should still be interpreted by vertices. On the other hand with second order logic, only the lower variables $(x, y, ...)$ should be represented by vertices while the upper variables ought be interpreted by something like loops around the appropriate collection of vertices. There is no need to add new vertices for the classes. Of course, this is only an illustration, but it helps to motivate the following question: why don't we need any extra objects in second order logic? The answer is that we *intend* there to be a unique standard second order model of any structure. We do not need to specify which collections are required, for the simple reason that we want *all* of them. Underlying this approach is a substantive philosophical thesis: that it is always intelligible to do this. We might make this clearer with the following semi-formal definition.

DEFINITION 4.1. *Given a structure, its* superstructure *is formed by taking the set of all subcollections of the domain and expanding the model accordingly.*

For example, given the standard model of arithmetic $\mathbb{N} = (\omega, 0, s, +, \times)$, its superstructure is formed by taking the powerset of ω and augmenting the signature of the structure with it: thus giving $\boldsymbol{\mathcal{N}} = (\mathcal{P}(\omega), \omega, 0, s, +, \times)$. If we were using a theory like ACA_0, we would let the class variables range over elements of $\mathcal{P}(\omega)$.

[9] We should stress that it is possible to prove a more traditional version of the categoricity theorem that uses second order logic but does not expose itself to the semantic regress. Rather than talking about theories and models, we take arbitrary relations N_1, S_1, 0_1, N_2, S_2 and 0_2 where we intend these relations to represent the domains, zero-elements and successor relations of two models of arithmetic. For brevity, we ignore addition and multiplication. We then suppose that both $N_1, S_1, 0_1$ and $N_2, 0_2, S_2$ satisfy the axioms of PA^2 and we do this by simply formulating the axioms of PA^2 using $N_1, 0_1$ and S_1in the first instance and then $N_2, 0_2$ and S_2 in the second. From here we can show that there is a relation R witnessing a bijection between N_1 and N_2 which maps 0_1 to 0_2 and preserves structure between S_1 and S_2. This is a straightforward theorem of second order logic in the sense that, unlike in the approach taken above, no frame theory needs to be taken up. For a detailed discussion of this approach, see Väänänen (2012). However, there is a sense in which we have lost the ability to properly talk about models. We were able to get away with not using a satisfaction predicate by representing a model by a short series of relations. However, in doing this we are no longer able to actually articulate categoricity as a theorem about a relation between a theory and a model because we have forgone the capacity to talk about theories. Shapiro's (1991) preference for theories capable of doing semantics would make this too high a cost. This is not necessarily a deep philosophical issue as we still appear to have shown that relations obeying PA^2 are unique up to isomorphism. But if we wanted to gain the expressive power to talk about theories, then we would need a satisfaction relation and then semantic regress could get started.
Moreover, no restriction on the cardinality of N_1 and N_2 is presupposed, so in this sense, we have a more natural representation of an arbitrary model. However for our purposes the admission of quantification over every possible relation pushes us to make too strong a commitment, too early. I do not want to suggest that this is a problem for the proof, merely that it will end up obstructing our attempt to evaluate the epistemic leverage provided by it. Our goal is to get a clearer idea of just what is required in order to obtain categoricity.

https://doi.org/10.1017/S1755020313000178 Published online by Cambridge University Press

We then propose the following thesis.

Thesis. *(SST) Every structure has a unique superstructure.*

My contention is that a proponent of the second order approach to categoricity should be an adherent of SST. In accepting the second order perspective with regard to some particular structure, they already accept that it makes sense to quantify over all subsets of the domain. Thus there should be no problem if we construct a multi-sorted structure in which the new class of objects are intended to represent every subset of the domain. Given a commitment to the second order logic approach, there is no reason to reject that any structure has a unique superstructure. To illustrate this, we might imagine a person who having conducted the categoricity proof for second order logic was asked to consider whether the standard model of arithmetic had a unique superstructure. Assuming the person has a relatively ordinary understanding of model theory, it does not seem plausible for them to reject this.[10]

