STUDIA UBB. PHILOSOPHIA, Vol. 63 (2018), 3, pp. 31-65
(RECOMMENDED CITATION)
DOI:10.24193/subbphil.2018.3.02

DETERMINACY OF REFERENCE, SCHEMATIC THEORIES, AND INTERNAL CATEGORICITY

ADRIAN LUDUȘAN*

ABSTRACT. The article surveys the problem of the determinacy of reference in the contemporary philosophy of mathematics focusing on Peano arithmetic. I present the philosophical arguments behind the shift from the problem of the referential determinacy of singular mathematical terms to that of nonalgebraic/univocal theories. I examine Shaughan Lavine's particular solution to this problem based on schematic theories and an 'internalized' version of Dedekind's categoricity theorem for Peano arithmetic. I will argue that Lavine's detailed and sophisticated solution is unwarranted. However, some of the arguments that I present are applicable, mutatis mutandis, to all versions of 'internal categoricity' conceived as a philosophical remedy for the problem of referential determinacy of arithmetical theories.

Keywords: *Determinacy of reference, Peano arithmetic, permutation argument, structuralism, Dedekind's categoricity theorem, schematic theories, internal categoricity*

The central problem[1]

The central problem of this article concerns the determinacy of reference for those mathematical theories whose intended subject matter is a certain mathematical structure[2]. More precisely, the philosophical problem that we are considering is how

* Department of European Studies and Governance, Emmanuel de Martonne no. 1, Cluj-Napoca, RO- 400090, adiludusan@gmail.com

[1] The philosophical issue that I will address in this paper is an instance of what Shaughan Lavine defined in his manuscript, *Skolem was wrong,* as the 'central problem'. Since Lavine's detailed and sophisticated argument will be the focus of my paper, I kept his way of naming the issue.

[2] Of course, there are mathematical theories, such as group theory, ring theory, etc whose axiomatizations are not supposed to pick up a unique structure modulo isomorphism. Following (Shapiro 1997), I shall call such theories, 'algebraic', leaving the characterization 'non-algebraic' for those mathematical theories whose axiomatization is supposed to determinately refer to a unique structure up to isomorphism, such as Peano Arithmetic, analysis, etc. For reasons of clarity (Button and Walsh 2018) contrast algebraic theories with univocal ones.

can a theory such as Peano Arithmetic (PA) manage to characterize, up to isomorphism, its intended subject matter, that is, the natural number structure that we all know and love. A few qualifications are needed in order to unpack the central problem, one methodological, and the rest philosophical.

(1) The problem arises for some profiles of positions in the philosophy of mathematics with some discernable epistemological and ontological features.

Ontologically, the problem arises for a structural realist in the philosophy of mathematics. I will call such a position platonism, although I am aware that that forces the label 'platonist'. What 'realism' means in this context is the combination of three traits, *existence*, *independence,* and *abstractness* of mathematical objects. The first two traits are formal, and concern the status of mathematical objects, while the latter is material and regards their nature. The belief that mathematical entities are *bona fide* existing objects with distinctive properties defines the *existence* trait, the belief that these objects are not our creation, defines the independence trait, and the belief that mathematical objects have a non-spatial, non-temporal, acausal nature forms the abstractness trait.

Epistemologically, the problem arises for what Button & Walsh[3] call a 'moderate' position. An easy way out of the problem of how we can determinately refer to mathematical structures or objects is to attribute to the mind some mysterious faculties, like a mathematical intuition, that enables the mind to glue the theories/ singular mathematical terms to the envisaged structures/mathematical objects. By contrast, a moderate position presupposes the rejection of any talk of intellectual or mathematical intuitions, or for that matter, any mysterious faculties of the mind, and focuses only on philosophical positions capable of offering naturalistically approved explanations. In our case, this means that the explanations have to be semantically traceable. Accordingly, from a moderate perspective, if anything fixes the reference, then the theory and its semantics ought to do it.

(2) I will talk of determinacy of reference of mathematical theories only up to isomorphism for reasons that I will develop and explore in the next two sections.

(3) I will construe the informal talk of 'mathematical structures' as isomorphism types, as is the practice of many mathematicians, thus restricting the analysis to what Button & Walsh[4] call modelism.

[3] (Button and Walsh 2018, 6.3)
[4] (Button and Walsh 2018, 38)

(4) The methodological framework in which I will conduct the analysis is the standard model-theoretic one and at times set-theoretic, customarily employed in textbook presentations of first and second-order logic. I assume that the reader is familiar with these frameworks.

In short, the philosophical setting is constituted by platonism, moderation, and modelism, and the instruments of analysis are the standard model-theoretic ones.

The 'push-through construction' and the permutation argument

There are two arguments for focusing on the referential determinacy of non-algebraic theories, rather than singular mathematical terms, and for considering structures only 'up to isomorphism' as such referential candidates. The first one is based on an elementary result from model theory, the 'push-through construction'[5], and it is known as 'the permutation argument'[6], while the second is based on technical results in set theory regarding different, but equivalent set-theoretic reconstructions of the natural number structure, and it is known as 'Benacerraf's identification problem'[7]. Let us develop the two arguments, with an emphasis on the first one.

Before outlining the permutation argument, we need to state some definitions and basic results in model theory.

In model-theoretic semantics, one typically assigns certain entities of the domain M to each item of the signature[8] $\mathcal{L}$:

i. to every constant $c_i \in \mathcal{L}$, an element $c_i^{\mathcal{M}} \in M$.
ii. to every n-ary relation symbol $R_i \in \mathcal{L}$, a subset $R_i^{\mathcal{M}} \subseteq M^n$.
iii. to every n-ary function symbol $f_i \in \mathcal{L}$, a corresponding n-ary function, $f_i^{\mathcal{M}}: M \to M$.

Variables v_i , $i \in \mathbb{N}$, are taken to range over the domain M.

Observe that these specifications can be viewed as a schematic referential explanation of the constitutive items of $\mathcal{L}$. More precisely, consider an $\mathcal{L}$-structure $\mathcal{M} = <M, c_i^{\mathcal{M}}, R_i^{\mathcal{M}}, f_i^{\mathcal{M}}>$. The structure explicates reference in a similar manner to that of natural languages like English, by assigning to each constant $c_i \in \mathcal{L}$ ($\mathcal{L}$'s

[5] The name originates with (Button and Walsh 2016, 284).

[6] Although permutation arguments have a long history – see (Button 2013, 25) our focus will be on the permutation argument developed by (Putnam 1981, 33–5, 217–18).

[7] (Benacerraf 1965)

[8] I will only consider at most countable signatures since nothing on the arguments involved in the subsequent analysis relies on the cardinality of the signature.

correspondent of a proper name) an element of the domain M, to each predicate $R_i \in \mathcal{L}$ ($\mathcal{L}$'s name of a property/relation) a certain subset, etc. In sort, reference for singular mathematical terms is fixed by stipulation and it has a non-descriptivist character. Based on the above specifications, one then recursively defines in model-theoretic terms the notions of *satisfaction* and *truth*. For subsequent discussions, it is important to note that truth is a relation between a structure $\mathcal{M}$ and an $\mathcal{L}$-sentence φ, usually symbolized like this, $\mathcal{M} \vDash \varphi$. The more general notion of *satisfaction* is a relation between a structure $\mathcal{M}$ with an assignment s from the set of variables to M, and a well-formed formula (wff from now on), symbolically $\mathcal{M}, s \vDash \varphi(\bar{v})$, where $\bar{v}$ is an n-tuple $<v_1, v_2, \dots, v_n>$ of free variables. In both cases, the relation $\vDash$ connects a model-theoretic structure with a proper linguistic construct. I assume that the reader is familiar with such definitions and with their generalization to $\mathcal{L}$-theories, not just particular sentences, in which case, we speak of the structure $\mathcal{M}$ as a model of any such $\mathcal{L}$-theory T. Note that an $\mathcal{L}$-model $\mathcal{M}$ of an $\mathcal{L}$-theory T, makes true – in the technical sense of model theory – all assertions in T which intuitively should be true, and false the assertions which intuitively should be false. Briefly stated, for any T-sentence φ, $\mathcal{M} \vDash \varphi$, if and only if (abbreviated *iff* from now on) φ is (intuitively) true.

Tacking stock, model theory provides explanatory referential schemas for $\mathcal{L}$ signatures, recursive definitions of truth and satisfaction, which enables the generalization to theories and models.

In model theory, one can easily construct an isomorphic copy of any such structure $\mathcal{M}$. The only requirements are that we have a set N with the same cardinality as M and a bijection $\pi : M \to N$ – but these are not serious issues since we can take $N = M$, and consider π a nontrivial permutation of M. A basic recipe for constructing an isomorphic copy is the following:

Push-through construction: Let $\mathcal{L}$ be any signature, $\mathcal{M} = <M, c_i^{\mathcal{M}}, R_i^{\mathcal{M}}, f_i^{\mathcal{M}}>$ any $\mathcal{L}$-structure, and $\pi : M \to N$ any bijection. Define another $\mathcal{L}$-structure, $\mathcal{N} = <N, c_i^{\mathcal{N}}, R_i^{\mathcal{N}}, f_i^{\mathcal{N}}>$ by:

i. $c_i^{\mathcal{N}} = \pi(c_i^{\mathcal{M}})$,
ii. $R_i^{\mathcal{N}} = \{<\pi(m_1), \pi(m_2), \dots, \pi(m_n)>/<m_1, m_2, \dots, m_n> \in R_i^{\mathcal{M}}\}$
iii. $f_i^{\mathcal{N}}(\pi(m_1), \pi(m_2), \dots, \pi(m_n)) = \pi(f_i^{\mathcal{M}}(m_1, m_2, \dots, m_n))$.

In these conditions, π defines an isomorphism, and we say that $\mathcal{M}$ and $\mathcal{N}$ are isomorphic structures, in symbols $\mathcal{M} \cong \mathcal{N}$.

Isomorphic models preserve the truth-values of all formulas (hence, in particular, of all sentences). If two $\mathcal{L}$-structures $\mathcal{M}$ and $\mathcal{N}$ satisfy exactly the same $\mathcal{L}$-sentences, we say that the structures are *elementarily equivalent*, in symbols $\mathcal{M} \equiv \mathcal{N}$.

