Does Set Theory Really Ground Arithmetic Truth?

Alfredo Roque Freire

Center for Research and Development in Mathematics and Applications (CIDMA), Department of Mathematics, University of Aveiro, 3810-193 Aveiro, Portugal; alfredo.roque.freire@ua.pt

Abstract: We consider the foundational relation between arithmetic and set theory. Our goal is to criticize the construction of standard arithmetic models as providing grounds for arithmetic truth. Our method is to emphasize the incomplete picture of both theories and to treat models as their syntactical counterparts. Insisting on the incomplete picture will allow us to argue in favor of the revisability of the standard-model interpretation. We start briefly characterizing the expansion of arithmetic 'truth' provided by the interpretation in a set theory. Interpreted versions of an arithmetic theory into set theories generally have more theorems than the original. This theorem expansion is not complete however. Using this, the set theoretic multiversalist concludes that there are multiple legitimate standard models of arithmetic. We suggest a different multiversalist conclusion: while there is a single arithmetic structure, its interpretation in each universe may vary or even not be possible. We continue by defining the coordination problem. We consider two independent communities of mathematicians responsible for deciding over new axioms for ZF and PA. How likely are they to be coordinated regarding PA's interpretation in ZF? We prove that it is possible to have extensions of PA not interpretable in a given set theory ST. We further show that the number of extensions of arithmetic is uncountable, while interpretable extensions in ST are countable. We finally argue that this fact suggests that coordination can only work if it is assumed from the start.

Keywords: foundations of mathematics; arithmetic; set theory

1. Overview

In this article, we study the idea of reducing arithmetic to set theory as a strategy for grounding arithmetic truth. The method of reduction we have in mind is interpretation. We say that a theory T_1 is interpreted in a theory T_2, when there is a uniform mapping of theorems of T_1 in theorems of T_2. This mapping should preserve the boolean structure and bound quantifiers of T_1 in a definable class of T_2. We will next indicate how model constructions can be understood as the establishment of interpretations between theories.

In what follows, we assume that mathematical structures exist independently of our ability to completely describe them. It is common practice, however, to refer to models as fully formed entities for which one can assert whether any formula is valid. This is generally done with Gödel-Tarski method within a set-theoretic metatheory. The fact that one can decide whether any formula φ is satisfied by a model M is simply given by the axiom of excluded middle in the metatheory. Although this strategy may help us to understand model-theoretic properties, it will not necessarily help us to concretely determine which are the valid formulas. For example, considering the standard model N of arithmetic built in a ZF metatheory, we indeed know that ψ = "twin prime conjecture" is satisfied or not by the model. But that "N satisfies ψ" can still be unprovable from the point of view of ZF.

This is the reason why we will consider models via their syntactical representation through interpretations. Understanding models in this way will allow us to distinguish more precisely the undecidable instances of the form "N satisfies ψ" in the chosen metatheory. Structures should not be treated as syntactical constructions nevertheless. One may refer to a set-theoretic structure V as a platonic collection of objects; and due to our limited

knowledge, the notion of satisfaction in V is vaguely defined. We can, however, define a precise notion of knowledge about satisfaction by fixing a set theoretic theory ST:

$$\text{We know that } V \vDash \varphi \text{ if, and only if, } ST \vdash \varphi \tag{1}$$

Now, each model definable in a given base model $V \vDash ST$ can be said be to the result of bounding the elements of V to a given interpretation I (this will be define precisely in the Section 2 with respect to arithmetic). By doing so, we can keep in mind our limited knowledge of models. Since, if $\mathcal{M}$ is definable in V (i.e., $\mathcal{M} = I^V$) and we do not know any other information about V other than that it satisfies ST, then

$$\text{We know } \mathcal{M} \vDash \varphi \text{ if, and only if, } ST \vdash \varphi^I \tag{2}$$

Furthermore, we investigate the grounding relation represented by interpreting PA in ZF. Notably, if one considers the standard interpretation of PA in ZF to be correct, then it expands what one known to be arithmetically true—i.e., many independent formulas in PA become theorems as we see them in ZF through the interpretation. But even though we expect that interpretations of PA in ZF expand knowledge of arithmetic truth, ZF does not completely decide on arithmetical formulas. Indeed, for every interpretation I of arithmetic in a recursive extension S of ZF, there is an arithmetical formula that S does not decide under this interpretation. At any stage in the development of ZF (a recursive extension), the concept of arithmetical truth will still be open. Some arithmetic formulas will be undecidable under the interpretation in any recursively extended set theory. Hence, it is possible to build two structures satisfying the set theory that disagree about the truth value of an arithmetic formula.

Taking a multiversalist view of set theory, Hamkins and others (see [1–3]) use a similar basis to advance a pluralist view of arithmetic. In [1], for example, Hamkins and Yang show that there are models of ZF that agree about what the standard model of arithmetic is and yet disagree about what is valid in the standard model. This (and other results) suggests that there are alternative models of arithmetic. In this article we use a different approach. Assuming we have good reasons to say that there is a unique arithmetic intended structure while maintaining a multiversalist view of set theory (this view is suggested by Koellner in [4]), we argue that the standard interpretation should be taken as revisable. Furthermore, it may happen that the structure of arithmetic is not definable in some set-theoretic universes.

It is due to this phenomena that we consider what we call the coordination problem: consider that there are two groups of mathematicians responsible for deciding over new axioms. The first will decide over axioms for arithmetic and the second for a set theory. How should we consider the relation between the two groups? Note that if we consider that the arithmetic group should conform to any development provided by the set theory group, it becomes hard to see in what sense the interpretation of arithmetic into set theory has any foundational role. This framework is indistinguishable from simply taking arithmetic to live in set theory.

If, however, the interpretation of arithmetic in set theory has a meaningful foundational role, it is important to consider the possibility of the coordination between the two theories to break. Is it possible that an extension of arithmetic not to be interpretable in any extension of a set theory? We show in Theorem 2 that for any extension A of PA and any extension S of ZF, there is an extension A^+ that is not interpretable in S. But, how likely is it to be the case? We will further show in Theorem 3 that there are uncountable consistent extensions of a recursive A, while only a countable number of interpretations of arithmetic in any set theory. For this reason, the addition of axioms to set theory and arithmetic by the two groups would preserve the interpretability relation only if coordination is assumed. We further conclude that this perfect coordination would empty the reductivist foundational role of set theory to arithmetic. Finally, we briefly explore an alternative foundational role that would avoid this problem.