Such a line of reasoning follows for small models regardless of our metatheory. But if we take this to the extreme, we reach the point at which the definition above should be regarded as merely semi-formal. This is because for some structures, the resultant superstructure may require a shift in frame theory, if we are to capture it formally. For example, we saw that we needed to move into a three sorted theory $\Delta^2_\infty - CA_0$ in order to be able to naturally represent the superstructure of a first order model of arithmetic. More generally, we note that it is a theorem of ZFC that every (set-sized) model has a unique superstructure. But it is not a theorem of ZFC that (proper) class models of ZFC have unique superstructures. It is in this sense, adherence to SST is a philosophical rather than mathematical position as there is, so to speak, no conventional mathematical position from which it can be articulated.[11] This discussion is not intended to argue one way or another whether SST is correct or not, merely to identify an acceptance of it with the second order perspective. In contrast, we shall identify a rejection of SST with what we have called the first order perspective in the ensuing discussion.

We should also mention a tangential worry that may emerge here. Given that SST is a trivial theorem of ZFC when we consider set-sized structures, it could appear to be utterly uncontentious and perhaps irrelevant to the issue at hand. We shall deal with this point thoroughly in Section §6. However, we remark that the theorem as articulated from the point of view of ZFC is not of much value to us. We have no more reason to think it picks out a unique structure, from the philosophical view, than ACA_0. Moreover, as we shall see later its increased ontological burden puts it in an arguably worse position. Our goal

[10] Shapiro (1990) draws a distinction between what he calls the "logical" and "iterative" conceptions of set, which is pertinent here. A supporter of the coherence of this distinction may claim that they do no need to support SST on the basis that SST makes use of "iterative" sets while the standard semantics for second order logic make use of "logical" sets. I contend that this distinction is, at heart, just another way of articulating SST. It is hoped, however, that this way of describing the situation provides a clearer way of investigating the issue. Moreover, the thought experiment above demonstrates that even if we take this distinction seriously, the "iterative" expansion of SST should still be accepted.

[11] We should note that we could move into higher order set theories which would allow us to talk about models of ZFC and their superstructures. Such a move is unpopular in both mathematics and philosophy, although for arguably different reasons. However, even if, for the sake of argument, we permit the move, we then have the problem of talking about structures and superstructure for that theory. Moreover, the indefinite pursuit of such a project is merely philosophically describable rather than formally so.

in the previous paragraph is merely to show that an adherence to ZFC in concert with a sympathy with second order logic will uncontentiously give us SST for small models. The situation for models of ZFC is more complex but outside the scope of this paper.

§5. Reviewing the goals of categoricity. We now review the goals set out in Section §2. in light of the discussion of the previous two sections. We shall review each goal from the perspectives of the first and second order perspectives respectively.

5.1. Demonstrate that there is a unique structure corresponding to some intuition or practice.

5.1.1. First order perspective. Suppose that we wished to demonstrate that, up to isomorphism, there was only one model corresponding to our arithmetic practice. As we have discussed, this is a desirable thing to know. Our mathematical practice is carried out in such a way that we appear to expect that this is the case, and it would be congenial to have some way of demonstrating it. In Section §3.1, we conducted a categoricity proof of PA from the perspective of the frame theory, ACA_0. The proof was successful in the sense that we showed that every well-founded model of PA was isomorphic to every other. At first blush, this may appear to have been exactly what we were after. But there was a problem. In order to produce the proof, we needed to increase the expressive power of our language and theory by introducing a new sort of object. The resultant theory was formulated in (a form of) first order logic and as such, it has nonstandard models. So although we have proven that the target theory is categorical, we have only achieved this from a frame theory whose fixity is itself open to question.

We then proposed a pathological solution: try to prove the categoricity of the frame theory using the same technique. If this had been successful, we would have shown that there is only one model for the frame theory and *a fortiori* only one model of the target theory. However, the same problem raised its head. We end up with another frame theory formulated in first order logic, which again has nonstandard models. Moreover, any continuation of this proposal ends up with the same result. We show categoricity of the target theory, but only from the perspective of a frame theory which is itself in question. Thus the pathological solution leads to a regress.