Resuming, we can say that if two structures are isomorphic, then they are elementarily equivalent, which is a basic result in model theory often stated as a corollary of the following theorem:

Theorem 1. Let $\mathcal{M}$, $\mathcal{N}$, be any two $\mathcal{L}$-structures such that $\mathcal{M} \cong \mathcal{N}$, with $\pi : M \rightarrow N$ the isomorphic bijection. For all $\mathcal{L}$-formulas $\varphi(\bar{v})$, $\mathcal{M}, s \vDash \varphi(\bar{v})$ iff $\mathcal{N}, \pi \circ s \vDash \varphi(\bar{v})$.

The proof of the theorem is by induction on the complexity of the formulas.

Now, the permutation argument is simply a philosophical usage of the push-through construction in order to undermine the determinacy of reference as explained above i.e. in the model-theoretic semantics. Suppose that one has formulated a nonalgebraic/univocal $\mathcal{L}$-theory *T*, such as Peano Arithmetic, with an intended model $\mathcal{M}$. Obviously, stipulation alone cannot fix the reference of singular terms such as c_i, $f(c_i)$, etc., we can always specify another referential schema in which the referents of all constants c_i, predicates and functions R_i, f_i of the $\mathcal{L}$-theory *T* are different from those in $\mathcal{M}$. A far better candidate for referential glue is represented by the truth-value of sentences. Maybe the truth-value of sentences in which a certain singular term occurs imposes the reference of that singular term. It is precisely this account of the determinacy of reference of singular terms that the permutation argument dismantles. In the intended model $\mathcal{M}$, each singular term has a definite referent; for example, the referent of c_1 in $\mathcal{M}$ is a certain object $c_1^{\mathcal{M}}$. Apply the push-through construction to this intended model, with $N = M$, and π a nontrivial permutation of *M*. In the generated model, call it $\mathcal{M}'$, at least one singular term has a different referent than the one assigned in $\mathcal{M}$, say the interpretation of c_1 in $\mathcal{M}'$ is a definite object $c_1^{\mathcal{M}'}$ which is different from $c_1^{\mathcal{M}}$, the interpretation of c_1 in $\mathcal{M}$. If the truth-values of sentences were enough to glue names to referents, then some truth-values of sentences containing c_1 will differ in the two models, $\mathcal{M}$ and $\mathcal{M}'$. But the push-through construction ensures us that $\mathcal{M}'$ is isomorphic to $\mathcal{M}$, and by the corollary to the theorem 1, $\mathcal{M}'$ is elementary equivalent to $\mathcal{M}$, that is the models are indiscernible with respect to the truth-values of all the sentences. To illustrate this procedure, suppose that the signature $\mathcal{L}$ contains the names of those celestial bodies within our solar system that have been named so far, and the predicate 'is a planet'(abbreviated *P*), while the intended structure $\mathcal{M}$ has a domain *M* that contains all celestial bodies within our solar system (either named or not) and the other ingredients of the signature interpreted in the usual manner. In $\mathcal{M}$, 'Mars' refers to the planet Mars, and the sentence *P(Mars)* is true, i.e. $\mathcal{M} \vDash P(Mars)$. Consider the nontrivial permutation π

that swaps Mars with Phobos. If the story ended here[9], then the truth value of sentences containing the name 'Mars' would enable one to pick out the intended referent, because, obviously $\mathcal{M} \vDash P(Mars)$, but $\mathcal{M}, \pi \nvDash P(Mars)$, where $\mathcal{M}, \pi$ is the model obtained from $\mathcal{M}$ by the π permutation of the domain *M*, without any other adjustments to the predicate *P*. However, the push-through construction induces a reinterpretation of the predicate *P*. In the pushed-through interpretation, *P* would apply to Phobos, and all other planets minus Mars (π leaves all named celestial bodies un-swapped, except for Mars and Phobos). In this permuted model, call it $\mathcal{M}^{\pi}$ in order to distinguish it from $\mathcal{M}, \pi$, the sentence *P(Mars)* is true, as expected. Moreover, by the corollary to the theorem 1, $\mathcal{M}^{\pi}$ attributes to all $\mathcal{L}$-sentences exactly the same truth-values as $\mathcal{M}$.

The permutation argument has the virtue of being easily extendable to other logics, and one such extension to logics with modal operators was, in fact, used by Putnam[10] to argue that *truth-conditions* of sentences, not just truth-values, underdetermine the reference of singular terms[11].

Concluding, the moral of the permutation argument is simply that truth-values and truth-conditions cannot fix the reference of singular terms, and, for our envisaged philosophical position, the question of w*hat, if anything, fixes the reference of terms* remains pertinent and unanswered.

Benacerraf's identification problem

Besides the permutation argument, there is another celebrated argument that poses a problem for the determinacy of reference of mathematical singular terms, although the main target of the argument is the ontological status of the intended referents of mathematical singular terms. To be more precise, the problem addresses the belief that the natural numbers are genuine objects.

The puzzle is properly stated in a set-theoretic foundationalist setting and it focuses on the structure of the natural numbers. Suppose that one endorses the project of reducing the whole mathematics to set theory. Such a project definitely has some attractive philosophical consequences, for example, it unifies the ontology of the whole mathematics, which just by itself is a significant philosophical achievement. In short, suppose that one is committed to the following thesis:

[9] That is, without any other compensatory reinterpretations of the signature $\mathcal{L}$.

[10] (Putnam 1981)

[11] For an elaborate discussion of this version of the permutation argument and two extensions of the push-through construction see (Hale and Wright 1998).

Set-theoretic foundationalist thesis (SF): Set theory is the foundation of mathematics.

As I mentioned at the beginning of this section, suppose that one also embraces the following:

Thesis (IT): The natural numbers are bona-fide objects.

Benacerraf's identification problem is the observation that there is an irreconcilable tension between (SF) and (IT), manifest in the particular case of the natural numbers. In a standard set-theoretic framework, one can reconstruct the natural numbers system in two elementary equivalent (modulo PA-truths), but referentially incompatible ways.

The sketches of the two reconstructions presuppose that the reader is familiar with Peano systems, specified as a triple $<N, 0, s>$, and with basic set-theoretic concepts and techniques.

(A) The first reconstruction is due to Von Neumann[12], and is by far the most popular one among working set-theorists. Concisely, in Von Neumann's reconstruction, we begin with the following definitions, $0 = \emptyset$, and $s_N(x) = x \cup \{x\}$. Consequently, we obtain the following equalities: $0 = \emptyset$, $1 = \{0\} = \{\emptyset\}$, $2 = \{0, 1\} = \{\emptyset, \{\emptyset\}\}$, $3 = \{0, 1, 2\} = \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}$ and so on. Let N_N to be the smallest set containing 0 and closed under the s_N function (the Von Neumann 'successor function'). Now, it can be proved that:

Theorem 2. $<N_N, 0, s_N> \vDash Th(PA)$

(B) The second reconstruction is Zermelo's[13], and basically consists in defining $0 = \emptyset$ and $s_Z(x) = \{x\}$. Obviously, in the zermelian reconstruction, $1 = \{\emptyset\}$, $2 = \{\{\emptyset\}\}$, $3 = \{\{\{\emptyset\}\}\}$ and so on. Let N_Z to be the smallest set containing 0 and closed under the successor function s_Z . Again, it can be proved that:

Theorem 3. $<N_Z, 0, s_Z> \vDash Th(PA)$

By theorems (2) + (3), $<N_N, 0, s_N>$ and $<N_Z, 0, s_Z>$ are elementary equivalent (modulo PA-truths), although referentially distinct: for example, the set corresponding to *2* in N_N is different from the set corresponding to *2* in N_Z; moreover, there are true statements, besides those of PA, which hold in one, but not the other: for example, $3 \in 4$ is true for $<N_N, 0, s_N>$, but not for $<N_Z, 0, s_Z>$.

Benacerraf's identification problem, as it is called, may be stated simply as 'Which set-theoretic objects are the natural numbers?'

[12] Hence the subscript *N* in the subsequent notation.
[13] Hence the subscript *Z* in the subsequent notation.

Enter structuralism. Exit reference

Both the permutation argument and Benacerraf's identification problem received a lot of philosophical attention and scrutiny, and several responses were proposed. For the purpose of this paper, I am going to state briefly and selectively the relevant (for our discussion) standard philosophical countermove to these problems, but before, I will mention a widely entertained consequence of the above arguments with regard to the determinacy of reference of singular terms.

In the particular case of the permutation argument, a widely embraced response[14] was to argue that causal constraints can, and do fix reference. However, in the case of mathematics, there seems to be no such causal constrains, so, the problem of the determinacy of reference holds ground in mathematics. Consequently, the reference of singular terms is taken to be genuinely indeterminate:

> For the objects of pure mathematics, there are no contingencies and no causal connections; so the inscrutability strikes us full force. Inscrutability of reference arises from the fact that our thoughts and practices in using mathematical vocabulary are unable to discern a preference among isomorphic copies of a mathematical structure[15].

The standard countermove, especially in the recent philosophy of mathematics, to the permutation argument and Benacerraf's identification problem was to resort to a structuralist conception of mathematics. Shapiro, Resnik, Hellman, Benacerraf, developed structuralists positions with different ontological, epistemological and semantical flavors. Each of these positions have, however, some common themes, which, for present purposes, are encapsulated as follows: (I) structures are the subject matter of mathematics, and (II) the 'objects'/places in a structure have no other properties except those prescribed by the structure itself[16].

A couple of important consequences follow from (I). From a structuralist point of view, it really does not matter whether two models or two set-theoretic reconstructions of the natural numbers are referentially incompatible, as long as

---

[14] The literature regarding Putnam's argument is impressive, I mention only few authors who developed this line of response: (Lewis 1984) (Devitt 1983), (Field1972), (Field 1975).

[15] (McGee 1997, 38)

[16] As a caveat, one should not think that (II) entails that all structuralists are committed to the existence of mathematical objects, or even structures. The commas in 'objects', and the alternative 'places in a structure' should pinpoint in the direction of a conditional/ontologically neutral reading. However, the structuralists who believe in the existence of mathematical objects, also think that these objects have no internal nature.

they are isomorphic. All that matters is that they have the same 'structural properties'. According to structuralism, then, it makes sense to talk about reference only 'up to isomorphism'[17], thus rendering objects and mathematical reference to objects irrelevant. Discarding objects does not pose a threatening problem insofar as truth is concerned, for structuralists can argue that

> If our thoughts and practices in using the vocabulary distinguish an isomorphism class of equally good candidates for what the terms refer to, this will be enough to establish a determinate truth value for each of the sentences, even though it doesn't pin down the referent of any term. Inscrutability of reference does not imply inscrutability of truth conditions.[18]

Briefly, truth-value determinacy follows from the determinacy of structures, construed as isomorphism types. The thesis that each sentence has a determinate truth-value is known as 'semantic realism', and a theory's semantical capacity to refer to a unique structure is ensured by categoricity. So, what McGee says is that, for structuralist purposes, categoricity is sufficient for ensuring semantic realism[19].