2. The Standard Model of Arithmetic

The strategy of offering set-theoretical models to describe objects of a theory comes from the work of Tarski, Mostowsky, and Robinson in the 1940s [5]. Ever since this date, mathematicians and philosophers often resort to this strategy. It is generally accepted that once we start talking about models, we put aside the formal aspects of the mathematical subject and start talking about its objects and truths. Nevertheless, because of Gödel's incompleteness theorem and Löwenhein-Skolem theorem, there is no formal way to fix the model of any recursive extension of Peano arithmetic. It is impossible to say that the only model that satisfies our descriptions of arithmetic is the intended model, no matter how extensively we describe it. Still, using a set-theoretical apparatus, we can describe the intended model as $N = \langle \omega, +, ., 0, s \rangle$ (called standard model). We can then show that a set theory like ZF is expressive enough to define a truth predicate for this interpretation.

The literature on this subject generally presents two approaches for fixing the standard model: (i) one should offer extra-logical (or second-order) reasons for choosing N from the myriad possible models for arithmetic; (ii) one should abandon the model-theoretical construction and find other ways to ground arithmetic truth. A renewed version of (ii) can be seen in Gabbay's defense of a new kind of formalism [6]; Moreover, others may abandon a privileged emphasis on N, because we must focus on mathematical practice (Ferreirós [7]) or because we must commit ourselves to a realistic multiverse (Hamkins [8]). Still, differences of opinion are more common as to how and why we should follow project (i). Those like Williamson [9] argue for metaphysical reasons for setting N, others like Maddy [10], Quine [11] or Putnam [12] advocate ways to naturalize the reasons for N. Finally, a recent approach by Rodrigo Freire grounds N in mathematical practice using a normative basis in place of the Platonist commitment to N [13].

The question of the adequacy of N is often overlooked. Though one may find a vast literature on non-standard models of arithmetic, these are generally regarded as 'deviant' or not intended. They are indeed existing structures that satisfy an arithmetic theory, but they are not the one true model of arithmetic. The assumption behind this is that if something is a model of arithmetic, then it is N. We may not know why this is the intended model or even deny that such a model exists, but the conformity to N is hardly questioned. However, presenting N as an object without further consideration is a category mistake. Notably, a similar category mistake would be to say that 'there have been two sun revolutions since so and so'. The phrase 'two sun revolutions' is used as quantity of time, even though it describes a movement in reference to the Sun. Hence, the statement would be a category mistake unless, for instance, an implicit reference to Earth and not Mars is assumed. Precisely stated, N is an interpretation of PA in the language of membership. It represents therefore a construction of objects for arithmetic in terms of objects of a given set theory. Hence, it is only when we fix the objects for a set theory that the objects expressed in the construction N gain life.

For any given model of set theory $V \vDash ZF$, an arithmetic interpretation I can be understood as a procedure for obtaining a model $\mathcal{N}$ for PA. The model $\mathcal{N} = \langle Obj, +^{\mathcal{N}}, .^{\mathcal{N}}, 0^{\mathcal{N}}, s^N \rangle$ is a set in the vaguely defined V with the appropriate meaning for the arithmetic symbols $+$ (sum), . (multiplication), 0 (constant zero) and s (successor function). The model $\mathcal{N}$ is built from the interpretation $I = \langle U, f_+, f_., f_s, zero \rangle$. The elements of I are formulas in the language of ZF: U is a formula with one free variable, f_+ and $f_.$ are formulas with three free variables, f_s is a formula with two free variables and $zero$ is a formula with one free variable. It is then necessary to prove in V that the formulas in $f_+(x, y, z), f_.(x, y, z), f_s(x, z)$ indeed represent functions with respect to the variable z and that $zero(x)$ is satisfied by a unique element in V. With these ingredients, we explicitly build in V the model $\mathcal{N}$:

1. $Obj = \{x \in V \mid V \vDash U(x)\}$.
2. $0^{\mathcal{N}} = a$ such that $V \vDash zero(a)$.
3. $+^{\mathcal{N}} = \{\langle x, y, z \rangle \mid x, y, z \in Obj \text{ and } V \vDash f_+(x, y, z)\}$.
4. $.^{\mathcal{N}} = \{\langle x, y, z \rangle \mid x, y, z \in Obj \text{ and } V \vDash f_.(x, y, z)\}$.

5. $s^{\mathcal{N}} = \{\langle x, y \rangle \mid x, y \in Obj \text{ and } V \vDash f_s(x, y)\}$.

We may refer to the model obtained from V using I as I^V. In this context, the standard interpretation $N = \langle U, f_+, f_., f_s, Zero \rangle$ is the case where $U(x)$ expresses in set theory 'x is an finite ordinal', $f_+(x, y, z)$ expresses 'z is the ordinal sum of x and y', $f_.(x, y, z)$ expresses 'z is the ordinal product of x and y', $f_s(x, z)$ expresses 'z is the ordinal successor of x' and $Zero(x)$ expresses 'x is the empty set'. We can then obtain that, independently on the choice of the base model $V \vDash ZF$, the model $N^V \vDash PA$.

Syntactically, we may use I to produce a uniform strategy for mapping formulas in the language of arithmetic $L(PA)$ to formulas in the language of set theory $L(ZF)$. As we assumed that $f_+(x, y, z)$ is a function in V, we may use, for simplicity, a function-like language defining $F_+(x, y)$ in ZF as "the z such that $f(x, y, z)$". Similarly, we define $F_.$, F_s and $Zero$. For every arithmetic formula φ, we define the partially interpreted formula φ^{I^*} by:

1. replacing in every atomic subformula of φ occurrence of the form $x + y$, $x.y$, $s(x)$, for $F_+(x, y))$, $F_.(x, y)$, $F_s(x)$ respectively;
2. replacing every occurrence of $\forall x(\psi)$ for $\forall x(U(x) \to \psi)$;
3. replacing every occurrence of $\exists x(\psi)$ for $\exists x(U(x) \land \psi)$;

If φ has free variables $x_1, x_2, \ldots, x_n$, the interpreted formula φ^I is defined as $(U(x_1) \land U(x_2) \land \ldots \land U(x_n)) \to \varphi^{I^*}$. With this, we can now say that ZF interprets PA with the standard interpretation N since every $\varphi \in PA$ is such that $ZF \vdash \varphi^N$.