Now, of course, the existence of these nonstandard models is a trivial consequence of the Löwenheim-Skolem theorems and we have long since learned to live with and indeed rejoice in these results. We grant this, but also observe that the issue at hand is not whether or not there are unique models corresponding to our practice, but whether the second order categoricity theorems help us establish this. And in this regard, we see that it does not.

To put the point into sharper focus, we observe that the regress of the pathological solution pushes us to consider theories whose categoricity is also open to question, but theories that are, at every step, more questionable than the previous level. At every step, we consider a theory which is staggeringly more complex than the previous step. I contend that at every metatheoretic step we take back, our epistemic confidence about categoricity and its coherence should weaken: the ice is getting thinner. To see this, we need to consider our goal. We are trying to show that our practice of, say arithmetic, makes sense. We are trying to show that all of those times when we acted as if there was a unique structure of arithmetic, up to isomorphism, that this was a sensible thing to do. To appeal to structures with regard to which we have even less confidence in order to establish this result is to put the cart before the horse.

We might appeal here to an analogy with another property of theories: consistency. Gödel showed us that the ideal of an absolute consistency proof in which we show the consistency of some theory within its own confines is unattainable. We can only obtain relative consistency proofs: proofs of the consistency of some target theory which are contingent upon the consistency of the frame theory from which they are conducted. As such, the epistemic value of such consistency proofs is generally limited by the fact that we have less confidence in the consistency of the frame theory from which the proof is conducted. It is a riskier proposition. Similarly, we should see the regress as a demonstration that absolute categoricity is not available, only relative categoricity.

The theorem is only as good as the metatheory from which it is conducted. For example, if we formalise our metatheory in the usual way (i.e. in ZFC) then the categoricity theorem tells use that arithmetic is categorical *modulo* the categoricity of ZFC. We thus support a generally accepted claim with a plausibly contentious one. In the base case of arithmetic, with regard to which there was almost universal assent to uniqueness, we had an antecedent belief in that uniqueness based on experience and plausible intuition. However, as we move further back our reasoning tends to operate via analogy and if our intuitions are not bankrupt, then they are at least testing the overdraft.

5.1.2. Second order perspective. On the other hand, in Section §3.2, we obtained a categoricity theorem that did not ensue in a explosive, ontological regress. We showed that from the point of view of $PA^2 + \Gamma$ that any two models of PA^2 are isomorphic. If our goal was to establish the uniqueness of structures corresponding to our arithmetic practice, then we appear to have achieved this at the modest price of an axiomatisation of satisfaction (as opposed to a layer of new classes). However, there is a problem here as well. To put it bluntly, given a commitment to SST, I contend that the categoricity result is unsurprising and as such, its value is minimal. Little or no epistemic leverage is gained on our problem. We defend this with two arguments.

First, we argue that there is little force for a claim that there is a unique structure corresponding to some practice which is supported by the claim that a much more complicated structure is unique. To illustrate this, we observe that there is a certain similarity between the content of SST and what we are trying to achieve. In the case of arithmetic, we are trying to show that there a *unique* structure corresponding to our arithmetic practice. The acceptance of SST in this situation amounts to claiming that any such structure has a *unique* superstructure. Both propositions are about the uniqueness of a structure of some kind. However, SST asserts the existence of a significantly more complex structure. In the case of arithmetic, we have seen that each of the new elements added in the superstructure is able to provide a natural representation of an example of the substructure. We used this fact to make class objects represent models of the substructure and prove the theorem. Thus there is a sense in which we are using the uniqueness claim about a complex structure to lever a result about a comparatively simple one. Now given that we are assuming acceptance of SST, there is no reason why we cannot put this claim to use. The issue is rather, whether we should be surprised by the result. But we see that the claim for the uniqueness of the structure of arithmetic rests on the antecedent belief in the uniqueness of a structure which, in some sense, already includes that of arithmetic.

