BREAKING THE TIE: BENACERRAF'S IDENTIFICATION ARGUMENT REVISITED

ARNON AVRON* AND BALTHASAR GRABMAYR**

*School of Computer Science, Tel Aviv University

**Department of Philosophy, University of Haifa

This article is forthcoming at *Philosophia Mathematica* (https://doi.org/10.1093/philmat/nkac022). This is a preprint; the published version may be subject to minor changes.

ABSTRACT. Most philosophers take Benacerraf's (1965) identification argument to successfully rebut the reductionist view that numbers are sets. This philosophical consensus jars with mathematical practice, in which reductionism continues to thrive. In this note, we develop a new challenge to Benacerraf's argument by contesting a central premiss which is almost unanimously accepted in the literature. Namely, we argue that — contra orthodoxy — there are metaphysically relevant reasons to prefer von Neumann ordinals over other set-theoretic reductions of arithmetic. In doing so, we provide set-theoretical facts which, we believe, are crucial for an informed assessment of reductionism.

1. Introduction

In his classic paper, Benacerraf [1965] sought to reject two realist positions regarding arithmetic. The first is *object realism* about arithmetic, i.e., the view that numbers are objects. The second is set-theoretic *reductionism* about arithmetic, i.e., the view that numbers are sets. While Benacerraf's argument against object realism is widely taken to be defective,[1] there is a broad consensus among philosophers that his paper is a successful rebuttal of set-theoretic reductionism.[2] Benacerraf [1998] later restated this anti-reductionist argument as follows:

(1) ... indefinitely many accounts satisf[y] all of the conditions [on an account of number];
(2) ... because 'the number n = the object o' has the form of a statement of identity, at most one such account could be correct;
(3) ... there [is] no principled way to choose among them, i.e. to decide which sets the numbers really were;

The order of the authors is alphabetical. The authors contributed equally to this work.

* Orcid.org/0000-0001-6831-3343. E-mail: aa@cs.tau.ac.il.

** Orcid.org/0000-0003-3326-8030. E-mail: balthasar.grabmayr@gmail.com.

[1] See Steiner [1975], Resnik [1997], Wetzel [1989], McLarty [1993], Balaguer [1998], Shapiro [2016].

[2] This reception corresponds to Benacerraf's (1998) retrospective concession that "the sweeping conclusions of section III [against object realism] were being advanced with considerably less confidence than were those of sections I and II [against set-theoretic reductionism]" (p. 50).

(4) … if one of the accounts were the correct one there would be a way to establish which one it was: 'the position that this is an unknowable truth is hardly tenable.' [Benacerraf, 1965, p. 62]
 and therefore
(5) … any feature of an account that identifies a number with a set is a superfluous feature of the account (i.e. not one that is grounded in our concept of number);
 consequently
(6) 'numbers … could not be sets at all.' [Benacerraf, 1965, p. 62]

[Benacerraf, 1998, pp. 52–53]

In this note, we will attempt to defend set-theoretic reductionism from Benacerraf's argument by showing that premise (3) is untenable.[3] That is, we will argue that there are reasons to prefer one set-theoretic reduction of numbers to another. *Prima facie*, our claim seems highly credible, given that all mathematicians who care about these foundational matters (like set theorists and mathematical logicians) unanimously choose one specific set-theoretic reduction of natural numbers, namely, von Neumann ordinals, over its competitors.[4] Yet, premiss (3) is one of the few assumptions in the philosophical literature which is accepted by proponents and opponents of Benacerraf's claim alike.[5] How can this glaring discrepancy between mathematicians and philosophers of mathematics be explained? Do mathematicians have no reasons for this choice? Or, are these reasons irrelevant to the metaphysical problem under scrutiny? In this paper we argue that — contra orthodoxy — the answer to both questions is negative. Thus, Benacerraf's argument is blocked and the reductionist is freed from the multiple-reductions problem.

2. Benacerraf's Argument Revisited

We start by making Benacerraf's argument precise. Premise (1) presents us with the first difficulty. There is no agreement in the literature as to what "*all* of the conditions on an account of the natural numbers" precisely amount to. Even Benacerraf himself seems to have changed his mind and provides different clarifications in his articles of 1965 and 1998 respectively. According to Benacerraf [1965], the conditions on an account of the natural numbers are exhaustively given as follows:

(I): To give definitions of '1', 'number', and 'successor', and '+', '$\times$', and so forth, on the basis of which the laws of arithmetic could be derived;
(II): To explain the "extra-mathematical" uses of numbers, the principal one being counting - thereby introducing the concept of cardinality and cardinal number.