Our idea is to insist on the incomplete picture of the set-theoretical representation of arithmetic. All we know about the vaguely defined V is that it is based on an incomplete theory ZF. Therefore, the picture of arithmetic obtained from reducing PA to V by N is also incomplete. In this context, it is worth paying attention to precisely what is decidedly valid in the standard construction with the syntactic notion $ZF \vdash \varphi^N$. If one only commits to the validity of the axioms of a set theory ST, the undecidable formulas in ST of the form φ^I are precisely the arithmetic formulas that one does not know if they are valid or not.

So to what are we committing in the case where we say that N is the standard model of arithmetic? As we will discuss in the next section, it depends on what is the chosen model V. It is in fact showing that the standard model has many representations (even isomorphic, though with different truth predicates), that Hamkins and Yang in [1] proposes a pluralist view of arithmetic. Notably, however, they still fix the standard interpretation–evaluating this interpretation in different structures of set theory. It seems like the single construction for the intended model of arithmetic is based on the idea condensed in the sentence: 'no matter which model of set theory one is assuming, the model of arithmetic would be given by N'. Indeed, the picture provided by the literature is that of **revisable** *truth for set theory and arithmetic*–but **unrevisable** reduction of arithmetic in set theory. In the next sections, we argue that to take the standard model to have a foundational role, one should assume the interpretation to be revisable. For now, we consider the characterization of arithmetic in set theory in more details.

Foundational Characterization of PA in ZF

Being I an interpretation of arithmetic in a set theory S, we call the set $A_I^S = \{\varphi \in L(PA) \mid S \vdash \varphi^I\}$ the expansion of arithmetic truth under the interpretation. Indeed some undecidable formulas φ of PA are 'true' in the standard model ($ZF \vdash \varphi^N$). This is the case for the Gödel formula, Goodstein's theorem and many others arithmetic results. We will thus consider more broadly the question of expansion of arithmetic truth from interpretations in set theories.

Given that I is an interpretation of an arithmetic theory A in a set theory S and $Th(A) = \{\varphi \mid A \vdash \varphi\}$, we expect to have $Th(A) \subsetneq A_I^S \subsetneq$ *Arithmetic truth*, as we see in Figure 1:
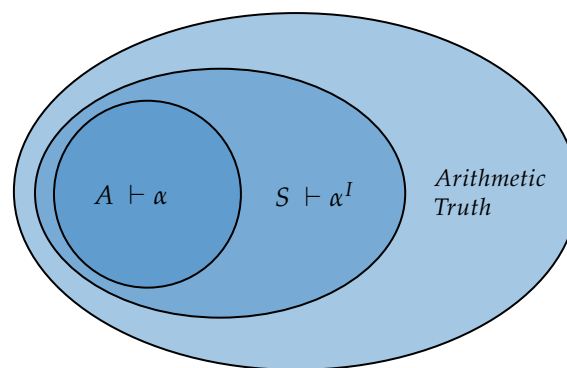
Figure 1. Expansion of validity under interpretation.

The reason for the expansion $Th(A) \subsetneq A_I^S$ is that, in the usual case, one expects to build a set-size model of arithmetic. Consequently, a consistency predicate for A should be expressed and proved in S. Consider the base case of PA and ZF with the standard interpretation N. Assuming a model V for ZF, we can build a model N^V satisfying PA. We then know that there are many valid formulas in N^V that are not provable in PA. The most immediate example is the consistency predicate $Con(PA)$; in fact, we know that the predicate is valid in N^V or, in other words, that $ZF \vdash (Con(PA))^I$.

Of course, from a given recursive extension S of ZF, one may simply choose the recursive arithmetic theory corresponding to the theorems in ST about the standard interpretation (i.e., A_N^S). But this is to *put the cart before the horse*, being open to the evaluation of extra valid formulas with respect to the current axiomatization of arithmetic (e.g., $\varphi \in A_N^S$ but such that φ is not proved in the current axiomatization of arithmetic) is a fundamental aspect in this study. In addition, there are important recent results that show fundamental mismatches between arithmetic and set theory. In fact, no subtheory of any extension of ZF is bi-interpretable with any extension of PA. This is a simple consequence of a theorem by Enayat and independently discovered by Hamkins and me: two different extensions of ZF can never be bi-interpretable [14–16] (the direct proof is done in the dissertation ([17] pp. 150–152). Together with the bi-interpretation of finite set theory and Peano arithmetic, the result follows. Hence, in order to obtain a set theory equivalent to PA we must add an axiom that contradicts ZF. Similarly, no compatible (with ZF) collection of set-theoretic concepts can perfectly mirror an axiomatization of arithmetic that extends PA.

We also note that the characterization of the foundation relation by theorem expansion relates to the mathematical practice. With the discovery of the Gödel's incompleteness theorem in [18], some resistance to the result was argued in the sense that the obtained undecidable statement had little mathematical meaning. Later on, Goodstein [19] proved that there are fast growing functions (called Goodstein sequences) that cannot be proved to be total in PA. The existence of these sequences is directly connected to the traditional Hydra problem, and thus it bears a clear mathematical meaning (see Caicedo's "Goodstein's function" [20]). Thus the question of foundation arises as to whether the interpretation of PA in set theory answers a significant arithmetical problem that was not possibly addressed by the axiomatization. And this is indeed the case as we consider Goodstein sequences.

Notably, important results in number theory have recently become so loaded with complicated techniques that mathematicians have begun to question whether the proofs extrapolated Peano's axioms. This is the case of Fermat's last theorem and the weak Goldbach conjecture, proved respectively by Andrew Wiles [21,22] and by Harald Helfgott [23]. This type of question is akin to the program of reverse mathematics and has drawn the attention of mathematicians like Harvey Friedman. However, the validity of those theorems, whether they depend or not on more axioms than PA, is hardly questioned. The choice is not commonly to add axioms to PA, but to investigate arithmetic truths in a theory that expands the extension of theorems. One is not however simply doing 'finite ordinal set theory' when dealing more loosely with arithmetic's axiomatization, as these 'stronger than PA' assumption should correspond to number theorists' intuitions about natural numbers.