This philosophical vein converges with the practice of mathematics: mathematicians seem to be uninterested in the ontological status and nature of the mathematical objects; they discern structures only up to isomorphism, especially algebraists, and focus on the truth of mathematical statements, rather than other ontological issues. I take these attributes to be marks of structuralism, of course, not exclusively.

Resuming, if structures are the focal point of mathematics, then all the philosophical problems related to objects are irrelevant or unwarranted. All batteries of concerns about the ontological status and nature of mathematical objects, as well as the problem of the determinacy of reference for singular terms that follows from viewing objects as such referential candidates, are benign (if not irrelevant or unwarranted) with respect to what really matters in mathematics, the truths that the structures entertain. Kreisel, as quoted by Dummett[20], aptly described this move as a move from the problem of the existence of mathematical objects to that of mathematical objectivity.

---

[17] A caveat is in order here; there is a version of structuralism, developed by Stewart Shapiro called *ante-rem structuralism*, which zooms in reference up to singular terms – for more details see (Shapiro 1997).

[18] (McGee 1997, 38)

[19] I explored the details and controversies regarding the connection between categoricity and semantic realism in (Ludușan 2015).

[20] "the problem is not the existence of mathematical objects but the objectivity of mathematical statements". (Dummett 1996, xxviii)

Now, it seems that structuralism manages to answer both philosophical problems regarding reference. First, it bypasses Benacerraf's identification problem by insisting that what matters in mathematics are structures, not objects, and secondly it rejects the problem of the indeterminacy of reference of singular terms by rendering it mathematically and philosophically insignificant.

Reference's new structuralist clothes. Enter categoricity

Mathematical structuralism seems to tackle a few philosophically significant problems by shifting the focus from objects to structures. In this way, structures become the bearers of all the mathematically and philosophically relevant properties, such as, for example, the determinacy of the truth-values of sentences, which, as I have mentioned in the previous section, now fully relies on the determinacy of structures. So, a considerable philosophical and mathematical load is placed on structures, which justifies the need for decent ontological, epistemological and semantic explanations regarding structures.

I will disregard the discussions around the ontological status of structures, and, as I have stated in the first section, I will adopt a moderate epistemological position. With this background, I will address a significant semantic problem concerning structures. The problem is the old conundrum about the determinacy of reference, pitched, this time, at the level of theories: *what, if anything, fixes the reference of nonalgebraic/univocal theories*? The reader will recognize this as the central problem. In accordance with the moderation assumption, the explanation cannot invoke innate faculties or intuitions that enable one to pin down the intended reference of such a theory. The explanation, if there is one, has to rely solely on the theory's transparent semantical capacities to determinately refer to a unique structure up to isomorphism. Now, the mathematical way in which one secures that a nonalgebraic/ univocal theory pins down a single structure up to isomorphism, is by proving that the theory is categorical.

> 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[21].

[21] (Meadows 2013, 524)

A theory *T* is categorical if any two models $\mathcal{M}$, $\mathcal{N}$ of *T* are isomorphic, $\mathcal{M} \cong \mathcal{N}$.

In conclusion, in order to fulfill the philosophical promises of structuralism, nonalgebraic theories have to refer determinately to unique structures, which, in turn, is secured by providing categoricity results for each such theory.

Categoricity and first-order logic

Categoricity theorems depend heavily on the logical frameworks in which they are conducted, and effectively this means moving beyond first-order logic. As it is well known, the defining properties of first-order logic make it an unsuitable candidate for proving the categoricity of theories that have models with infinite domains. Model-theoretic results characterizing first-order logic tell us that categoricity in first-order logic can only be obtained for theories with finite models. Suppose that a first-order theory *T* expressed in a language of cardinality λ, $\lambda \geq \aleph_0$, has an infinite model of cardinality κ, $\kappa \geq \lambda$. The upward Löwenheim–Skolem theorem tells us that *T* has models of every cardinality κ', $\kappa' \geq \kappa$ while the downward Löwenheim–Skolem theorem tells us that *T* has a model of cardinality $\aleph_0$. Consequently, the two theorems indicate that such a theory *T* cannot be categorical.

In the case of PA such negative results are reinforced by the use of compactness theorem in order to produce continuum-many pairwise non-isomorphic structures with the same cardinality that satisfy PA[22].

A caveat should be addressed here: of course, we can resort to first-order set theory as the metatheory in which we can prove the categoricity of PA, but the standard argument against this maneuver is that this will push the problem from the categoricity of PA to that of the first-order set theory. First-order set theory has non-isomorphic models, non-standard models, and the categoricity of PA proved in this setting only ensures the uniqueness of the referential structure of PA within each model of set theory, not across different models.

Parsons and Lavine certainly recognize this fact:

> Thus, of the set theory in which we have proved Dedekind's theorem, there will also be nonisomorphic models. And nonisomorphic models of set theory can give rise to nonisomorphic models of arithmetic. Consider now two models M1,

[22] For details regarding the construction of such models see (Kaye 1991) and the responses of Joel David Hamkins and Andreas Blass on the following thread on mathoverflow: https://mathoverflow.net/questions/92099/how-many-models-of-peano-arithmetic-are-isomorphic-to-the-standard-model-and-how.

> and M2 of set theory, and let ω1 and ω2 be their sets of natural numbers. Dedekind's theorem is a theorem of set theory; hence it is true in each of M1, and M2. But what that tells us is that within M1 any structure satisfying [PA2] is isomorphic to ω1 (with the obvious structure), and similarly for M2. But it does not tell us that ω1 is isomorphic to ω2; indeed, since non-well-founded models of set theory can be constructed […], they need not be isomorphic[23].

> Take two models of set theory with nonisomorphic systems of natural numbers, and the proofs of [DCT and quasicategoricity of ZFC] carried out within each one of them only shows that any models of $PA^{(+)}$ or $ZFC^{<(+)}$ within that one must be isomorphic. Those proofs do not show that the natural numbers in the sense one [sic!] of the two models need be isomorphic to those in the other, let alone that the sets in the sense of one of the two models need be isomorphic to those in the other[24].

In short, appeal to categoricity means moving beyond strictly first-order logic.

The mathematics of Dedekind's categoricity theorem

A natural medium for proving categoricity theorems is second-order logic, which has enough resources to categorical characterize not only Peano Arithmetic, but also endless mathematical structures. From now on, I will focus on the structure of the natural numbers and its standard axiomatization encapsulated in Peano systems (see below).

Moving to second-order logic with standard semantics[25], also called full second-order logic, enables us to fix categorically Peano Arithmetic (PA2)[26]. Dedekind already proved[27] in 1888 the categoricity of PA2, formulated in what we today would regard as full second-order logic. In order to have a better grasp of what the categoricity proof presupposes I will present Shapiro's[28] modern reconstruction of Dedekind's original proof restricted[29] to Peano systems.

---

[23] (Parsons 1990: 17)

[24] (Lavine 1999, 65-66)

[25] In second-order logic with standard semantics we allow the second-order quantifiers to range over the powerset of the domain of the first-order variable.

[26] As formulated in second-order logic, of course.

[27] (Dedekind 1901)

[28] (Shapiro 1997, 82-83)

[29] This restriction is for simplicity purposes, the rest of the operations and relations of Peano Arithmetic can easily be defined in Peano systems and proved to obey their standard Peano axioms.

Definition 1. A Peano system is a triple P = <N, 0, s > which satisfies the following conditions (PA2):

i) $\forall x\, \neg(0 = s(x))$
ii) $\forall x \forall y\, ((s(x) = s(y)) \rightarrow (x = y))$
iii) $\forall X (X0 \wedge \forall x (Xx \rightarrow Xs(x)) \rightarrow \forall x Xx)$, where $X \subseteq N$.

Note that the only significant change between a Peano system formulated in first-order logic and one formulated in second-order logic is the induction axiom.

Theorem 4. Dedekind categoricity theorem (DCT): If $P_A \vDash PA_2$, and $P_B \vDash PA_2$, then $P_A \cong P_B$.

Proof: Let $P_A = \langle N_A, 0_A, s_A \rangle$ and $P_B = \langle N_B, 0_B, s_B \rangle$ be two Peano systems. Define

$$F = \bigcap\{I \subseteq N_A \times N_B \nless 0_A,\ 0_B > \in I \ and\ if < x, y > \in I\ ,$$
$$then\ < s_A(x), s_B(\ y) > \in I\ \}$$

It is clear that F is not empty, for the Cartesian product $N_A \times N_B$ itself would constitute such a set, and, further, by $\prod{}^1_1$ comprehension such a set exists.

Now, let's prove that F is an isomorphism between N_A and N_B. We divide the proof in two parts. First, we show (A) that F is a bijective function, and then (B) that it is isomorphic.

(A) For F to be a bijective function $F: N_A \rightarrow N_B$, we must first show that it is a function, i.e. to show that

(1) $dom(F) = N_A$.
(2) If $\langle x, y \rangle \in F$ and $\langle x, z \rangle \in F$, then $y = z$.

(1) We begin by defining the domain of F,

$$dom(F) = \{x \in N_A \,/\, \exists y \in N_B \text{ such that } \langle x, y \rangle \in F\}.$$

By induction on $dom(F)$ we will prove that $dom(F) = N_A$. *Base case*: obviously, $0_A \in dom(F)$ [for there is 0_B such that $< 0_A,\ 0_B > \in F$]. *Induction step*: assume that $x \in dom(F)$; accordingly, there is an element $y \in N_B$ such that $\langle x, y \rangle \in F$. It follows, by the definition of F, that $\langle s_A(x), s_B(y) \rangle \in F$, which, by the definition of $dom(F)$, let us conclude that $s_A(x) \in dom(F)$. By induction, we get that $dom(F) = N_A$.