More than 30 years later, Benacerraf [1998] reformulates these conditions as follows:

> Broadly speaking, having identified the target — arithmetic as the theory of $\{0,'\}$, or $\{0,',+,\times\}$, successfully axiomatized by Dedekind, Peano et al., and its applications — you need only

[3] A more common strategy to evade Benacerraf's argument is to contest premise (4). See, for example, Paseau [2009] and Mount [2019]. For a rejection of premise (2) see White [1974].

[4] Although we appeal here to this fact merely for motivational purposes, we strongly believe that in mathematics such "sociological" facts are always based on mathematical insights and good mathematical reasons. We shall return to this issue in Section 3.

[5] Steinhart [2002] is a notable exception. He attempts to provide a "mathematical demonstration" (p. 343) that numbers are von Neumann ordinals. We believe that Steinhart's arguments are unconvincing. Metaphysical questions cannot be settled by mathematical proof alone. See Ginammi [2019], D'Alessandro [2018] and Ebels-Duggan [2022] for extensive criticism of Steinhart's arguments.

> (a) identify *the successor function* with some set-theoretic function f and *zero* with some element e such that the ancestral of f with respect to e is provably a *progression*, i.e. an ω-sequence,
> (b) *the numbers* as the ancestral of f with respect to e, and
> (c) make explicit provision in the account for the expression of the cardinality relation (e.g. by defining the relation $C(x, n)$ between a number n and a set x of cardinality n in some way such that '$C(x, n)$' holds iff there exists a 1-1 correspondence between x and the numbers less than n). [Benacerraf, 1998, emphasis in the original]

There are obvious significant differences between these two formulations.[6] Moreover, Benacerraf's formulations are insufficiently precise. For example, he does not specify whether he has a first- or second-order framework in mind.[7] In fact, Benacerraf [1998] deliberately dodges this question when he speaks about "arithmetic as the theory of $\{0,'\}$, or $\{0,', +, \times\}$" (p. 46): the first option makes sense only if one uses second-order logic, while the second option is necessary only in the first-order case. But then what exactly are all of the conditions on an account of the natural numbers, i.e., how can premiss (1) be made precise?

One possible explanation of this lack of precision is that Benacerraf's argument is sufficiently robust to accommodate several explications of an account of number. We can thus understand Benacerraf's argument schematically, i.e., relative to how the conditions on an account of number are precisely understood. Given any such way of making these conditions precise, let $\mathcal{A}$ denote the class of set-theoretic reductions of arithmetic, which satisfy all of the conditions on an account of number. For instance, on one plausible reading, $\mathcal{A}$ may be taken to be the class $\mathrm{Mod}(\mathbf{PA}_2)$ of set-theoretic models of second-order Peano Arithmetic $\mathbf{PA}_2$.[8] For what follows, it is sufficient to assume that at least the ω-sequences based on von Neumann and Zermelo ordinals are contained in $\mathcal{A}$. Benacerraf's argument can then be reconstructed more precisely as:

- (R1) If numbers are sets, then exactly one $R \in \mathcal{A}$ is a correct set-theoretic reduction of arithmetic;
- (R2) There are multiple set-theoretic reductions of arithmetic, i.e., $|\mathcal{A}| \geq 2$;
- (R3) For no $R \in \mathcal{A}$ do we have reasons to believe that it is the correct set-theoretic reduction of arithmetic;

[6] These differences include: (1) In Benacerraf [1965], an important part of the first condition is to adequately define +, $\times$, exponentiation, "and so forth". In Benacerraf [1998] the inclusion of + and $\times$ is an option, while no 'so forth' is demanded, or even mentioned. (2) In Benacerraf [1965] the natural numbers start with 1, while in Benacerraf [1998] they start with 0. (3) The conditions in Benacerraf [1998] explicitly refer to being a *progression* as the most essential feature of the natural numbers. This demand is not included in (I)-(II) of Benacerraf [1965]. Nevertheless, it is repeated in that paper many times. It is also important that the vague notion of "progression" is explicated in Benacerraf [1998] by the more exact notion of a ω-sequence — a notion which is not mentioned in Benacerraf [1965]. (4) In Benacerraf [1998] it is demanded to have a *proof* that a proposed ω-sequence is indeed an ω-sequence. There is no such condition in Benacerraf [1965]. In contrast, in Benacerraf [1965] the proposed ω-sequence is required to be recursive. Benacerraf [1996, 1998] rejects this requirement. (5) Benacerraf [1998] makes explicit and significant reference to the crucial notion of ancestral. Benacerraf [1965] does not even mention it. (6) Benacerraf [1965] does not talk about theories. He only refers to "the laws of arithmetic". In contrast, "the theory of $\{0,'\}$, or $\{0,', +, \times\}$, successfully axiomatized by Dedekind, Peano et al." has a crucial role in Benacerraf [1998].