We have discussed that interpretations of arithmetic in set theories generally expand what may be taken to be arithmetical truth ($Th(A) \subsetneqq A_I^S$). Yet this expansion is not necessarily complete (A_I^S = arithmetic truth). A confusion in this regard is due to the idea that model constructions in set theories offer venues for defining truth for interpreted theories. Each interpretation I represents the appropriate model construction such that the grounding set theoretic model V can provide the notion of satisfaction $I^V \vDash \varphi$ for any formula. Eventually, we would have that for any formula γ, either $V^I \vDash \gamma$ or $V^I \vDash \neg\gamma$. However, as we have already discussed, a more syntactical approach makes it clear that this is simply the expression of the *excluded middle*. Indeed, "either $V^I \vDash \gamma$ or $V^I \vDash \neg\gamma$" should be syntactically represented by

$$ZF \vdash \gamma^I \vee \neg\gamma^I \tag{3}$$

Instead, what is really wanted is a notion like

$$ZF \vdash \gamma^I \text{ or } ZF \vdash \neg\gamma^I \tag{4}$$

As we suppose a base model V for ZF, we are at hand with an interpretation for ZF itself or with a loosely defined model. In this case, the notion of truth in a model is represented by "either $I^V \vDash \gamma$ or $I^V \vDash \neg\gamma$". However, if our supposition of a model V is not informed by any specific information other than $V \vDash ZF$, the interpretation works simply as the identity. Therefore, we return to the problem of establishing a notion as in (4).

However, Equation (4) is not achievable for any recursive extension of ZF. For a given interpretation I of arithmetic in a recursive extension S of ZF, there will be formulas of $L(PA)$ that are undecidable about arithmetic in S, that is, formulas φ in $L(PA)$ such that $S \nvdash \varphi^I$ and $S \nvdash (\neg\varphi)^I$. One may think that this is a direct consequence of Gödel's incompleteness for PA, as S could be seen as a recursive extension of PA. But this is false. As mentioned before, no subtheory of an extension of ZF is bi-interpretable with any extension of PA. Indeed, PA is bi-interpretable with the theory ZF_{fin} composed of ZF without axiom of infinity and with the addition of negation of infinity and transitive closure (see [24]). However, no extension of ZF_{fin} can be S, since S asserts the existence of infinite sets. In view of this, we prove the very simple theorem:

Theorem 1. *For a given interpretation I of PA in a recursive extension S of ZF, there will be formulas of $L(PA)$ such that $S \nvdash \varphi^I$ and $S \nvdash (\neg\varphi)^I$.*

Proof. To prove this, we should reinternalize the provability predicate under the interpretation. Let as consider $A = \{\varphi \mid S \vdash \varphi^I\}$. Notably, $PA \subseteq A$ and thus A can produce arithmetization for arithmetic formulas and for set-theoretic formulas. Let $\ulcorner\varphi\urcorner$ be the Gödel number of any formula φ in A or in S and $\ulcorner\langle\varphi_1, \varphi_2, \ldots, \varphi_n\rangle\urcorner$ the Gödel number of any sequence of formulas $\langle\varphi_1, \varphi_2, \ldots, \varphi_n\rangle$ in A or in S (as done in ([25] pp. 122–126)).

Since S is recursive, "$\ulcorner\langle\varphi_1, \varphi_2, \ldots, \varphi_n\rangle\urcorner$ is a proof in S" is recursive. From the representation theorem (see [25] pp. 126–128), there is a predicate $Pr_S(x, y)$ such that

$$A \vdash Pr_S(\ulcorner\langle\varphi_1, \varphi_2, \ldots, \varphi_n\rangle\urcorner, \ulcorner\psi\urcorner) \iff \langle\varphi_1, \varphi_2, \ldots, \varphi_n\rangle \text{ is a proof in S and } \psi \text{ is } \varphi_n \tag{5}$$

Moreover, the statement "ψ is the φ^I of some φ" is recursive. Then, from the representation theorem, there is a predicate $Fml_I(x)$ such that

$$A \vdash Fml_I(\ulcorner\psi\urcorner) \iff \psi \text{ is the } \varphi^I \text{ of some } \varphi \tag{6}$$

Defining $Th_S^A(y)$ as $\exists x (Pr_S(x, y) \wedge Fml_I(y))$, we can then use the diagonal lemma for the formula $\neg Th_S^A(y)$, obtaining a formula G such that

$$A \vdash G \leftrightarrow \neg Th_S(\ulcorner G\urcorner) \tag{7}$$

If $S \vdash G^I$, then $A \vdash Fml_I(\ulcorner G \urcorner)$ and $A \vdash Th_S(\ulcorner G \urcorner)$ from (5) and (6). From (7), we have $A \vdash \neg Th_S(\ulcorner G \urcorner)$, contradiction. To obtain a contradiction from $S \vdash \neg G^I$, we should reformulate the proof using the Rosser trick, although it will also work the same way as in ([25] pp. 131–132). Then the formula G obtained in the diagonalization for the equivalent Rosser-Gödel predicate is the undecidable arithmetic formula in S. □

This theorem can be understood as a **very small** expansion of Gödel's incompleteness theorem as we consider decidability under relations between theories. Moreover, it relates to results available in *Satisfaction is not absolute* [1]. In this article, Hamkins and Yang considered the idea that there may be arithmetical formulas ρ that two models of ZF disagree–even as these same models agree on what is the standard model for arithmetic. Though very important in the context of this paper, the result lacks a construction for the ρ formula. This formula is obtained as the existential for a number representing a formula. In fact, exhibiting ρ is not possible, since it would imply the inconsistency of ZF.

Put another way, we have shown a similar phenomenon in which disagreement can be exhibited. To make it possible, we considered a foundational view that accommodates our incomplete understanding of set theory and arithmetic. Thus, agreement on arithmetic is to be understood as having similar sets of known arithmetical truths $\{\varphi \mid S \vdash \varphi^N\}$, S being some stage (or alternative stage) in the development of ZF. In this sense, there is a formula ρ that would be true in some possible development of S and false in some other possible development of S. As a reviewer pointed out, Ali Enayat [26] has recently studied this phenomenon in a similar light. He points out that $N^{ZF} \subsetneq N^{ZFI}$, where I indicates the existence of an inaccessible cardinal. Interestingly, he also creates a natural way of describing the S's expansion of arithmetic. If $\theta_0, \theta_1, \ldots, \theta_i, \ldots$ is an enumeration of formulas of S and $S_n = \{\theta_i \mid i < n\}$, the resulting arithmetic obtained from S is PA together with statements $\varphi \to Con(S_n \cup \{\varphi^N\})$. Enayat later shows a series of results on how and to what extent set theory models can disagree over the standard model of arithmetic. The limit of his method for the purposes of the present article is that his main concern is a model-theoretical characterization of 'nonstandard' models (with respect to some background V) that are obtained in some S using the standard interpretation.