Second, there is a sense in which, given SST, we are already committed to the existence of the substructure. To see this, let us take the superstructure thesis further. Beginning a putative model of arithmetic, let us take its superstructure. Then let us take the superstructure of that and continue the process indefinitely. This gives us something like the simple theory of types (Church, 1940). Moreover, we could commence instead with the empty set and iterate the superstructure process indefinitely into the absolute. This then yields the cumulative hierarchy V. At each level we are committed to the uniqueness of the superstructure of the previous one. Moreover following Shepherdson (1952), we get categoricity at every level.[12] Now given an acceptance of the uniqueness of the cumulative hierarchy (or even just a large segment of it), the acceptance of a unique model of arithmetic (or any other set sized model) is a triviality. Nothing new is being said.

This is not to suggest that that acceptance of a unique cumulative hierarchy is in any way wrong, but it is to say that the categoricity theorem fails to demonstrate in any philosophically satisfying way that there is a unique structure corresponding to our practice. The proof is useless for this purpose unless you already believe that there is only one model of your frame theory, in which case you ought to already believe that there is only one model of arithmetic.

5.2. Demonstrate that some theory picks out a unique structure. We now turn to the second of our goals for categoricity proofs. In contrast to the previous goal, we are now uninterested in showing that there is a unique structure corresponding to our practice. For whatever reason, we shall just take it that there is one. But it is, of course, still desirable to have some means of referring to this structure. Our goal is to show that a particular theory is capable of doing this.

5.2.1. Second order perspective. Assuming that we accept SST, then the categoricity theorem does tell us something valuable. Returning to the example of arithmetic, it tells us that any two models of PA^2 are isomorphic. Moreover we shall assume a belief that the theorems of PA^2 are correct in that structure. So while we have not tried to demonstrate that *there is* a unique structure satisfying the axioms, we have been able to demonstrate that any two structures satisfying the axioms are isomorphic. Thus we have shown that our axiomatic enterprise has been successful. Given an antecedent belief that there is a unique structure of arithmetic we have shown (using SST) that PA^2 successfully picks out that structure. Following Read (1997), we might interpret the goals of Frege and Dedekind in this way and thus that these projects are vindicated.

So we should concede the value of categoricity proofs in this regard. We may, however, wonder how interesting a fact this is nowadays. In a broader mathematical context, it is really just part of a demonstration that a theory is well defined. A consistency proof may be understood as demonstrating the existence of structures satisfying the theory, while the categoricity proof may be understood as proving uniqueness. It is certainly good to know: it gives us, in a sense, a licence to refer to *the* model of arithmetic, analysis or the simple theory of types. It is not, however, a particularly unusual result.

5.2.2. Demonstrating every sentence has a determinate truth value. Now let us consider a philosopher who is not concerned about whether there is a unique structure corresponding to our practice. She may may want to take a step back from what may be

[12] There is a sense in which I'm offering a crude absolute categoricity theorem here. We have categoricity in the sense that every *well-founded, inner model* of ZFC which is *closed under subsets* is isomorphic to V. Using second order logic we can articulate a theory, ZFC^2 (see Shapiro, 1991), in which we can express what it means for a model to be well-founded and closed under subsets, but there is no way of ensuring that a model will be inner and thus exhaust the ordinals. This, of course, raises the question of what 'all the ordinals' means, but we defer this question to a future paper which focuses more directly on the issues of categoricity and set theory.

called ontological realism. For her purposes, it may be sufficient to demonstrate merely that every sentence has a determinate truth value. For some philosophical purposes, this may prove sufficient for the defence of the intelligibility of mathematical discourse and practice: McGee (1997) may be interpreted as arguing along these lines. Given this we may wonder whether the categoricity theorem could be useful for this goal. The fact that any two models of some theory are isomorphic tells us that they are elementary equivalent, so this seems promising. We argue that it is not.

Consider again the case of arithmetic. First we observe that Theorem 3.4 for PA also applies to arbitrary extensions of PA, including extensions which are pairwise inconsistent. Moreover, we can show that such extensions have models. For example:

PROPOSITION 5.1. *(ZFC) There exists a model $\mathcal{M}$ of ACA_0 in which there is a class model $\mathcal{A}$ such that*

(i) $\mathcal{A} \models PA + \neg Con(PA)$*; and*
(ii) $\mathcal{M} \models$ "*$\mathcal{A}$ is well founded.*".