[7] This ambiguity regarding "Peano's axioms" in Benacerraf's papers was also noted by Ginammi [2019].

[8] For example, McLarty [1993] explicitly takes "the theory of **N**" to be $\mathbf{PA}_2$. Rouilhan [2016] refers to '**PA**' as the theory of **N**, but he too clearly means $\mathbf{PA}_2$.

(R4) If $R \in \mathcal{A}$ is the correct set-theoretic reduction of arithmetic, then we have reasons to believe that it is the correct set-theoretic reduction of arithmetic;

Therefore:

(R5) No $R \in \mathcal{A}$ is the correct set-theoretic reduction of arithmetic;

Consequently,

(R6) Numbers are not sets.

Premise (R1) spells out the core tenet of the realist reductionist view, as understood in this paper (see premise (2) of the original argument). (R2) is true by our assumption on $\mathcal{A}$. That is, at least the ω-sequences based on von Neumann and Zermelo ordinals are set-theoretic reductions of arithmetic. (R3) is usually accepted by proponents and opponents of Benacerraf's argument alike, given that there are multiple set-theoretic reductions of arithmetic (see (R2)). (R4) is a contentious assumption which has been disputed in the literature (cf. footnote 3). Paseau [2009], for example, rejects (R4) by denying that speakers have transparent knowledge of the referents of their numerals (p. 35). (R5) follows from (R3) and (R4) and (R6) is an immediate consequence of (R1) and (R5).

3. Breaking the Tie

As noted above, almost all proponents and opponents of Benacerraf's argument accept premise (R3). Paseau [2009] compares the inability of a set-theoretic reductionist to prefer one set-theoretic reduction to another with Buridan's ass, which "died of inanition as a result of not breaking the tie between a bucket of water to its left and a stack of hay an equal distance to its right, at a time when it was equally hungry and thirsty. The two choices were perfectly symmetric, and there was no reason to prefer one to the other" (p. 38). While Paseau [2009] goes on to argue that rationality requires an arbitrary choice rather than no choice, we argue that the choice is not arbitrary to begin with. For simplicity (and following Benacerraf), we focus on the two most canonical and attractive set-theoretic competitors, namely, the reductions based on von Neumann and Zermelo numbers. We will argue that these reductions are not on par with each other, since we have reasons to believe that the von Neumann account is superior to the one by Zermelo. Hence, by closely examining mathematical properties of set-theoretic reductions, we can reject premise (R3).

A natural starting point to assess the alleged symmetry of set-theoretic reductions of arithmetic is mathematical practice. Consider, for example, the related claim that we do not have any reasons to prefer one arithmetic reduction of finite strings to another. Drawing on mathematical practice, this claim could be corroborated by the observation that logicians treat the choice of a Gödel numbering as a highly arbitrary matter [Visser, 2011, 2016]. Indeed, a plethora of different numberings can be found in the literature. However, the situation is fundamentally different in the case of set-theoretic reductions of arithmetic. Here, the mathematical community settled on one specific set-theoretic reduction of numbers, namely, von Neumann ordinals. The advocate of Benacerraf's argument is committed to believe that mathematicians have no principled reasons for this choice. For example, she might hold that the prevalence of von Neumann ordinals is merely a matter of historical contingency, without reflecting any mathematical or conceptual superiority of von Neumann's account. This view seems rather unconvincing, given certain historical facts. Firstly, the preference of von Neumann ordinals clearly was not determined by historical priority, since Zermelo numbers were introduced first in Zermelo [1908]. Several years later von Neumann, Mirimanoff and Zermelo(!) himself independently discovered von Neumann ordinals [Hallett, 1986, pp. 276–280].[9] Since the 1930s, von Neumann ordinals prevail as the standard choice of set-theoretical reductions of numbers. If Zermelo's definition, as maintained by Benacerrafians, were on par with von Neumann's, there would have been no need for the later introduction of von Neumann ordinals.[10] In other words, if there are no conceptual reasons to prefer von Neumann to Zermelo ordinals, then how can this collective change in mathematical practice in the early 1930s be explained? Secondly, if set-theoretic reductions of numbers are as arbitrary as arithmetic reductions of finite strings, then why did mathematicians settle on a unique choice with regard to the former, yet since the 1930s continue to use multiple choices with regard to the latter? We take these considerations from mathematical practice to strongly suggest that there are reasons to believe that the von Neumann account is superior to its competitors, like that given by Zermelo. In what follows, we present several such reasons.