There are indeed various important open statements of finite set theory. The recent book "Extremal problems for finite sets" ([26] pp. 211–215) deals with some of those systematically: Erdős matching conjecture, Chvátal conjecture, Frankl's union-closed conjecture and so on. If some of these turn out to be undecidable in ZF (or ZFC), they will correspond to undecidable statements of arithmetic under the standard interpretation. The question we would like to propose is this: assuming that the standard interpretation of PA in ZF produces true arithmetic statements, should we simply say that if some set theorists decide to include some of those conjectures as axioms, then should number theorists accept the corresponding statements as arithmetic truths?

In particular, there has been an important debate regarding the multiversalist picture of set theory. Many set theorists today consider that there are indeed equally legitimate non-isomorphic set theoretic models. The motivations for this are various (see [8]). But do those motivations apply to arithmetic? With set theory, there is a fundamental limitation generally accepted even by many conservative set theorists: whenever we deal with a model of set theory, we should always set a limit to an ordinal level in the cumulative hierarchy. Therefore, there is at least a multiverse of set-theoretic models with respect to ordinal levels. Nothing similar to this is found in arithmetic intuitions. Natural numbers are precisely those one can effectively count and there is little to no reason to take a pluralist view with respect to arithmetic. Notice, however, that by accepting the multiversalist view of set theory together with the view that the one true reduction of arithmetic to set theory is the standard interpretation, we are consequently subscribing to a pluralist view of arithmetic. And this is precisely the conclusion drawn by Hamkins. Now, if there is only one model of arithmetic and many legitimate set theoretic models, it becomes fundamentally important to consider that the interpretation of arithmetic in set theory is revisable and that the model of arithmetic may not even characterizable in some set

theoretic models. It is in view of this consideration that we should now investigate what we call *the coordination problem*.

3. The Coordination Problem

Let us consider the following fictional scenario for the development of set theory and arithmetic. There are two groups of mathematicians who would decide about new axioms for set theory and arithmetic. The first G_s is responsible for one (among possibly many) set-theoretic universe, and the second G_a for the arithmetic structure. Let us further assume that G_a agrees with the standard expansion of arithmetic in ZF (A_N^{ZF} is considered valid for G_a). How should we frame the relation between the two groups?

Consider that G_s have decided in favor of new axiom α to set theory ZF. In particular, this would expand the set of arithmetic truths in $A_N^{ZF+\alpha}$. Should G_a consider this new set to be true? This being the general attitude towards arithmetic means that the standard reduction determines new truths for arithmetic. In what sense does the standard interpretation provide a foundation for new arithmetical truths? If we think that the standard interpretation does this, it seems like we have simply assumed that arithmetic lives in set theory, without any further considerations. After all, this framework bounds the expansion of arithmetic truth to the expansion of set-theoretic truth. Therefore, G_a would not have any authority over new arithmetic axioms after all.

In order to make room for this setting, one should consider that we have a better understanding on how arithmetic is reduced to set theory than we have for each of the theories. And, for this to work in general, we should consider the reduction of arithmetic in set theory **unrevisable**.

Very often we consider ourselves to have a good understanding on relations between things that we may not have a good understanding. This is the case for translating a sentence like "Napoleon was an emperor". We may have a lot of doubts about the ontological status of the words used in this sentence and still be confident about how to translate it into Chinese.

Indeed, we may be more confident about the way we reduce arithmetic to set theory than about the truth in these theories. Yet this is not sufficient to assume the unrevisability of the reduction relation. After some investigation over the concept of emperor, one has realized that the standard translation of emperor in Chinese does not really represents what English speakers refer with 'emperor'. For instance, emperor is usually translated as 'Huangdi' in Chinese, even though this word associate the monarch with his divinity. In English, although often associated with divinity, the word emperor can be used without divine association. So a more intricate description as 'Napoleon was the non-divine man who ruled over the French empire' would be better (even if it is not practical).

If there are grounds for taking N to be a privileged interpretation, those would be based on partial representations of arithmetic and set theory. Therefore, the idea that N correctly works as a connection between the theories may be simply because we have not advanced the theories enough. This would be a similar case if a Chinese working in the translation of a western modern history book has been translating 'Emperor' as 'Huangdi'. It seems perfectly fine if he believed this to be a general translation, given that the only time he applied the translation was for the 'Emperor of the Holy Roman Empire'. But as he starts translating the Napoleonic period, the broader picture would force him to reconsider the generality of the translation.

A different picture would be the case where the Chinese translator invented a language where w means 'blue chair'. Finding someone else using w to refer to a red chair, he could correctly accuse the person to be using the word incorrectly. So this would be similar to the case where we consider arithmetic to be a definition inside set theory. But this being the case would imply that there is no foundational gain in studying the relation between the theories.

Whereas set theory has a foundational role for arithmetic, we may now consider that the standard interpretation is a good yet revisable set-theoretic inspection over arithmetic.

It is precisely because we assume the interpretation to be revisable that a foundational relation can be argued. As truth expands in both theories, we evaluate conflicts and revise, if necessary, the interpretation to accommodate changes. A summary of the steps in the coordination of G_a and G_s can be:

1. Every addition of axioms to one theory should provoke an inspection over the adequacy of the current interpretation of arithmetic in set theory.
2. If a conflict arises in the development of the theories, the two groups should meet to adjust the interpretation to prevent the conflict.
3. The adequacy of an interpretation should have reasons for itself apart from accommodating the interpretation.

As we see in Step 2, the two communities should sit together and reevaluate the state of the reduction, if necessary. Hopefully, these conferences would hardly occur. But we should allow some independence to each group. Otherwise, their development, especially on arithmetic, would turn out to be assumed by definition in the development of the other.

We have added some life to the grounding relation by allowing it to fail. However, there is still a deeper problem. The following scenario is still possible:

(i) Each instance of development allows one to fix the interpretation between the theories.
(ii) And at least one of the extensions of any state of arithmetic is not possibly interpreted in set theory.