Proof. ACA_0 is a conservative extension of PA, so there is a model

$$\mathcal{M} = (M^2, M^1, 0^{\mathcal{M}}, s^{\mathcal{M}}, +^{\mathcal{M}}, \times^{\mathcal{M}}, \in^{\mathcal{M}})$$

such that $\mathcal{M} \models ACA_0 + \neg Con(PA)$. We then define a class model

$$\mathcal{A} = (M^1, 0^{\mathcal{M}}, s^{\mathcal{M}}, +^{\mathcal{M}}, \times^{\mathcal{M}}).$$

This is a model of $PA + \neg Con(PA)$. Moreover, since $\mathcal{M}$ satisfies the induction axiom, there are no classes from M^2 which witness a nonwell-founded subset of M^1. Thus, from the point of view of $\mathcal{M}$, $\mathcal{A}$ is well founded. □

REMARK. *We use ZFC as the background theory in the proof above for convenience. Also note that $\mathcal{A}$ is intended to be a class model, while $\mathcal{M}$ is an ordinary model. Observe that $\mathcal{M}$ is not an ω-model of ACA_0.*

Now from the point of view of $\mathcal{M}$ we see (using Theorem 3.4) that every well founded model of $PA + \neg Con(PA)$ is isomorphic to every other. But clearly, this is problematic. We wanted to show that every sentence of PA has a determinate truth value but presumably $Con(PA)$, is in fact, true and here we have model in which it has come out false. Moreover we may wonder what such an $\mathcal{M}$ *thinks* about models of $PA + Con(PA)$: the models which are in some sense closer to being correct. The following corollary of categoricity helps here.

COROLLARY 5.2. *(ACA_0) If $\mathcal{A}$ and $\mathcal{B}$ are such that $\mathcal{A} \models PA + T$ and $B \models PA + S$ where $PA + T$ and $PA + S$ are inconsistent with each other, then at most one of $\mathcal{A}$ and $\mathcal{B}$ can be well founded.*

Proof. Suppose $\mathcal{A}$ and $\mathcal{B}$ are well founded. Then since $\mathcal{A}, \mathcal{B} \models PA$, $\mathcal{A} \cong \mathcal{B}$ by categoricity. Then we have $\mathcal{A} \equiv \mathcal{B}$ which contradicts the assumption that $PA + T$ and $PA + S$ are inconsistent with each other. □

Thus in the example above, $\mathcal{M}$ thinks that every model $\mathcal{B}$ of $PA + Con(PA)$ is nonwell-founded. Moreover if $\mathcal{M}$ could find an infinite descending $<$-chain in $\mathcal{B}$, then we would continue to find such a chain if we moved to larger frame theories. Thus assuming $Con(PA)$ is true, $\mathcal{M}$ does contain not any class model which represents our intended model of

https://doi.org/10.1017/S1755020313000178 Published online by Cambridge University Press

arithmetic.[13] The situation is thus much the same as in Section 5.1.1. and similarly the only way around the problem is to commit to SST.

5.3. Classify different types of theory. Finally, we consider the more modest goal of using the categoricity proof as a means to classify types of theory. We considered earlier the distinction between algebraic and nonalgebraic theories Shapiro (1997). Examples of algebraic theories included the theories of groups, fields and Euclidean geometry. On the other hand, our canonical examples of nonalgebraic theories included theories like PA, analysis and set theory. The distinguishing feature between them was that in the first case, we intended there to be multiple pairwise nonisomorphic instantiations of the theory, whereas in the latter case, we intended that there be just one. However, if we start to ask what the distinction is founded upon, then our only means to cleave them apart appears to be via our *intention*. We intend some theories to be uniquely instantiated and others we do not. We propose that the categoricity theorem could be used as a means to formalise this distinction. Thus if we are able to characterise our intended model up to isomorphism then the categoricity proof can be seen as showing us that the theory is nonalgebraic. Metaphysical questions about whether or not there is some unique structure are beside the point.