(2) As in the previous case, the proof is by induction. Define the set[30]:

$$X = \{x \in N_A \;/\; \exists y \in N_B \; such\ that < x, y > \in F \; and \; \forall z \in N_B, \; if < x, z > \in F, then\ y = z\}$$

Base case: Suppose that $0_A \notin X$. By definition, $<0_A, 0_B> \in F$, so, if $0_A \notin X$, there must be a $z \neq 0_B$ such that $<0_A, z> \in F$. Consider the set $Y = F - \{<0_A, z>\}$; clearly $Y \subset F$. We will prove that $F \subseteq Y$. (I). Obviously, $<0_A, 0_B> \in Y$ [since $<0_A, 0_B> \in F$ and $z \neq 0_B$]. (II). If $<x, y> \in Y$, then $<s(x), s(y)> \in Y$ [since by *i*) of *definition 1*, $s(x) \neq 0_A$]. By (I) and (II), $F \subseteq Y$, contradicting the fact $Y \subset F$. In conclusion, $0_A \in X$.

Induction step: From the supposition that $x \in X$, we'll prove that $s(x) \in X$. So, assume that $x \in X$. This means that there is a unique y such that $<x, y> \in X$. By the definition of F, $<s(x), s(y)> \in F$, so, if we suppose that $s(x) \notin X$, then there is $z \neq s(y)$ such that $<s(x), z> \in F$. Now, consider the set $Z = F - \{<s(x), z>\}$. As in the previous case, we will prove that $F \subseteq Z$, thus contradicting the fact $Z \subset F$. (III). $<0_A, 0_B> \in Z$ [again, $<0_A, 0_B> \in F$ and by *i*) of *definition 1*, $s(x) \neq 0_A$]. (IV). Assume that $<a, b> \in Z$. Then $<a, b> \in F$. By the definition of F, $<s(a), s(b)> \in F$. Now, there are two possibilities: either $a = x$, or $a \neq x$. If $a \neq x$, then, by *ii*) of *definition 1*, $s(a) \neq s(x)$, so $<s(a), s(b)> \in Z$. If $a = x$, then, since $x \in X$, there is a unique y such that $<x, y> \in X$, so $b = y$. But, by the assumption that $s(x) \notin X$, there is $z \neq s(y) = s(b)$ such that $<s(x), z> \in F$, so $<s(a), s(b)> \in Z$.

By (III) and (IV), $F \subseteq Z$ which contradicts the fact that $Z \subset F$. In conclusion, if $x \in X$, then $s(x) \in X$.

(1) and (2) assure us that F is a function, $F: N_A \rightarrow N_B$. Now, it remains to prove that F is bijective. This is done, as in the previous poof, in two steps, proving that:

(3) F is injective.

(4) F is surjective.

(3) Consider the set $X = \{x \in N_A \;/\; \forall y \in N_A \; ((F(x) = F(y)) \rightarrow (x = y))\}$

The proof is by induction on N_A along the same lines as in the first proof given above.

(4) Consider the set $X = \{ y \in N_B \;/\; \exists x \in N_A \wedge F(x) = y\}$

The proof is by induction on N_B along the same lines as in the first proof given above.

(B) The isomorphism of F follows directly from its definition.

---

[30] This set corresponds to the property that characterizes a function i.e. that there is just one element from the codomain corresponding to each element from the domain, or, as we expressed this condition, if $<x, y>$ and $<x, z>$, then $y = z$.

The philosophy of Dedekind's categoricity theorem

Dedekind's categoricity theorem, as conducted in second-order logic (SOL), is riddled with worries about its philosophical significance. The literature on the relevance of DCT is impressive and still in the making. I will only mention several philosophical worries that I discerned, emphasizing on the one that will concern us further.

There are ontological worries, based on Quine's criterion of ontological commitment[31], that adopting full PA2 means committing not only to the existence of numbers, but of arbitrary sets of numbers, in virtue of the semantics of the second-order quantifiers[32].

There are epistemological worries, first, about the infinitary set-theoretic presuppositions implied in the adoption of full PA2, and secondly, that commitment to full SOL presupposes the determinacy and intelligibility of the powerset operation, which is problematic[33].

The worry that interests us is that of the relevance of DCT insofar as it establishes the referential determinacy of PA, i.e. as it responds to the central problem.

DCT can provide a definitive answer to the central problem if the background theory in which it is conducted, SOL, is determinate. As it is well known, SOL has two distinct types of model-theoretic semantics: the full semantics, or standard semantics, in which the proof of the theorem was carried, and the Henkin semantics. Without delving too much into the technicalities and subtleties of the differences between the two types of semantics, I will present the significant differences between them, first, in terms of the fundamental feature that distinguishes the two approaches, and secondly, in terms of the difference of metatheoretical properties of SOL equipped with the two semantics.

The standard model-theoretic semantics presupposes that the second-order variables X^n, $n \geq 1$, range over the entire powerset $\wp(M^n)$, $n \geq 1$, of the corresponding domain M^n. In contrast, in Henkin semantics, this presupposition is relaxed by considering the domain of quantification for second-order variables X^n, $n \geq 1$, a subset M^n_{rel} of the corresponding powerset $\wp(M^n)$, of M^n, $M^n_{rel} \subseteq \wp(M^n)$. As one can observe, Henkin semantics are more general than standard semantics; in fact, standard semantics is just a limit case of Henkin semantics, precisely when $M^n_{rel} = \wp(M^n)$, for all, $M^n_{rel,}$ $n \geq 1$.

---

[31] "A theory is committed to those and only those entities to which the bound variables of the theory must be capable of referring in order that the affirmations made in the theory be true" – (Quine 1948, 33).

[32] See (McGee 1997).

[33] See (Weston 1976), (Field 2001, 352-354), (Field 1994).

As a caveat, let us note that although there is just one standard semantics, there are numerous incompatible Henkin semantics.

The central feature that distinguishes the two model-theoretic semantics, namely, the domain of the second-order quantifiers has a significant impact on the defining properties of SOL with full models or Henkin models. In standard second-order logic, the three defining properties of first-order logic, *compactness, Löwenheim–Skolem*, and *completeness*, fail, while SOL with Henkin models is characterized by all three properties. It is for this reason that Henkin models are closer to first-order logic than full SOL.

Obviously, the defining properties of SOL with Henkin models disrupt the appeal to DCT as a solution for the central problem. The situation is similar to that described three sections above, concerning DCT as proved in first-order set theory. There, we emphasized that such a result establishes the categoricity of PA only within, but not across different models of set theory, i.e. DCT is relevant modulo models of first-order set theory.

The considerations that led to such a diagnosis namely that by the three defining properties of first-order logic, *compactness, Löwenheim–Skolem*, and *completeness*, any theory couched in first-order logic (so, in particular set theory) has unintended models, apply to Henkin models also.

Now, the simple availability of two types of semantics for SOL should not be a problem for establishing the referential determinacy of PA2, if one can provide an explanation with moderate epistemological credentials as to why full models are preferable to Henkin models. But, unfortunately, it is doubtful that such an explanation is even possible. Remember, the moderate cannot appeal to any idiosyncratic capacities that would tie the mind to full models instead of Henkin models; all her available resources are restricted to theories and their semantics. So, the moderate has to explain her preference of full models by introducing more mathematical theory. However, this move is highly problematic, firstly, because the further we move from the referential determinacy of PA2, to that of the metatheoretical background in which it was proved, and to that of the metametatheoretical background and so on, the more philosophically dubious the supposed determinacies become. Secondly, such a move is vulnerable to the initial objection: the introduced explanatory mathematical theory is subjected to the same unintended reinterpretations as the previous (meta)theories were. The latter line of arguing is Putnam's *just more theory maneuver*. Speaking of Putnam, he concisely described the problem with the philosophical relevance of categoricity theorems in SOL:

> the 'intended' interpretation of the second-order formalism is not fixed by the use of the formalism (the formalism itself admits so-called 'Henkin models' […]), and it becomes necessary to attribute to the mind special powers of 'grasping second-order notions'[34]

Internal categoricity

The previous section highlighted that the use of DCT as a solution for the central problem is bound to the determinacy of the semantics or of the models of the metatheoretical background, which, in turn, is bound to the determinacy of higher order mathematical concepts and/or theories, and such a regress seems unbreakable. It is for this reason that in the '90's a somehow radical solution[35] was proposed: to reconstruct categoricity theorems in the purely 'syntactic'/deductive environment of the metatheory, thus bypassing any semantic notions. That means, for example, to reconstruct DCT as a 'pure' theorem in SOL, and refrain from engaging in semantic considerations about DCT or SOL. Such a move amounts to a certain confinement of the categoricity results within the metatheoretical framework, hence the name 'internal categoricity'. Button & Walsh describe the manifesto of this internalization movement like this:

> ***The internalist manifesto***. For philosophical purposes, the metamathematics of second-order theories should not involve semantic ascent. Instead, it should be undertaken within the logical framework of very theories under investigation. Our slogan is: METAMATHEMATICS WITHOUT SEMANTICS![36]

The plan for the rest of the paper is to focus on one such particular form of internalism, that of Shaughan Lavine, as it is articulated in *Skolem was wrong*. Here is how Lavine presents the rationale of his internalism:

> In order to escape the apparent impasse [that DCT is dependent upon the semantics of the metatheoretical background], it will be necessary to formulate and prove categoricity theorems that do not make use of a background set

[34] (Putnam1980, 481)

[35] The main figures of this movement are Charles Parsons, Van McGee, Stewart Shapiro, and Shaughan Lavine.

[36] (Button and Walsh 2018, 227)

> theory. The bulk of the rest of the book will be devoted to solving that problem, but the key idea is simple: No one ever actually compares set-theoretic universes; we compare theories of sets, which are syntactic, not set-theoretic, entities. When we ascend to the level of language and ask the question, "When are two theories syntactically theories of isomorphic structures"?, we shall see that that is a question that has perfectly clear purely syntactic sufficient condition for a positive answer that is free of any need for a background set theory[37].

In this context, I will discuss the potential philosophical uses of the internal categoricity of arithmetic related to the central problem. The philosophical achievements of this particular form of internalism are, however, applicable to internalism itself.

The interesting discussion is whether the internal categoricity results can solve in a satisfactory manner the central problem or something akin to the central problem. I say 'something akin' to the central problem, because internal categoricity does not seem to have any bearings on the central problem, as I formulated it: the central problem has a semantic character, regarding the relationship between PA and its intended referent, while the central feature of internal categoricity has a 'syntactical' character, and couples syntactic entities. Nevertheless, I will argue that the proponents of internalism advanced such arguments, thinking that internal categoricity can solve the central problem. I will argue that such a move is unwarranted. Next, I will go on to consider whether internal categoricity can establish the determinacy of PA's internal-structures. Again, the result is negative.