Allowing both of these possibilities weakens the edifice of the grounding relation. Each moment in the development of the theories is an incomplete stage in which we cannot anticipate the impossibility of reductions occurring further in the development of the theories. From (i), any addition to the theories allows one to find (or keep) an interpretation of arithmetic. However, from (ii), finding those interpretations does not add to the idea that arithmetic is indeed reducible to a given set theory. This scenario is possible, as we will see in the next theorem.

Theorem 2. *Let S be a consistent extension of ZF and A a consistent recursive extension of PA, then there is a consistent extension A^* of A that is not interpretable in S.*

Proof. We extend the theory A by generating a sequence of theories that are not interpretable in S by a particular interpretation I. Being these theories compatible with each other, the union of them will not be interpretable in S.

Let $A_0 = A$ and $\{I_1, I_2, \ldots\}$ be an enumeration of all interpretations from PA language to ZF language. We generate a sequence of theories $A_0 \subseteq A_1 \subseteq \ldots \subseteq A_n \subseteq \ldots$ by adding **one** formula in each step. It should be noticed that the proof here is not constructive, meaning we are not using a recursive method to determine the new formula added to A_i to obtain A_{i+1}. Nonetheless, since every theory A_i will be the addition of i formulas to the recursively axiomatized A_0, then A_i is also recursively axiomatized. In this case, for every i, there is a formula G_i obtained by the Rosser-Gödel diagonalization argument. With this in mind, we define the A_i's as follows (abbreviation: $T \leq_J T'$ represents "T is interpreted in T' by J"):

Let $\varphi_0, \varphi_1, \ldots, \varphi_k, \ldots$ be an enumeration of arithmetic formulas.

1. If $A_i \leq_{I_i} S$ and there is a least k such that $A_i \nvdash \varphi_k$ and $S \vdash \varphi_k^{I_i}$, then

$$A_{i+1} = A_i \cup \{\neg \varphi_k\}$$

2. Otherwise,
$$A_{i+1} = A_i \cup G_i$$

Let $A^* = \bigcup_{i \in \omega} A_i$. We note that A^* is a consistent extension of A because in each step we add an unprovable formula.

Suppose A^* is interpretable by I in S, then $I = I_k$ for some natural number k. Notably, if a theory T is interpreted in a theory T', then any subtheory of T is interpreted in T' by the same interpretation. Thus the entire sequence of theories $\{A_1, A_2, \ldots\}$ is interpreted in S by I_k. In particular, we have $A_k \leq_{I_k} S$ and $A_{k+1} = A_i + \neg\varphi_q$ or $A_{k+1} = A_k + G_k$ as in the definition. If $A_{k+1} = A_i + \neg\varphi_q$, then option 1 in the definition was used and we have $S \vdash \varphi_q^{I_k}$. However, since S also interprets A_{k+1} with I_k, we have the contradiction $S \vdash \neg\varphi_q^{I_k}$. If $A_{k+1} = A_k + G_k$, then option 1 is not applied and we have either (i) $A_k \not\leq_{I_k} S$ or (ii) that, for all n, $A_i \vdash \varphi_n$ if, and only if, $S \vdash \varphi_n^{I_k}$. Note that (i) contradicts $A_k \leq_{I_k} S$. Moreover, since $A_i \not\vdash G_i$, it follows from (ii) that $S \not\vdash G_k^{I_k}$–which, in turn, implies the contradiction $A_{k+1} \not\leq_{I_k} S$. Therefore, A^* is not interpretable in S. $\square$

Let $A = A_N^{ZF}$, Ak be the Ackermann interpretation of membership in arithmetic language and consider that a formula φ is equivalent to $(con(ZF))^{Ak}$ in A. Suppose also that the group G_a considers φ to be valid. Notably, this formula would represent a relation between natural numbers such that the standard interpretation stops being a correct interpretation of arithmetic. Similar constructions can be used to generate a myriad of examples. However, each of these examples can be subject to a 'contrary to intuition' kind of criticism. In the case presented, one may suggest that $(con(ZF))^{Ak}$ means that we are adding an axiom representing the consistency of ZF in the arithmetic without doing the same in the set theory. Simply adding the axiom $con(ZF)$ to our set theory would make the standard interpretation work again nicely. Nevertheless, we note that the phenomenon presented in the theorem is not exactly to add isolated axioms, but to add an enumeration of axioms to the arithmetic. Our suggestion is therefore that a bundle addition of axioms may force the theories to loose coordination. We also note that we do not impose the set theory S to be recursive. For this reason, one may simply consider that S is a complete extension of ZF. In this case, no addition to the set theory would possibly allow the theories to recover the interpretability relation.

We argued that it is possible for ZF and PA to part ways along the path of development. Although disturbing, this may simply account for the meaningfulness of the question about the reduction between the two theories. We have considered that we should conceive it to fail (even fatally, as in this case) in order not to take for granted that the reduction works. Note further that this pays tribute to the idea that by interpreting arithmetic in set theory we should inform something that was not simply given, i.e., that arithmetic lives in the realm of set theory. Nonetheless, we should now show the simple (and not a novelty) result that the number consistent extensions of PA is uncountable. Meanwhile, the number of interpretations is trivially countable. This means that we are in a situation similar to that of choosing a random number in the Real line expecting to find a natural number. Our claim is that, for this reason, the coordination between the systems can work only if the coordination is assumed from the beginning and as a principle.

Theorem 3. *Let A be a consistent recursive extension of PA, then there is a uncountable number of consistent extensions of A.*

Proof. From the incompletness theorem, there is a formula G that is undecidable in PA. Thus, both $PA + G$ and $PA + \neg G$ are consistent. Notably, this is still true for the addition of any finite number of new axioms $\alpha_1, \alpha_2, \ldots, \alpha_n$. There is a formula $G_{\langle i \rangle}$ that is undecidable in $A_{\langle i \rangle} = PA + \{\alpha_1, \alpha_2, \ldots, \alpha_n\}$ since $A_{\langle i \rangle}$ is a recursive extension. Let us then index PA extensions with binary codes (i.e. sequences of 0's and 1's) in the following way:

1. $A_{\langle 0 \rangle} = PA$.
2. If $G_{\langle i \rangle}$ is the undecidable obtained with Rosser-Gödel technique $A_{\langle i \rangle}$, then $A_{\langle i1 \rangle}$ is $A_{\langle i \rangle} + G_{\langle i \rangle}$ and $A_{\langle i0 \rangle}$ is $A_{\langle i \rangle} + \neg G_{\langle i \rangle}$. (where $i1$ and $i0$ are the binary extension of the code i with the digits 1 and 0)
3. Let *FinBin* be the set of all finite binary codes, the set $\Sigma = \{A_{\langle x \rangle} \mid x \in FinBin\}$ is a subset of finite extensions of PA.