Further, we may also be interested in weaker forms of categoricity. For example, as is well-known, the second order version of ZFC, ZFC^2 is merely almost categorical. If $\mathcal{A}$ and $\mathcal{B}$ are models of ZFC^2, then either $\mathcal{A} \cong \mathcal{B}$ or one of them is isomorphic to an initial segment of the other. This phenomena is not restricted to set theories and can be found in any theory which contains a merely semi-discrete linear order.[14] A better understanding of this phenomena may assist in understanding the problem of absolute categoricity.

§6. Unique superstructures. Having reviewed the philosophical goals, we now look to a possible problem for much of the discussion above. We have assumed that a rejection of, or at least doubt with regard to, SST is philosophically viable. For example, with regard to the first goal, this motivated our problems from the first order perspective and also informed the argument regarding the limitation of value of the second order perspective. But what if there is no reason to doubt SST? Then all this posturing is baseless. Moreover, while the value of the categoricity proofs may sometimes be overstated, no real harm can come from this since this extra philosophical assumption is correct.

6.1. An objection and a response. In the following passage Shapiro defends what I take to be a version of SST. His contention is that the meaning of *all* the subsets of some domain d is unambiguous. Thus, in the language above, any structure has a unique superstructure.[15]

---

[13] Of course, this example is a little contrived. Following Feferman's reflection programme, we see from our practice that we have good reason to say that $Con(PA)$ is true (Feferman, 1964). However, if we move to a theory like third order arithmetic, we are free to select a Σ^2_1 statement like CH whose decidability one way or another, or at all, is a more contentious matter.

[14] By semi-discrete, we mean that every object has an immediate successor but not necessarily an immediate predecessor. If we admit an appropriately formulated second order induction axiom or restrict ourselves to well founded models of such a theory, then we can only obtain quasi-categoricity for it. The problem for ZFC^2 has little to do with the ambitiousness of the theory.

[15] We should note that Shapiro frames the problem in terms of logical, as opposed to iterative, sets. For the reasons discussed in footnote 10, we shall pass over this point.

> I submit that what it means for a collection c to be a (logical) subset of d is clear and unambiguous: c is a subset of d if and only if every member of c is a member of d. What is at issue here is whether the totality (or range) of subsets of d is itself clear and unambiguous. Let $P1$and $P2$ be two candidates of the range of the second order predicate quantifiers of T (vis-a-vis d). That is let $P1$ and $P2$ be two candidates for the logical powerset of d. I suggest that if $P1 \neq P2$, then there is clear sense in which (at least) one of them is not the powerset of d. Indeed suppose that there were a collection c such that $c \in P1$ but $c \notin P2$. I take it that (for a classical mathematician) it is determinate whether every element of c is an element of d . If every element of c is in d then $P2$ is not the powerset of d; otherwise, $P1$ is not the power set of d. ... I do not make the absurd claim that any or all of the properties of the powerset (such as its cardinality, or whether it contains a nonconstructible element) are known. (Shapiro, 1985)

Essentially, we use a (logical version of) extensionality to argue that given any two candidates $P1$, $P2$ for being the set of all subsets of d, at most one of them can be it. Thus we appear to have proven the uniqueness claim and demonstrated that SST is correct. This argument is very plausible. We shall reject it, but our response illustrates clearly what is at stake in a rejection of SST. First we note that there is nothing wrong with the moves made in the argument. The claims follow in a logical fashion; indeed, if we formalised it, it would be a straightforward theorem of ZFC. The problem is the first premise: that the meaning of 'c is a subset of d' is unambiguous. But how could this be? What does it mean for c to be a subset of d? c is a subset of d just means that all the members of c are members of d. This is, I think, the right answer and beyond it I do not think any more can be intelligibly said. Nonetheless I do not think that this sense of the term *unambiguous* will suffice for the application Shapiro has in mind. Yes, we appear to have unambiguously said what it means for something to be a subset of another in that we have provided a definition of this, but is this enough? Have we provided enough to be certain that this definition only means one thing? The problem is that a definition is only as good as the concepts from which it is constructed. We have given a definition of subset *modulo a conception of membership*. Moreover we do not define what it means for something to be a member of b; rather, we characterise its behaviour with a set of axioms. The axioms of second order logic make an approximation of what it means for a to be an element of b. However, they do not complete this job, and it is well-known that no addition of further axioms can do so. Thus we claim that the appropriate response to the argument above is to observe that the membership relation itself may be ambiguous and that without this possibility, there could be no reason to doubt SST.