Schematic theories

The logical medium in which Shaughan Lavine proves an internalized version of DCT is the full schematic theory of Peano Arithmetic, $PA^{(+)}$. I should mention that all three major figures of internalism use schematic induction and comprehension to the effect of proving an internalized version of DCT, so the subsequent analysis applies in a large degree to all versions of internalism. The apparatus of the schematic theories that Lavine employs was theorized and developed to a different end by Solomon Feferman[38]. I will begin sketching the idea behind $PA^{(+)}$ by distinguishing several PA theories (schematic and ordinary). To this end, I will define in general terms the composition of ordinary and schematic theories, and then, using this definitional template I will discern and focus exclusively on different types of PA.

[37] (Lavine 1999, 39)

[38] (Feferman 1991)

Definition 2. An axiomatic theory $S = <\mathcal{L}_S, Ax_S, Rule_S>$ *is taken to be specified by three sets: the signature* $\mathcal{L}_S$ *of S, the special axioms* Ax_S *of S (those in addition to the logical axioms – Log* $Ax(\mathcal{L}_S)$*), and the special rules* $Rule_S$ *of S (those in addition to MP and GEN*[39]*).*

Definition 3. An ordinary (axiomatic) theory S is one for which $Rule_S$ *is empty i.e.* $S = <\mathcal{L}_S, Ax_S>$

Definition 4. By a schematic (axiomatic) theory S = S(P), we mean one for which

(i) $\mathcal{L}_{S(P)}$ *is of the form* $\mathcal{L}_S \cup \{P\}$*, for some base language* $\mathcal{L}_S$*, and*

(ii) $Rule_{S(P)}$ *consists of the single rule:*

$\mathcal{L}_{S(P)}$*-Subst: From* $\Phi(P)$ *infer* $\Phi(\varphi)$*, in symbols,* $\Phi(P)/\Phi(\varphi)$*, for any* $\Phi(P)$ *and* φ *in* $Form_{\mathcal{L}S(P)}$*, where* $Form_{\mathcal{L}S(P)}$ *= the set of all* $\mathcal{L}_{S(P)}$*-wffs.*

Let us distinguish three types of PA according to these definitions.

Let PA_0 be the Peano Arithmetic base theory defined by the usual axioms that state that 0 is a first element, the successor function is injective, and defines addition and multiplication:

PA_0*– (base theory)*

(i) $\forall x\, \neg(0 = s(x))$
(ii) $\forall x\ \forall y\, ((s(x) = s(y)) \rightarrow (x = y))$
(iii) $\forall x\, ((x + 0) = x)$
(iv) $\forall x\ \forall y\, ((x + s(y)) = s(x + y))$
(v) $\forall x\, ((x \cdot 0) = 0)$
(vi) $\forall x\ \forall y\, ((x \cdot s(y)) = (x \cdot y)+x))$

The discerning factor between several PA theories is the schematic induction axiom, formulated using a schematic variable symbol *P*:

$Ind(P)$: $(P(0) \wedge \forall x(P(x) \rightarrow P(s(x)))) \rightarrow \forall x(P(x))$

The induction axiom is accompanied by a corresponding substitution rule, which will define a hierarchy of theories. Its basic template is:

$\mathcal{L}$*-subst:* $Ind(P)/Ind(\hat{x}\varphi(x))$*, where* $Ind(\hat{x}\varphi(x))$ *indicates the result of substituting* $\varphi(t) \in \mathcal{L}$ *for each occurrence of* $P(t)$ *in* $Ind(P)$*, renaming bound variables of* $Ind(P)$ *and* φ *in order to prevent collisions with the free variables of t.*

[39] Short for *modus ponens* and *generalization*.

The difference between ordinary PA, schematic $PA^{(s)}$ and full schematic $PA^{(+)}$ manifests itself as soon as we consider extensions of signatures, say an extension $\mathcal{L}$ of $\mathcal{L}_{PA0}$, $\mathcal{L} \supseteq \mathcal{L}_{PA0}$:

Ordinary PA in an extension $\mathcal{L} \supseteq \mathcal{L}_{PA0}$ is: $<\mathcal{L}_{PA0}, PA_0 \cup \{Ind(\hat{x}\varphi(x)), \varphi(x) \in \mathcal{L}_{PA0}\}>$[40]

Ordinary schematic theory $PA^{(s)}$ in an extension $\mathcal{L} \supseteq \mathcal{L}_{PA0}$: $<\mathcal{L}_{PA0(P)}, PA_0 \cup \{Ind(P)\}, \mathcal{L}_{PA0}$-*subst* rule$>$

Full schematic theory $PA^{(+)}$ in an extension $\mathcal{L} \supseteq \mathcal{L}_{PA0}$: $<\mathcal{L}_{PA0(P)}, PA_0 \cup\{Ind(P)\}, \mathcal{L}_{(P)}$-*subst* rule$>$

As one can observe, the full schematic theory $PA^{(+)}$ is the only one that allows derivations of instances of induction containing open wffs from the extended signature $\varphi(x) \in \mathcal{L}$. The ordinary PA_0 contains an infinite number of induction axioms, one for each $\varphi(x) \in \mathcal{L}_{PA0}$, and is immune to extensions of the signature, while $PA^{(s)}$ contains a single induction axiom, from which one can infer only instances containing open wffs in the old signature $\varphi(x) \in \mathcal{L}_{PA0}$. It is no surprise, then, that *PA* and $PA^{(s)}$ are deductively equivalent, although different in elegance (*PA* has an infinite number of axioms, while $PA^{(s)}$ only a finite number). So, the preference of *PA* over $PA^{(s)}$ is a matter of aesthetics. Also, let us note that the induction schema with its associated generous substitution rule from $PA^{(+)}$ behaves as an open-ended induction schema that can be defined as follows:

$$(Ind_{1/2}): (\varphi(0) \wedge \forall x(\varphi(x) \rightarrow \varphi(s(x)))) \rightarrow \forall x(\varphi(x)), \text{ for all } \varphi(x) \in \mathcal{L} \text{ and all } \mathcal{L} \supseteq \mathcal{L}_{PA}$$

In a similar fashion, one can adapt the second-order comprehension schema:

$$CS_2: \exists X\ \forall x(X(x) \leftrightarrow \varphi(x)), \text{ for all } \varphi(x) \in \mathcal{L} \text{ such that } X \notin FV(\varphi)$$

to obtain:

$$CS(P): \exists X\ \forall x(X(x) \leftrightarrow P(x)), \mathcal{L}_{(P)}\text{-subst rule, with the proviso that for all } \varphi(x) \in \mathcal{L}_{(P)} \text{ and all } \mathcal{L} \supseteq \mathcal{L}_{0(P)}, X \notin FV(\varphi)$$

In the literature around internal categoricity, this open-ended character of schemas is the focal point of discussions and critiques. Note that all instances of $(Ind_{1/2})$, or for that matter of Ind(P) in $PA^{(+)}$ are first-order. This is relevant for our moderation assumption.

[40] As one can easily observe, there is no $\mathcal{L}$-*subst* rule.

Now, Parsons[41], McGee[42], and Lavine[43] all argue for adopting a full schematic perspective as a way of bypassing all the philosophical shortcomings of PA2. The reasons for adopting a schematic perspective are both philosophical and technical. On the philosophical side, McGee and Lavine argue that schematic induction is the only one that accords with arithmetical practices.

> Note that insofar as the theory $PA^{(+)}$ differs from the theory $PA^{(s)}$, it is the former theory that is a superior codification of our informal intentions concerning arithmetic: we intend induction to apply to any predicate of numbers, not just those definable in elementary number theory. No one ever hesitated to apply induction in the context, for example, of analytic number theory, as they should have done if our intentions were better codified by the theory $PA^{(s)}$, a theory that fails to foreclose the intuitively absurd possibility of our coming to define a noninductive predicate of the natural numbers, that is, a predicate W of the natural numbers such that $W(0) \land \forall x(W(x) \to W(Sx) \land \exists x \neg W(x)$ holds on the natural numbers.[44]

Note that what Lavine is implying here is that only the full schematic $PA^{(+)}$ can prohibit the definition and incorporation of such 'an intuitively absurd' predicate, thus, that only $PA^{(+)}$ can characterize the standard model of arithmetic. This will become relevant for the argument that I will develop after the next section. McGee also insists on the virtues of open-ended schemas, arguing that in a rational reconstruction of how we learn arithmetic, a fundamental step, if not the fundamental step, is precisely mastering $(Ind_{1/2})$[45].

Now, the technical reason. The fundamental technical reason for adopting $PA^{(+)}$ is that it enables the addition of new predicates with appropriate axioms in the schema of induction. In fact, precisely this type of extensions motivate the adoption of $PA^{(+)}$. Suppose that one is trying to define by primitive recursion a function (and prove the legitimacy of such a definition), say, natural number exponentiation on a group. Then, one can do that in $PA^{(+)}$ in a series of steps:

(i) enlarge the signature so that it consists of the signature of Peano Arithmetic, PA, Group theory, GT, and two predicates, U and U' corresponding to the 'intended' domains of the two theories: $\mathcal{L} = \mathcal{L}_{PA} \cup \mathcal{L}_{GT} \cup \{U, U'\}$

[41] (Parsons 1990)
[42] (McGee 1997)
[43] (Lavine 1999)
[44] (Lavine 1999, 15-16)
[45] (McGee 1997)

(ii) relativize the quantifiers, constants, and function symbols to the predicates U and U'. The resulting theories are symbolized as PA^{+U} and $G^{U'}$.

(iii) add the axioms governing the new relation symbol *E* (For readability reasons I skipped the relativization procedure):

$\forall x\, E(0, x, e),$

$\forall n\ \forall x\ \forall y(E(n, x, y) \rightarrow E(s(n), x, y \otimes x)$.

With this device active,

> The proof that E is a function can now be carried out in the familiar way in the theory that is the union of PA^{+U} and $G^{U'}$, and the definition of E. There is no need for any additional background theory, and the success of this hybrid theory, which requires the full induction schema, is compelling evidence that $PA^{(+)U}$ is the appropriate formalization of arithmetic: surely no one will try to claim that natural number exponentiation on groups is intrinsically set theoretic[46].

The internal categoricity of $PA^{(+)U}$

In the course of analysis of Lavine's detailed argument for internal categoricity, I am going to follow closely his presentation. This will be helpful not only for the accuracy of the analysis, but also for also for pointing precisely my critiques.