3.1. **Relating natural numbers to ordinals.** Ordinals have originally been introduced by Cantor as a generalization of the natural numbers in their role as ordinal numbers which is applicable to arbitrary sets (or at least those that can be well-ordered). Since then, one of the basic modern views of the natural numbers construes them as *finite ordinals*. And indeed, this is exactly what they are according to von Neumann's definition. In contrast, there is no non-artificial way to extend Zermelo's definition of the natural numbers to infinite ordinals. In fact, von Neumann's definition is the *only* reasonable set-theoretical definition of the natural numbers that has this crucial property, because von Neumann's definition of an ordinal is the only reasonable set-theoretical definition of this notion.[11] We believe that this alone is a principled reason to single it out from all the possible set-theoretical candidates for identifying the natural numbers.[12]

---

[9]Von Neumann introduced his ordinals in his 1923. Zermelo discovered von Neumann ordinals in unpublished work in 1915 and used them in his 1930. (See also [Bernays, 1941, p. 6] and [von Neumann, 1928, p. 374, footnote 2].) So Zermelo himself actually preferred "Von Neumann" ordinals to "Zermelo" ordinals…

[10]The reader may object that even though von Neumann and Zermelo's definitions are on par as accounts of natural numbers, the former might have been preferred over the latter in virtue of features external to the concept of number. See Section 4 for a response to this objection.

[11]Similar observations can be found in Maddy [1990], Clarke-Doane [2008], Mount [2019], D'Alessandro [2018], Steinhart [2002] and Ebels-Duggan [2022].

[12]An anonymous referee raised the worry that requiring numbers to be finite ordinals might be a double-edged sword, since there are reasons not to identify ordinals with sets, and instead to conceive of ordinals as order-types of well-ordered classes. However, this pre-theoretic view of the ordinals directly leads to the Burali-Forti paradox, and is highly problematic [Copi, 1958]. It is precisely the rigorous treatment of the notion of an ordinal within set theory which has provided us with a coherent concept of infinite ordinals. (Compare our discussion in Section 3.2 of the different ways "new abstract objects" are defined in mathematics.) What is more: the multiple reduction argument against reductionism about *ordinals* is widely taken to be much less promising than Benacerraf's argument against reductionism about *numbers* (see, e.g., Mount [2019]). In fact, von Neumann's reduction practically has no competitor.

An additional way to address the referee's worry is to recall the dialectical stage at which our discussion takes place. For argument's sake, Benacerraf has granted that reductionism is correct. Moreover, it has been observed that there are different set-theoretical reductions of arithmetic. The current stage of the argument is concerned with the problem of choosing among them. As Benacerraf claims, there are no reasons to prefer one reduction over another, which leads to his multiple reduction argument against reductionism. Hence, our rebuttal of premise (R3) takes for granted a provisional acceptance of reductionism about numbers. The additional assumption that also reductionism about ordinals is correct, to wit, to view ordinals as sets, seems to be harmless and in accordance with the present dialectal context.