Note that each member of Σ is an extension of PA with the addition of a finite number of formulas. Now we build infinite extensions of PA from Σ. Let $M : FinBin \to \Sigma$ be the map between binary codes and extensions in Σ. We say that $C : \omega \to FinBin$ is a chain in *FinBin* when $\forall x, y \in \omega(x \leq y \to (C(y)$ extends the code of $C(x)))$. Also, if $x \in FinBin$, we write

$$x(n) = \begin{cases} \text{n'th digit of } x, & \text{if there is the n'th digit} \\ 0, & \text{otherwise} \end{cases}$$

If C is a chain in *FinBin*, then $dig_C = \langle (C(0))(0), (C(1))(1), \ldots, (C(n))(n), \ldots \rangle$ is an infinite binary code associated with the extension Ex_C obtained by

$$\bigcup \{C(i) \mid i \in \omega\}$$

We define Π as the set

$$\{\langle dig_C, Ex_C \rangle \mid C \text{ is a chain in } FinBin\}$$

Note that Π is a function from the set of all binary infinite codes to extensions of PA. Since infinite binary codes are uncountable, we need only to show that Π is injective and that the image of Π is composed of consistent extensions of PA.

Suppose that some Ex_C is not consistent; then there is a finite proof of the inconsistency of Ex_C. Hence, there is $n \in \omega$ such that $Ex_C^n = \bigcup\{C(i) \mid i \in n\} = C(n)$ is inconsistent. But this is false, since each $C(i+1)$ obtained by adding an unprovable formula to $C(i)$ and $C(0) = PA$ is assumed consistent.

Suppose that $\Pi(dig_{C_1}) = \Pi(dig_{C_2})$ and that $dig_{C_1} \neq dig_{C_2}$. Then there is the least i such that $dig_{C_1}(i+1) \neq dig_{C_2}(i+1)$. This means, without loss of generality, that $C_1(i+1) = C_1(i) + G_{C_1(i)}$, $C_2(i+1) = C_2(i) + \neg G_{C_2(i)}$ and $C_1(i) = C_2(i)$. Therefore, $\Pi(dig_{C_1})$ contains the formulas $G_{C_1(i)}$ and $\neg G_{C_1(i)}$. This is absurd, as we just showed that the image of Π is composed of consistent extensions of PA. $\square$

We note that the same can be obtained, even if the starting point includes all theorems of the set theory S under the interpretation. Indeed, we can include the theorems under a given interpretation at any point without interfering with the result.

Although extensions like A^+ are in general not interpretable in S, the process of generating these theories is internalizable in S. Therefore, we may say that *S* proves the consistency statement for all these extensions. This is not enough to claim a proper foundational relation. The model construction emerging from this type of consistency proof is simply given by the existence of a model as in the Henkin canonical construction. Thus, the foundational model one can generate provides little more information than saying that the theory is consistent (see [27]). Therefore, we should not consider those cases as a path to avoid the problem discussed in this section.

As developed in this section, we should not consider that the addition of new axioms to the systems is, in principle, coordinated. Instead, the reducibility of arithmetical truth should be a result of the expressiveness of set theory. However, assuming that the choices of the two groups G_a and G_s would result in a interpretable arithmetic is similar to expect that a random choice of a real number to be a natural number (which has probability zero). It follows that coordination between the groups of mathematicians can only occur in principle. Hence, the reduction of arithmetic truth to set theory is not attainable unless assumed and the foundational relation should be based on other grounds.

To further elaborate on this conclusion, let us consider a metaphor. Picture the situation in which we have the unstable equilibrium of a sphere on a hill with a very small slope. We would like to say that the appearance of equilibrium represents our intuitions about the reduction between the theories being correct. Indeed, we have put the sphere in a position that appears to be an equilibrium. As the slope of the hill is very small, our perception of equilibrium works really well. However, even if it takes a long time, it will become evident that the interpretation of PA in ZF is not in equilibrium. We are, nonetheless, in a

better position if we accept the multiversalist view of set theory. Under this assumption, we should thus say that there are indeed some universes perfectly coordinated with arithmetic under the standard interpretation, and there are some universes perfectly coordinated with arithmetic under other interpretations. However, these universes are only a small portion among a much larger multitude of possible universes of set theory.

The ideas developed in the present article, especially in Theorem 3, bring attention to the fact that we are talking about an unstable hill. No matter how the sphere appears to be at rest, we know that eventually it will gain traction and fall. The project of using N for grounding arithmetic truth is equivalent to finding the equilibrium peak of the hill. It seems to be a good project as we focus on the movement of the sphere–but an analysis of the geography of the hill is already sufficient to conclude this hill to be unstable. We should not base our foundational investigations on the guarantee that we have the correct interpretation in a fixed set theory. Instead, we should use the interpretations as it informs about arithmetic concepts and as it considers bundles of arithmetic formulas in the very expressive environment of set theory.

Our position is not that the standard interpretation N cannot play a foundational role. Alternatively, the very possibility of investigating expansions of arithmetic propositions provided by analyzing N (or other interpretations) is all the ground we need. In place of using foundational relations to establish 'arithmetic truth', we propose using the N interpretation to understand how bundles of arithmetical propositions relates to each other. In this case, we use the technical apparatus and the expressiveness of theories like ZF to analyze arithmetical concepts rather than fixing its truth.

4. Final Remarks

Rather than manipulating models of PA, we considered interpretations of PA in ZF. Our goal was to accommodate the incomplete picture of the set-theoretical metatheory into our analysis of the foundations of arithmetic. The standard interpretation expands what we may consider true in arithmetic: many undecidable formulas in PA become theorems when examined under the interpretation in ZF. This is a general phenomenon. For every well founded interpretation of recursive extensions of PA in extensions of ZF, the interpreted version of arithmetic has more theorems than the original. This shows that studying arithmetic inside set theory can be significant. As one considers these interpretations, one explores **expansion of arithmetic truth** and how the addition of bundles of axioms plays out.