Essentially, Lavine project is to prove internal categoricity in the same manner as the one just described: enlarge the signature to include a copy of the signature of PA^{+U}, relativize the quantifiers, constants, functions to their corresponding 'domains' say U, U', merge the two theories so that we end up with a theory that is $PA^{+U} \cup PA^{+U'}$, and then add a relation *I* that defines an internal or syntactical isomorphism between PA^{+U} and $PA^{+U'}$.

The addition of $PA^{+U'}$ doesn't raise any consistency or satisfiability problems, for it can be easily specified in an extension by definitions of PA^{+U} thus:

$\forall x(U'x \leftrightarrow Ux)$
$0' = 0$
$\forall x(s'(x) \leftrightarrow s(x))$
$\forall x\, \forall y\ ((x +' y) = (x + y))$
$\forall x\, \forall y\ ((x \cdot' y) = (x \cdot y))$

46 (Lavine 1999, 20)

Generally, the expansion of a schematic theory with a new relation is not uncontroversial, as Lavine himself notes[47]. Consequently, a fundamental challenge for the schematic approach to internal categoricity is precisely to formulate general acceptable conditions for such an expansion. Informally, Lavine's proposal is that the conditions of acceptability have to be such that the addition of new relations leaves the domain of the theory intact:

> It is therefore natural to ask when the addition of a theory A to a full schematic theory $T^{(+)}$ adds a new relation without changing the domain. I shall call such an addition an *acceptable addition*.

Technically, the conditions of an acceptable addition that Lavine proposes imply extending the schematic theory to include a form of inflationary fixed-point logic, particularly, inflationary fixed-point logic that includes monotone fixed-point operators defined by positive formulas. In this way, one obtains a minimal extension of first-order logic that allow closure under inductive definitions.

With this setting in place, Lavine defines *I* to be:

$$\forall x \forall x' (I(x, x') \leftrightarrow \forall y{<}x\ \exists y'{<'}x'(I(y, y') \wedge \forall y'{<'}x'\ \exists y{<}x)\ I(y, y')).$$

As Lavine remarks, 'the definition of I is a definition of a fixed point of an operation defined by a positive formula, and that it is therefore an acceptable definition'[48].

Now, in order for a relation *I* to be a *syntactic or internal isomorphism* between two $PA^{(+)}$ systems it has to satisfy the following conditions[49]:

1. $(\forall x)(U(x) \rightarrow \exists\, y\ (U'(y) \wedge (\forall z)(U'(z) \rightarrow (I(x, z) \leftrightarrow z = y))))$
(I is a function from U to U')
2. $(\forall x)(U'(x) \rightarrow \exists\ (U(y) \wedge (\forall z)(U'(z) \rightarrow (I(z, x) \leftrightarrow z = y))))$
(I is one-to-one and onto from U to U')
3. $I(0, 0')$,
4. $(\forall x)(\forall x')(\forall y)(\forall y')(U(x) \wedge U'(x') \wedge U(y) \wedge U'(y') \wedge I(x, x') \wedge I(y, y') \rightarrow s(x) = y \leftrightarrow s'(x') = y')$
5. $(\forall x)(\forall x')(\forall y)(\forall y')(\forall z)(\forall z')(U(x) \wedge U'(x') \wedge U(y) \wedge U'(y') \wedge U(z) \wedge U'(z') \wedge I(x, x') \wedge I(y, y') \wedge I(z, z') \rightarrow x + y = z \leftrightarrow x' +' y' = z')$
6. $(\forall x)(\forall x')(\forall y)(\forall y')(\forall z)(\forall z')(U(x) \wedge U'(x') \wedge U(y) \wedge U'(y') \wedge U(z) \wedge U'(z') \wedge I(x, x') \wedge I(y, y') \wedge I(z, z') \rightarrow x \cdot y = z \leftrightarrow x' \cdot' y' = z')$.

---

47 (Lavine 1999, 45-46)
48 (Lavine 1999, 57)
49 As Lavine defines them.

Conditions 3-6 define the homomorphism property of *I*, and, taken together, the conditions 1-6 define properly speaking the internal or syntactical isomorphism between $PA^{(+)U}$ and $PA^{(+)U'}$.

Now everything is in place to prove:

Theorem 5. (internal-DCT): Let T be the theory with some language $\mathcal{L}^2$ that is the union of the theory $PA^{(+)U}$ and the theory $PA^{(+)U'}$. Then one can acceptably define a new binary relation I such that I is a syntactic isomorphism.

The proof consists, basically, in mimicking DCT's proof outlined in a previous section.

This theorem is theorem 4.6 in Lavine's manuscript. He also proves using the same recipe the quasicategoricity of ZFC; that is his theorem 4.7.

The philosophy of the internal categoricity of $PA^{(+)U}$

So, the question is, 'What is the philosophical relevance of (this version of) internal categoricity?' What does Lavine expect to get from such a result? This section focuses on the analysis of Lavine's arguments for the philosophical significance of his version of internal categoricity.

The fundamental benefit of using this particular internalized version of DCT is, as Lavine emphasizes throughout his manuscript, that it involves only an uninterpreted first-order background:

> Our actual theorems 4.6 and 4.7 are theorems of the first-order predicate calculus that do not presuppose any set-theoretic notions. Unlike [DCT and Zermelo's quasicategoricty theorem], which had proofs that relied on notions of a background set theory, theorems 4.6 and 4.7 may—and indeed must—be viewed as theorems of an uninterpreted background first-order logic, one introduced without benefit of a background set theory.

This, in itself, represents an important philosophical achievement, especially in the context of a moderate epistemic position. However, this is only the start of the real philosophical challenge, which is to show if and how this particular version of internal categoricity provides a solution to the central problem. Now, as I have mentioned previously, it is a type confusion[50] to think that an internal categoricity result can provide a solution to a problem expressed in 'external' semantic terms.

[50] As (Button and Walsh 2018, 226) put it.

However, Lavine's manuscript is an elaborate argument for the legitimacy of such a connection: he articulates a general form of our 'central problem', and constructs a detailed argument for solving it, based on internalizing DCT and Zermelo's quasicategoricity theorem; as a side note, the title of the manuscript should be a significant giveaway. So, I will first argue generically that such a connection is unwarranted, then I will reconstruct Lavine's main argument for solving the central problem via the categoricity of $PA^{(+)U}$ and show that his argument fails.

Let us begin by noting that internalized-DCT is about the behavior of the predicates U, U', constant symbols, 0, 0' and function symbols, *s, s*', +, +' etc, which all are syntactic entities. Any such system (U, 0, s, +, ·) behaving according to the Peano axioms constitutes the syntactical counterpart of a PA structure; accordingly, I will call it an internal-structure. Consequently, what internalized-DCT effectively shows is that any two internal-structures of $PA^{(+)}$ must behave in the same arithmetical way. Or, in other terms, internal-DCT shows that one cannot accept two different internal-structures inside the same $PA^{(+)}$.

Now, how is this going to help solving the central problem? Well, one line of thought is that once we have proved internalized-DCT, we can bestow a semantical dimension to $PA^{(+)}$, prove that the only model of $PA^{(+)}$ is the standard model of arithmetic and, thus, solve the central problem. Unfortunately, this strategy does not work, for as soon as one engages semantical attributes, the old problems of the semantical relativity of the metatheory come back. In this context, what internal-DCT establishes is, at best, categoricity within $PA^{(+)}$ models, not across such models. Lavine seems to engage in such considerations[51], for example when he qualifies internal-DCT as a stronger theorem than its (external) counterpart –proved as theorem 3.3 in his manuscript – or when he explicitly says that "[o]nce we have proved it, we shall be able to use theorem 4.6 in place of theorem 3.3, thereby avoiding the use of a background set theory"[52]. One can infer legitimately that the internal versions of the categoricity theorems are stronger than the external ones if one engages in semantical considerations (but not exclusively – see below another interpretation): as I have just mentioned, internal-DCT establishes categoricity within Henkin models or $PA^{(+)}$ models, so, in particular, it establishes the categoricity in 'full' models also, for full models are limit-cases of Henkin models.

[51]"Theorems 4.6 and 4.7 should not be confused with the weaker theorems that look just like them and are proved in verbatim the same way that presuppose a background set theory". (Lavine 1999, 64)

[52] (Lavine 1999, 51, fn 7)

Lavine's argumentative strategy, however, is different. The way in which he connects internal-categoricity to the central problem, if I understand him correctly, is the following.

The fundamental assumption that his argument is based on is the neutral, prior, and independent character of first-order logic. Explicitly, this assumption presupposes that we have a cogent understanding of first-order logic, prior to any semantic considerations, and that first-order logic is unproblematic[53]. From this assumption it follows that understanding first-order logic precedes any set-theoretic or model-theoretic perspectives, which are always an afterthought. It is decisive for Lavine's argument that we should carefully distinguish between 'pure', stronger results, obtained within first-order logic by deductive means alone, and their 'weak' counterparts, polluted by a set-theoretic or model-theoretic interpretation. Whenever first-order logic is embedded in a semantic environment, the pure results become contaminated, and, thus, weak – because of their dependence upon the semantic environment. It is in this way that I construe Lavine's remarks about the strength of the internal-categoricity theorems – as indicative of the distinctness and strength of the first-order results. As exegetical evidence, I will quote Lavine's eighth footnote:

> In that it is central to my solution of Skolem's problem that the categoricity theorems are outside any prior model or system of set theory and can therefore be applied to any of them, I am implicitly endorsing Wright's "diagnosis": "there is an informal set-theoretic result . . . which we can prove about this model, which is not to be identified with the corresponding result within the system when the latter is interpreted in terms of this model . . . " [Wri85, p. 132]. The result to which Wright is referring is Cantor's theorem[54].

I further interpret this prevalence and distinctness of the 'pure' results obtained in first-order logic as playing a pivotal role in the development of our various mathematical conceptions[55]. The subsequent argument that I am going to develop against Lavine's strategy for solving the central problem does not essentially depend on this latter interpretation; nevertheless, I will assess whether the interpretation

[53] "In providing a solution to the central problem I may therefore presume that there is a clear antecedent understanding of first-order logic and that first-order logic is free of unwarranted presuppositions". (Lavine 1999, 7)

[54] (Lavine 1999, 65, fn 8)

[55] In this, I take it, he follows Crispin Wright's proposal that Cantor's diagonal argument "plays a role in the formation of our conception of what the intended interpretation of set theory is. Its role is […] to lead the determination of an inchoate concept of set in a particular direction". (Wright 1985, 132-133)

can save Lavine's strategy, and show that the argument so construed is sound, but points to a different conclusion, that is, it misses its intended target, the central problem. Resuming, all the results obtained by means of first-order logic alone have a cogent character, with universal applicability, that should be sharply distinguished by the same results when interpreted in semantical terms. Allow me to emphasize this argumentative joint: because of the autonomous, cogent, understanding of first-order logic, results proven in such a setting 'can therefore be applied to any [model or system]'[56], and should not be confused with the same results after adding a semantical dimension. In particular, any first-order result concerning theories of arithmetic precedes and subverts the same result interpreted in model-theoretic/set theoretic terms. Consequently, the categoricity of arithmetic, established in first-order logic, takes antecedence to any model-based (post)interpretation.