We should stress that this is not intended to be a refutation of Shapiro's position, merely his argument. Our goal is not to demonstrate that SST is false, or true, but rather to defuse an argument to the effect that we are obliged to accept it.

6.2. Where does this leave us? In presenting this response to a number of people, I have noticed that, in general, people have found the corner in which the SST doubter finds himself to be somewhat unappealing: sometimes too much so. I cannot say that I share this intuition. So whilst remaining neutral with regard to SST, I would like now to make some remarks which may serve to soften the impact of my claim. First we should note that this response does not commit us to some kind of nonclassicality via intuitionism or three-valued indeterminacy. There is a sense in which we are saying that there is no fact of the matter whether or not some object is a member of a particular set. But this failure of excluded middle is a result of the ambiguity of membership: of indeterminacy with regard to which membership relation is under consideration. With regard to any particular membership relation, there is a fact of the matter. In a sense, we are understanding the theory of membership as algebraic rather than nonalgebraic. It is algebraic in the sense that we would accept that there *are* multiple interpretations of it. We might say that it is *ontologically algebraic*.[16] Now we do not think of the algebraic theory of groups as being nonclassical in that there is no fact of the matter whether addition is commutative. We can thus extend the same courtesy to the theory of membership.

Second we observe that this response does not relegate us to the realms of anti-realism or formalism. For example, Hamkins has developed a theory of the multiverse which takes seriously the idea that forcing extensions present us with genuine collection of alternative set theoretic universes (Hamkins, 2012; Hamkins, 2009; Gitman & Hamkins, 2010; Hamkins, 2007). To give a flavour of the theory the following is taken as an axiom:

- (Forcing Extension Axiom) For every universe and any forcing notion there is another universe which is a forcing extension of the other.

The full philosophical ramifications of this theory are yet to be probed, but it is certainly not intentionally formalist. As opposed to Mostowski (1967), we do not end up in a pragmatic position regarding our selection of a theory of sets. Rather, every universe is to be taken as seriously as any other and the resultant multiverse is regarded as a fresh mathematical domain, ripe for exploration. Indeed Hamkins has developed the means of navigating part of it using modal logic (Hamkins, 2007). Once again, the point here is not to reject belief in the uniqueness of V or to reject Shapiro's position outright. Rather our goal is to look more closely at the the bullet that needs to be bitten and to show that it may not be so unpalatable after all.

6.3. Conclusion. We summarise the main points of the paper. If we adopt the first order perspective and do not take up SST, then the most we can philosophically gain from the categoricity theorem is a means of classifying algebraic and nonalgebraic theories. On the other hand, if we take up the second order perspective and with it SST, we can verify that a theory picks out the structure in which we have some antecedent belief. However, there is no value in moving in the other direction and using the categoricity as evidence that a unique structure exists which instantiates that theory. We have claimed that the cost of doubting SST is to doubt the unambiguousness of the membership relation. And in closing we have tried to show that this cost might not be as dear as usually imagined.

§7. Acknowledgments. I would like to thank the *Plurals, Predicates and Paradox* project, the Arché Research Centre and the Melbourne Logic Group for the opportunity to present and hone the material discussed in this paper. In particular, I would like to thank Greg Restall, Øystein Linnebo, Stewart Shapiro, Allen Hazen, Hartry Field, Joel Hamkins,

[16] On the other hand, we intend, when we talk about, say, models of second order arithmetic from the perspective of third order arithmetic, that it is nonalgebraic: that there is only one correct interpretation of it. We might say that it *intentionally nonalgebraic*, at least in these restricted contexts.

https://doi.org/10.1017/S1755020313000178 Published online by Cambridge University Press

Stephen Read, Sean Walsh, Salvatore Florio, Zach Weber, Simon Hewitt and Conrad Asmus for their invaluable comments and assistance.