Let us add here that Benacerraf [1965] would have failed to come up with a convincing Johnny-Ernie tale concerning ordinals,[13] had he tried to do so. Moreover, even in his original tale, the education given to Johnny by his parents would be rather poor (from the parents' own militant logicist point of view) if they did not introduce him to the notion of an ordinal and relate it to the notion of a natural number (cf. Section 4). As for Ernie - his education would surely follow now one of the modern relevant undergraduate textbooks, like Devlin [1993] or Kunen [2009]. In such books the general notion of an ordinal is introduced first, and only then N, the set of natural numbers is defined as a certain subset of the class On of ordinals. Thus N may be defined as the set of finite ordinals, or as the minimal initial segment of On which is closed under S (the successor operation, which is defined on the whole of On), or the maximal initial segment of On in which the operation $+$ (which is defined on the whole of On too) is commutative.[14]

3.2. **The identity of finite ordinals and cardinals.** The dual role of the natural numbers as ordinal and cardinal numbers is one of their most basic features. It is so important that Benacerraf [1998] himself requires an account of the natural numbers to

> make explicit provision in the account for the expression of the cardinality relation (e.g. by defining the relation $C(x, n)$ between a number n and a set x of cardinality n in some way such that '$C(x, n)$' holds iff there exists a 1-1 correspondence between x and the numbers less than n). (p. 47)

Here, Benacerraf practically identifies the cardinal number n with the set of numbers less than n. To understand why, we first note that (finite and infinite) cardinal numbers are abstract objects that are somehow induced by the equipollence relation on sets. This is just a particular case of a general principle: abstract objects are usually introduced in mathematics (but not only in mathematics) by turning statements of the form aRb, where R is an equivalence relation, to equivalent statements of the form $F(a) = F(b)$. (Here the values of F represent/explicate the corresponding "new abstract objects".) There are two standard methods in mathematics of doing this. One is to let $F(a)$ be the equivalence class of a with respect to R. The other is to choose $F(a)$ to be some canonical representative of that equivalence class. Frege's (1884) definition of cardinality (which is mentioned and criticized by Benacerraf) was an attempt to apply the first method. However, this approach has turned out to be futile.[15] Therefore the second method is now commonly used for defining cardinal numbers in set theory. As noted above, this method is also the method implicitly suggested by Benacerraf himself for the finite case: in the above quoted passage he in fact implicitly takes the set of numbers less than n as the canonical representative of the class of sets "of cardinality n". This is indeed an obvious choice, since we *immediately* see, for instance, that the set of numbers less than $10^{10^{10}}$ has $10^{10^{10}}$ elements.[16] In other words, cardinal numbers are commonly taken to be initial

[13]Benacerraf's paper opens with a fictitious tale in which two boys, Johnny and Ernie, are taught arithmetic by their "militant logicist" parents according to the set-theoretic reductions due to *Ernst* Zermelo and *John* von Neumann respectively. This tale occupies about half of the relevant sections I-II of Benacerraf [1965] (See Footnote 2), and its role is to provide quite persuasive support to the philosophical argument that follows it.

[14]The first definition relies on a prior independent notion of "finite set". The other two do not, and so they allow us to define a finite set as a set which is equipollent with some natural number.

[15]At least, the resulting notion of cardinality cannot be used together with $\mathbf{ZFC}$.

[16]The first author of this paper is convinced that the set of numbers less than $10^{10^{10}}$ is the *only* set we immediately know to have the cardinality $10^{10^{10}}$. Note that this knowledge relies on an inductive assumption, according to which all the numbers less than $10^{10^{10}}$ have already been recognized as cardinalities of some sets! Thus, ultra-finitists (also called ultra-intuitionists — see e.g. Van Dantzig [1956], Esenin-Volpin [1970]) deny

segments of the less-than relation $<$ on ordinal numbers. If, moreover, the dual role of natural numbers is accounted for by identifying finite ordinals with finite cardinals, then von Neumann's definition is clearly superior to Zermelo's approach. What is more, von Neumann's account is the *only* set-theoretical definition of a natural number such that the (finite) ordinal number n and the (finite) cardinal number n are identical and such that the cardinal number n is the set of ordinal numbers less than the ordinal number n. More precisely, the von Neumann reduction of numbers to sets is the unique injective function f mapping numbers to sets such that $f(n) = \{a \mid a <_f f(n)\}$ for all $n \in \omega$, where $<_f$ is a relation on the image of f given by $a <_f b \Leftrightarrow f^{-1}(a) < f^{-1}(b)$.[17] This provides another reason to prefer von Neumann ordinals over alternative set-theoretic reductions.