Now, the second assumption of Lavine's argument, and I cannot overstate its importance, is that referential indeterminacy is always a byproduct of model-based considerations. It is only when we add a set-theoretical/model-theoretical dimension to a schematic theory $T^{(+)}$, that the indeterminacy of reference for $T^{(+)}$ strikes.

Thus, the argument goes, all worries regarding the referential determinacy of arithmetic, which arise exclusively from semantic considerations, dissipate, for the first-order categoricity of arithmetic is prior to any such considerations, and, as such, takes antecedence. The referential indeterminacy of arithmetic is a byproduct of embedding the arithmetical theory in different models, and, as such, is insolubly tied with the semantical perspective. But the internal-DCT, being a first-order result, undercuts the ulterior, model-based, problem of the referential determinacy of $PA^{(+)}$. As one can easily observe, the only missing piece of the argument is a first-order proof of DCT. And this is exactly what internal/syntactical categoricity of $PA^{(+)}$ is supposes to provide. This, I believe, is an accurate gloss of Lavine's argument:

> Our actual theorems 4.6 and 4.7 are theorems of the first-order predicate calculus that do not presuppose any set-theoretic notions. […] Since the theorems are prior to any choice of any system of natural numbers or of any theory of sets, they can be used to compare any proposed systems and theories whatever. The theorems thus guarantee that if we even regard it as coherent to raise the possibility that either $PA^{(+)}$ or $ZFC^{<(+)}$ could fail to characterize its subject matter, and therefore grant that it is coherent to contemplate multiple copies of $PA^{(+)}$ or $ZFC^{<(+)}$, that alone is enough to prove that the requisite characterization has been achieved. I am inclined to take the argument just

[56] (Lavine 1999, 65, fn 8)

> given at face value: I think that it does show that Skolem was wrong— $PA^{(+)}$ and $ZFC^{<(+)}$ characterize the natural numbers and the sets up to isomorphism, and do so in a non-question-begging way.

Now, I will raise two distinct types of critiques to Lavine's argument. The first type of critique regards the justifications provided for the purely internal character of DCT, and the second type regards the soundness of Lavine's argument, even conceding that internal-DCT is a 'pure' first-order result.

I will begin with the former critique. To this end, let me summarize Lavine's argument that internal-DCT is a first-order result as is developed through the manuscript. (1) He defines from the very beginning a model-theoretic semantics for $PA^{(+)}$ and proves that the only model of $PA^{(+)}$ is the standard model of arithmetic, acknowledging that the philosophical relevance of the theorem is dependent upon the semantics' set-theoretic assumptions. Consequently, he proceeds to reconstruct the proof of the theorem in a set-theoretic-free environment. (2) To this end, he engages in setting the conditions of acceptable additions (of relations and theories) to $PA^{(+)}$ so that (3) he can define the relation *I* and show that it is acceptable, and, finally, (4) prove that *I* is a syntactic isomorphism, i.e. establishing internal-DCT.

Now, depending on one's philosophical views, all steps have weak spots, but I am going to concentrate on the first two, that are more relevant for Lavine's particular version of internalism. First, Lavine's model-theoretic sketch of the proof of the categoricity of $PA^{(+)}$ is extremely dubious[57]. The proof has two parts, the first one consists in observing that the standard model of PA is a model of $PA^{(+)}$, and the second consists in proving by *reductio* that $PA^{(+)}$ cannot have a nonstandard model $\mathcal{M}$. To that effect, Lavine presupposes that $\mathcal{M}$ is a nonstandard model of $PA^{(+)}$ and then considers an expansion $\mathcal{M}[{}^{I}{}_{N}]$ for a $\mathcal{L}_{PA(+)} \cup \{I\}$ signature, where *I* is interpreted as the standard part $N \subset M$ of the domain M of $\mathcal{M}$. Of course, applying Ind(P) to *I*, by the corresponding substitution rule, yields that $M \subseteq N$. Contradiction.

This proof, I must confess, confounds me, for if one has at her disposal a predicate *I* which determinately refers to the standard part *N*, then why the detour through schematic theories and/or induction in order to establish the categoricity of PA? One can just add the predicate *I* with its intended interpretation to PA1 and prove in whatever metatheory she prefers the categoricity of arithmetic, by rejecting all nonstandard models. The point is that once one has at her disposal the means for referring to the standard model of arithmetic, one also has free of charge the referential determinacy of PA, so the argument based on the proof begs the question. It is like including in the logical vocabulary the predicate *N* with its intended

[57] This line of critique is similar to the one presented by (Field 2001, 355) in another context.

interpretation; of course, this maneuver will single out the standard model of arithmetic, but nothing substantial was proved, you already had the referential determinacy of PA.

The second point of the critique is two-folded. First, there is the issue of the justification of the choices that led to the particular formulation of the conditions of an acceptable addition to $PA^{(+)}$, and then that of their accurate statement or definition. The driving idea that underlines the choices for what constitutes an acceptable addition to $PA^{(+)}$ is to singularize the standard model as the unique referential structure of $PA^{(+)}$. This explains why Lavine considers informally that an acceptable addition of a relation should preserve the domain of the model. This is also why he specifies formally[58] that sets of universal formulas (i.e. formulas that are of the form $\forall\bar{x}\phi(\bar{x})$) are acceptable additions to a schematic theory: for universal formulas are preserved under substructures, and, obviously, the standard model is the smallest model of all possible models, i.e. is the initial segment of all models. So, it is clear that all the choices involved in setting the conditions of an acceptable addition to $PA^{(+)}$ were *a priori* biased in favor of the standard model. Again, a case of begging the question. And, again, it has less to do with schematic theories and more to do with the model-theoretic ways in which we beefed up schematic theories for a particular goal.

The model-theoretic means employed in specifying the conditions of an acceptable addition are the subject of my second critique. As one can easily observe, in all instances, the formulation of the conditions of an acceptable addition is set-theoretic and the proofs involved are model theoretic. Lavine acknowledges this as a shortcoming of his approach and solves it by appeal to another schematic theory, PAPR (Peano arithmetic with primitive recursion), which combines PA with primitive recursive arithmetic. Now, the assessment of that solution constitutes the topic of another paper, and I am not going to add anything to that discussion here. The issue is a fragile joint of Lavine's argument, for it is extremely problematic to maintain the 'pure' syntactic first-order character of internal-DCT, fundamental to the argument, yet, in proving the result to rely extensively and heavily on model-theoretic or set theoretic specifications. Here is how Lavine summarizes the discussion:

> For present purposes, [the criteria for determining what can be added to full schematic theories] are to be regarded merely as *ex post facto* justifications, and perhaps generalizations, of principles concerning acceptable additions that we take as basic, intuitive, and well-established parts of mathematical practice:

[58] in his theorem 4.3

> We can add any universal theory consistent with arithmetic to arithmetic, and we can add fixed points of operations defined by positive formulas to any full schematic theory. […] The notion of acceptability is an intuitive one that cannot be made mathematically precise without set-theoretic apparatus to which I am not entitled at this stage of the argument, but all I shall use in the rest of this book is that the definition of a fixed point of an operation defined by a positive formula is an acceptable addition to any full schematic theory.[59]

So, Lavine argues that the model-theoretic/set-theoretic infused formulations and proofs of the conditions of an acceptable addition are to be seen as mathematically rigorous articulations of intuitive principles of mathematical practice, but this is far from being a sound or even convincing argument. I must confess, I find it difficult to base the intuition behind the standard model of arithmetic on the intuition behind the acceptability of adding to a schematic theory T^+ sets of universal formulas consistent with the base theory T, or the intuition behind the model theoretic/set theoretic devices that allow formulations of inductive definitions. The history of mathematics shows pretty clearly that the structure of the natural numbers is the source of our concepts of induction and recursion, not the other way around. However, he deploys another dodging maneuver: even though he is not entitled to set-theoretic resources in this stage of the argument, he can in the last resort, prove the quasicategoricity of ZCF^+ and then safely use the set-theoretic apparatus needed for the formulations and proofs of the conditions. For example, he states that after proving the internal quasicategoricity of set theory,

> one can just introduce the other intended structures using familiar second-order axiomatizations, with the second-order quantifiers now explained without circularity in terms of the set theory that has already been introduced. Thus, set theory is the central case.[60]

This is somehow ironic. Lavine's main argument of the manuscript is that the categoricity of arithmetic can be proved independently of any set theoretical background. Nevertheless, it seems that in his own project, in order to prove the categoricity of arithmetic, one has to establish first the categoricity of set theory. Besides the irony, the point of my critique is that once one has proved the quasicategoricity of set theory, the categoricity of arithmetic follows immediately, but that has nothing to do with schematic theories, nor with the pure syntactical first-order-logic character of the proof. Once we have established the quasicategoricity

[59] (Lavine 1999, 54-55)
[60] (Lavine 1999, 40)

of set theory, we can prove DCT easily in any adequate background we like, including in a set theoretical background. This will erase the difference between the latter ‘traditional’ proof and Lavine’s internal one: for proving the categoricity of arithmetic, both take the detour through the quasicategoricity of set theory. Nothing significant has been achieved. Before concluding this type of critique, let me point to another difficulty related to the last remarks: the conditions of an acceptable addition are used in the proof of the quasicategoricity of ZCF^{+}, and there one can conspicuously perceive their circularity, for there isn’t any other theory whose categoricity once established allows the use of the resources in discussion.

The second type of argument regards the relevance of Lavine’s argument granting that he successfully proved the internal categoricity of $PA^{(+)}$ in a non-question begging way, using only first-order logic resources. Well, if my gloss of Lavine’s argument is accurate, then, I will argue that Lavine’s particular solution of the central problem fails.