BIBLIOGRAPHY

Chang, C. C., & Keisler, H. J. (1973). *Model Theory*. Amsterdam, the Netherlands: North Holland Publishing Company.

Church, A. (1940). A formulation of the simple theory of types. *Journal of Symbolic Logic*, **5**(2), 55–68.

Dedekind, R. (1963). *Essays on the Theory of Numbers*. New York, NY: Dover Publications.

Feferman, S. (1964). Systems of predicative analysis. *Journal of Symbolic Logic*, **29**(1), 1–30.

Feferman, S. (1999). Does mathematics need new axioms? *American Mathematical Monthly*, **6**, 401–446.

Gitman, V., & Hamkins, J. D. (2010). A natural model of the multiverse axioms. *Notre Dame Journal of Formal Logic*, **51**(4), 475–484.

Halbach, V., & Horsten, L. (2005). Computational structuralism. *Philosophia Mathematica*, **13**, 174–186.

Hamkins, J. (2007). The modal logic of forcing. *Transactions of the Americal Mathematical Society*, **360**(4), 1793–1817.

Hamkins, J. (2009). Some second order set theory. In Ramanujan, R., and Sarukkai, S., editors, *Logic and Its Applications: Lecture Notes in Computer Science*, Vol. 5378. Heidelberg, Germany: Springer-Verlag, pp. 36–50.

Hamkins, J. (2012). The set-theoretic multiverse. *The Review of Symbolic Logic*, **5**, 416–449.

Kaye, R. (1991). *Models of Peano Arithmetic*. Oxford, UK: Oxford University Press.

Kreisel, G. (1969). Informal rigour and completeness proofs. In Hintikka, J., editor. *The Philosophy of Mathematics*. London, UK: Oxford University Press.

Marker, D. (2002). *Model Theory and Introduction*. New York, NY: Springer.

Martin, D. A. (2001). Multiple universes of sets and indeterminate truth values. *Topoi*, **20**(1), 5–16.

Mayberry, J. P. (2000). Review of J. L. Bell: A primer of infinitesimal analysis. *British Journal of the Philosophy of Science*, **51**, 339–445.

McGee, V. (1997). How we learn mathematical language. *The Philosophical Review*, **106**(1), 35–68.

Mostowski, A. (1967). Recent results in set theory. In Lakatos, I., editor. *Problems in the Philosophy of Mathematics*. Amsterdam, The Netherlands: North Holland Publishing Company.

Read, S. (1997). Completeness and categoricity: Frege, Gödel and model theory. *History and Philosophy of Logic*, **18**(2), 79–93.

Shapiro, S. (1985). Second-order languages and mathematical practices. *The Journal of Symbolic Logic*, **50**, 714–742.

Shapiro, S. (1990). Second-order logic foundations and rules. *The Journal of Philosophy*, **87**(5), 234–261.

Shapiro, S. (1991). *Foundations without Foundationalism: A Case for Second Order Logic*. Oxford, UK: Oxford University Press.

Shapiro, S. (1997). *Philosophy of Mathematics: Structure and Ontology*. Oxford, UK: Oxford University Press.

Shepherdson, J. C. (1952). Inner models for set theory—Part II. *Journal of Symbolic Logic*, **17**(4), 225–237.

Simpson, S. G. (1999). *Subsystems of Second Order Arithmetic*. Berlin, Germany: Springer.

Simpson, S., & Yokoyama, K. (2012). On the reverse mathematics of Peano categoricity. Unpublished manuscript.

Väänänen, J. (2012). Second order logic or set theory? *Bulletin of Symbolic Logic*, **18**, 91–121.

Weston, T. (1976). Kreisel, the continuum hypothesis and second order set theory. *The Journal of Philosophical Logic*, **5**, 281–298.

SCHOOL OF MATHEMATICS
UNIVERSITY OF BRISTOL
E-mail: toby.meadows@gmail.com