We continued by introducing the coordination problem. We considered two independent communities of mathematicians responsible for deciding over new axioms of ZF and PA. Using this setting, we studied the possibility of coordinating PA with PA's interpretation in ZF. Nonetheless, we showed that it is possible to have extensions of PA that are not interpretable in a given set theory S. Moreover, we consider a given recursive extension A of PA and an extension S of ZF. Here, we prove that there are uncountable extensions of A while countable interpretations of arithmetic in S. This last result implies that the coordination between the two communities of mathematicians should be coordinated from the start. However, we argued that this would empty the foundational role of set theory over arithmetic.

We have, therefore, set a framework to criticize the notion of grounding truth between theories such as arithmetic and set theory, specially with respect to the idea of fixing an interpretation between the systems. Indeed, the multiversalist propagates their pluralism from set theory to arithmetic by relying on the standard interpretation. We reject this conclusion, arguing that it is the interpretation that should be revised. By allowing the interpretation of arithmetic into set theory to change, we make compatible the set theoretic pluralism with the view that there is a single arithmetic structure.

However, this is not to be understood as a general criticism of the idea of using set theory to investigate foundational matters regarding arithmetic. Instead, we have solely shown that it may be flawed to assume that a single set theory can really provide grounds

for arithmetic truth or a definitive description of the universe of numbers. Our suggestion is therefore to consider a foundational relation that aims primarily at conceptual clarification of the concepts involved in the studied theory. An expressively rich environment such as set theory is armed with tools to study arithmetical relations in wider settings than it would be possible without leaving its deductive apparatus.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hamkins, J.D.; Yang, R. Satisfaction is not absolute. *Rev. Symb. Log.* 2014, 1–34, *accepted*. Available online: http://jdh.hamkins.org/satisfaction-is-not-absolute/ (accessed on 18 May 2022).
2. Hamkins, J.D. The modal logic of arithmetic potentialism and the universal algorithm. *Mathematics* **2018**, 1–35, *under review*. Available online: http://jdh.hamkins.org/arithmetic-potentialism-and-the-universal-algorithm/ (accessed on 18 May 2022).
3. Hamkins, J.D.; Linnebo, O. The modal logic of set-theoretic potentialism and the potentialist maximality principles. *Rev. Symb. Log.* **2022**, *15*, 1–35. [CrossRef]
4. Koellner, P. Hamkins on the multiverse. In *Exploring the Frontiers of Incompleteness*; Harvard University: Cambridge, MA, USA, 2013.
5. Tarski, A.; Mostowski, A.; Robinson, R.M. *Undecidable Theories*; Elsevier: Amsterdam, The Netherlands, 1953; Volume 13.
6. Gabbay, M. A formalist philosophy of mathematics part I: Arithmetic. *Stud. Log.* **2010**, *96*, 219–238. [CrossRef]
7. Ferreirós, J. *Mathematical Knowledge and the Interplay of Practices*; Princeton University Press: Princeton, NJ, USA, 2015.
8. Hamkins, J.D. The set-theoretic multiverse. *Rev. Symb. Log.* **2012**, *5*, 416–449. [CrossRef]
9. Williamson, T. Absolute provability and safe knowledge of axioms. In *Gödel's Disjunction: The Scope and Limits of Mathematical Knowledge*; Phillip Books: New York, NY, USA, 2016; pp. 243–252.
10. Maddy, P. A second philosophy of arithmetic. *Rev. Symb. Log.* **2014**, *7*, 222–249. [CrossRef]
11. Quine, W.V. Ontological reduction and the world of numbers. *J. Philos.* **1964**, *61*, 209–216. [CrossRef]
12. Putnam, H. Mathematics without foundations. *J. Philos.* **1967**, *64*, 5–22. [CrossRef]
13. Freire, R.A. Interpretation and Truth in Set Theory. In *Contradictions, from Consistency to Inconsistency*; Springer: Cham, Switzerland, 2015.
14. Enayat, A. Variations on a Visserian Theme. In *Liber Amicorum Alberti: A tribute to Albert Visser*; College Publications: London, UK, 2016; pp. 99–110.
15. Freire, A.R.; Hamkins, J.D. Bi-interpretation in weak set theories. *J. Symb. Log.* **2021**, *86*, 609–634. [CrossRef]
16. Hamkins, J.D. Different set theories are never bi-interpretable. In *Mathematics and Philosophy of the Infinite*; Routledge: London, UK, 2018.
17. Freire, A.R. Estudo Comparado do Comprometimento Ontológico das Teorias de Classes e Conjuntos. Ph.D. Thesis, Universidad Estudal de Campinas, Campinas, Brazil, 2019.
18. Gödel, K. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshef. Math. Phys.* **1931**, *38*, 173–198. [CrossRef]
19. Goodstein, R.L. On the restricted ordinal theorem. *J. Symb. Log.* **1944**, *9*, 33–41. [CrossRef]
20. Caicedo, A.E. Goodstein's function. *Rev. Colomb. Matem.* **2007**, *41*, 381–391.
21. Wiles, A. Modular elliptic curves and Fermat's last theorem. *Ann. Math.* **1995**, *141*, 443–551. [CrossRef]
22. McLarty, C. What does it take to prove Fermat's Last Theorem? Grothendieck and the logic of number theory. *Bull. Symb. Log.* **2010**, *16*, 359–377. [CrossRef]
23. Helfgott, H.A. The ternary Goldbach conjecture is true. *arXiv* **2013**, arXiv:1312.7748.
24. Kaye, R.; Wong, T.L. On interpretations of arithmetic and set theory. *Notre Dame Form. Log.* **2007**, *48*, 497–510. [CrossRef]
25. Shoenfield, J.R. *Mathematical Logic*; Addison-Wesley Reading: London, UK, 1967; Volume 21.
26. Enayat, A. Standard models of arithmetic. In *Idées Fixes*; Kaså, M., Ed.; University of Gothenburg Publications: Gothenburg, Sweden, 2014; pp. 55–64.
27. Freire, A.R. Translating non Interpretable Theories. *S. Am. J. Log.* **2020**, *10*, 1–21.