I start by reiterating the fundamental assumptions of Lavine’s argument, 1) the prior, autonomous, semantic-free character of first-order logic, and 2) that referential indeterminacy is a byproduct of ulterior, model-based considerations. To this skeleton, add the meat of producing a first-order proof of the internal categoricity of $PA^{(+)}$. The result is that the internal-categoricity of $PA^{(+)}$ takes precedence, so that the indeterminacy-inducing interpretations derived from embedding $PA^{(+)}$ in a set theoretic or a model-theoretic environment have no effect. It is this last part of the argument that I find highly problematic, so much so, I will argue, that it leaves the central problem unanswered. I begin my argument constructing a scenario involving a schematic theory $T^{(+)}$ and a model of set theory. Consider an unaware inhabitant of such a model that accepts $T^{(+)}$. She endorses Lavine’s assumptions about the prior and autonomous character of first-order logic, and of the ulterior model-based referential indeterminacy of $T^{(+)}$. She proves the internal categoricity of $T^{(+)}$, thus assuring herself that $T^{(+)}$ manages to refer to a unique intended structure. However, in light of the model-based considerations that proliferate deviant, nonstandard models and structures, she would like to expose the referential mechanism by which $T^{(+)}$ pins down its referent. Note that she is not driven by skepticism regarding the referential determinacy of $T^{(+)}$, she firmly believes that $T^{(+)}$ manages to successfully refer to its unique intended structure. She just wants to explain how $T^{(+)}$ accomplishes this. Of course, the mechanism of reference should not appeal to enigmatic faculties of the mind, but be restricted to moderate-approved resources, i.e. those that the theory and its semantics consist of. The moderate means by which $T^{(+)}$ selects the intended structure from all deviant referential competitors consists in utilizing the first-order proof of the internal-categoricity of $T^{(+)}$. Now, the central problem shows its teeth, for the internal categoricity of $T^{(+)}$ is consistent with there being many models with non-isomorphic structures as perfectly legitimate referential candidates. She cannot resort to the first-order character of the proof of internal categoricity, for this is also consistent with the existence of models of set theory containing non-standard models of $T^{(+)}$. That is, nothing in Lavine's assumptions or argument precludes the possibility that she is living on such a set-theoretic multiverse.

Mathematically, in such a set-theoretic multiverse, everybody could easily establish internal-DCT for $PA^{(+)}$, thus, establish the isomorphism of all the structures inside their models corresponding to $PA^{(+)}$, without establishing external, 'true' isomorphism. In fact, this is the distinctive mark of internal categoricity as Jouko Väänänen[61] defines it.

So, although she buys everything Lavines argues, she still cannot exclude, by any referential means offered by the internal categoricity in first-order logic, the set-theoretic possibility of there being more than one up to isomorphism structure as the referent of $T^{(+)}$. Note, again, that she does not doubt that $T^{(+)}$ refers to the intended structure, and that all other concocted structures are deviant, non-intentional ones. She is not motivated by skepticism. She just wants to clarify the referential means by which $T^{(+)}$ accomplishes this selection task. It is at this point that she acknowledges that all the available referential mechanisms fail to glue $T^{(+)}$ to its intended referent. The reason, again, is that the available referential mechanisms are consistent with a scenario in which $T^{(+)}$ refers to a concocted nonstandard structure, even though she recognizes the artificiality of the nonstandard structure and its dependence on the standard one in its construction.

Let me conclude my critique by presenting the gist of the argument in other terms. One can illustrate the point of my argument using Kripke's[62] Wittgensteinian paradox involving plus-quus, or Goodman's[63] green-grue puzzle. I will choose the former. Suppose that someone has learned in a standard, normal, way, how to perform additions. She is confident that her use of 'plus' or '+' denotes the standard mathematical function of addition. This stage of my illustration corresponds to learning that $T^{(+)}$ has a first-order internal categoricity proof by a corresponding character. Returning to Kripke's example, imagine that by an encounter with a bizarre skeptic, our heroine learns about the deviant referential candidate of '+', name it 'quus'[64], a function that agrees with addition up to the largest number used in her past computations, but deviating form addition for all other larger numbers. At this stage of the illustration, the corresponding character from my argument

[61] (Väänänen 2012, 98-99), (Väänänen and Wang 2015, 125)
[62] (Kripke 1982)
[63] (Goodman 1955)
[64] I follow Kripke's baptism of the deviant function.

learns about the existence of deviant, nonstandard, but adequate referential candidates for $T^{(+)}$. Returning to Kripke's example, the non-skeptical problem that she begins to contemplate is what is the referential mechanism by which '+' denotes the addition function and not the quus function. Again, she is not skeptical, she doesn't believe that the referent of '+' is quus, she just wants to provide an explanation for the referential relation between '+' and addition. But all the moderate-available means at her disposal could not pick addition as the sole referent of '+'. There is nothing in the usage of '+' that could discern between addition and quus. Similarly, all the referential moderate means– the internal categoricity of $T^{(+)}$ in first-order logic – available to the corresponding character are consistent with many non-isomorphic referents of $T^{(+)}$. The point is that the afterthought concerning referential determinacy always comes back to haunt the pre-established harmony of internal categoricity.

Now, I have to tie one more loose end. Remember, in the interpretation that I proposed above any result obtained in first-order logic informs and permeates our conceptions and our subsequent considerations, because first-order logic is this prior, autonomous, unproblematic, devoid of any semantical assumptions, medium. I don't believe that resorting to such an interpretation solves the conundrum of the referential determinacy of arithmetic. In fact, it misses the target, and leaves the conundrum posed by the central problem unanswered. What Lavine accomplished, at best, is to indicate that the natural number structure is a presupposition, and not a philosophical thesis to be argued for. In mathematical practice, the standard model is regarded as a presupposition, $\mathbb{N}$ just is the structure for which induction holds for all $X \subseteq N$. This, of course, is a resolution by stipulation, and in that quality, it needs no further justification. Well, if this is so, then what Lavine's argument shows is that our conception of arithmetic is from the very beginning bound by certain constraints to admit just one structure. That may be so, but then, how can such an argument solve the central problem? The central problem is about how the resources of a theory of arithmetic can pin down the structure of the natural numbers, not about how our conception of arithmetic is so shaped that the uniqueness of the natural numbers is already built in.

In conclusion, Lavine's detailed and sophisticated argument misses its intended target, the central problem. First, the argument fails to adequately respond to the challenge raised by the central problem. Secondly, the argument is riddled with philosophical question-begging or relevance difficulties, which, I argued, are insurmountable.

REFERENCES

Benacerraf, Paul. 1965. "What numbers could not be." *Philosophical Review* 74 (1): 47-73.

Button, Tim. 2013. *The limits of realism.* Oxford: Oxford University Press.

Button, Tim, and Sean Walsh. 2018. *Philosophy and model theory.* Oxford: Oxford University Press.

Button, Tim, and Sean Walsh. 2016. "Structure and categoricity: Determinacy of reference and truth value in the philosophy of mathematics." *Philosophia Mathematica* 24 (3): 283-307.

Dedekind, Richard. 1901. *Essays on the theory of numbers.* Translated by Woodruff W Beman. Chicago: Open Court.

Devitt, Michael. 1983. "Realism and the Renegade Putnam: A Critical Study of Meaning and the Moral Sciences." *Noûs* 72 (2): 291–301.

Dummett, Michael. 2014. *Origins of analytic philosophy.* London, New York: Bloomsbury Academic.

—. 1996. *Truth and other enigmas.* Cambridge, MA: Harvard University Press.

Feferman, Solomon. 1991. "Reflecting on Incompleteness." *The Journal of Symbolic Logic* 56 (1): 1-49.

Field, Hartry. 1994. *Are our logical and mathematical concepts highly indeterminate?* 19: 391-429.

Field, Hartry. 1975. "Conventionalism and instrumentalism in semantics." *Noûs* 9 (4): 375-405.

Field, Hartry. 1972. "Tarski's Theory of Truth." *The Journal of Philosophy* 69 (13): 347-375.

—. 2001. *Truth and the absence of fact.* Oxford: Oxford University Press.

Frege, Gottlob. 1953. *Foundations of Arithmetic.* Oxford: Blackwell.

Goodman, Nelson. 1955. *Fact, fiction, and forecast.* Cambridge, MA: Harvard University Press.

Hale, Bob, and Crispin Wright. 1998. "Putnam's model-theoretic argument against metaphysical realism." In *A companion to the philosophy of language*, edited by Bob Hale and Crispin Wright, 427-458. Oxford: Blackwell.

Kaye, Richard. 1991. *Models of Peano arithmetic.* Oxford: Clarendon.

Kripke, Saul. 1982. *Wittgenstein on rules and private language. An elementary exposition.* Cambridge MA: Harvard University Press.

Lavine, Shaughan. 1999. *Skolem was wrong.* manuscript.

Lewis, David. 1984. "Putnam's Paradox." *Australasian Journal of Philosophy* 62 (3): 221–236.

Linnebo, Øystein. 2017. *Philosophy of mathematics.* New Jersey, London: Princeton University Press.

Ludușan, Adrian. 2015. "Categoricity, open-ended schemas and Peano arithmetic." *Logos & Episteme* VI (3): 313-332.

McGee, Vann. 1993. "A semantic conception of truth?" *Philosophical Topics* 21 (2): 83-111.

McGee, Vann. 1997. "How We Learn Mathematical Language." *The Philosophical Review* 106 (1): 35-68.
Meadows, Toby. 2013. "What can a categoricity theorem tell us?" *The review of symbolic logic* 6 (3): 524-544.
Parsons, Charles. 1990. "The Uniqueness of the Natural Numbers." *Iyyun* 39: 13-44.
Putnam, Hilary. 1980. "Models and reality." *Journal of Symbolic Logic* 45 (3): 464–482.
—. 1981. *Reason, truth and history.* Cambridge: Cambridge University Press.
Quine, Willard van Orman. 1948. "On What There Is." *The Review of Metaphysics* 2 (1): 21-38.
Shapiro, Stewart. 1997. *Philosophy of mathematics: Structure and ontology.* New York: Oxford University Press.
Väänänen, Jouko. 2012. "Second order logic or set theory?" *The bulletin of symbolic logic* 18 (1): 91-121.
Väänänen, Jouko, and Tong Wang. 2015. "Internal Categoricity in Arithmetic and Set Theory." *Notre Dame Journal of Formal Logic* 56 (1): 121-134.
Weston, Thomas. 1976. "Kreisel, the continuum hypothesis and second order set theory." *Journal of Philosophical Logic* 5 (2): 281-298.
Wright, Crispin. 1985. "Skolem and the skeptic." *Proceedings of the Aristotelian Society, Supplementary Volumes* 59: 117-137.