3.3. **Explanatory power.** One of the best possible signs that a certain identification of mathematical objects is correct — or at least more likely to be correct than its competitors — is if it provides explanations of facts and phenomena that alternative identifications cannot account for. The two previous items may be viewed as cases in point with respect to the identification of the natural numbers with finite von Neumann's ordinals. But there is an even stronger case. As is well-known, there is a striking similarity between the standard classifications of the formulæ of the first-order languages of Peano Arithmetic and set theory. In both arithmetic and set theory, formulæ are classified in exactly the same way, yielding an infinite hierarchy of classes with practically the same associated theorems. The most basic class consists of the *bounded* (or Δ_0-) formulæ. They are defined in both cases as the class of formulæ which are recursively obtained from atomic formulæ using the propositional connectives and *bounded* quantifiers. The only difference between the two cases is with respect to the definition of the latter. In set theory the bounded quantifiers have the form $\forall x \in y$ or $\exists x \in y$, while in formal number theory they have the form $\forall x < y$ or $\exists x < y$. This strong analogy between the bounded formulæ of set theory and arithmetic respectively might seem quite mysterious. First, the relation $\in$ between sets and the relation $<$ between natural numbers have very different properties — even if we focus our attention just on the finitary fragment of the universe of set theory: HF, the collection of hereditarily finite sets. For example, $<$ on N is a linear order, while $\in$ on HF is not even a partial order. (Accordingly, the structures $\mathcal{HF} = \langle HF, \in \rangle$ and $\langle N, < \rangle$ are not isomorphic.) Second, when we compare the *full* intended universe $\mathcal{V}$ of set theory and the intended universe $\mathcal{N}$ of number theory, we find two universes which are *very* different in size and nature. In particular: the set $\{x \mid x < y\}$ is finite for every natural number y, while the set $\{x \mid x \in y\}$ is infinite for almost every set y. This fact makes the similarity even more mysterious. However, the mystery almost disappears once we construe $\mathcal{N}$ as a transitive part of $\mathcal{V}$ and view the less-than relation $<$ as actually *identical*

that numbers like $10^{10^{10}}$ exist, i.e. are cardinals of actual sets. There is no point to give them as an example a set like (say) the numbers between $10^{10^{10}} + 1$ and $10^{10^{10}} + 10^{10^{10}}$, since the existence of that set relies on the existence of sets of cardinalities greater than $10^{10^{10}}$...

[17] Upon completion of this manuscript, we learned that Ebels-Duggan [2022] independently uses this characterization to uniquely single out the von Neumann reduction. However, Ebels-Duggan's philosophical assumptions and considerations greatly differ from our presentation. For a more extensive discussion we refer the interested reader to Ebels-Duggan [2022, section 7].

to the restriction of the $\in$-relation to finite ordinals.[18] This is possible, of course, *only* if we identify the natural numbers with the finite von Neumann ordinals.[19] [20]

To be sure, this mystery can be also explained without identifying numbers with von Neumann ordinals. Such explanations may resort to the fact that the structures $\langle N, < \rangle$ and $\langle N_Z, <_Z \rangle$ of *sui generis* numbers and finite Zermelo ordinals respectively are *isomorphic* to a transitive substructure of $\langle \mathcal{V}, \in \rangle$. Note, however, that there is just one such transitive substructure: the set of finite von Neumann ordinals. Hence, von Neumann ordinals play an indispensable role in explanations of this interesting connection between numbers and sets. Thus, von Neumann's account provides the most immediate or best explanation of the mystery described above and so fares better than Zermelo's definition.

Now in set theory the crucial property of Δ_0-formulæ is that they are *absolute*, while in number theory they induce *decidable* relations on the natural numbers. *Prima facie*, the set-theoretic notion of absoluteness is of very different nature than the computability-theoretic notion of decidability. Yet, the identification of $\in$ and $<$ indicates that there might be strong connections between these two notions. This is indeed the case. We present an example of such a connection which is due to Avron [2008]. Here, we assume that our arithmetical language is purely relational. That is, we replace each n-ary function symbol by an $(n+1)$-ary relation symbol. We start by defining a special case of the notion of absoluteness, which is concerned with transitive substructures of $\mathcal{N}$.

Definition 3.1. We say that an arithmetical formula ϕ is $\mathcal{N}$*-absolute*, if for any assignment v in N, ϕ gets the same truth value in all initial segments of N (including N itself) which contain the values assigned by v to its free variables.[21]

We then have:

Theorem 3.2.

- *If ϕ is $\mathcal{N}$-absolute, then ϕ induces a decidable relation.*
- *A relation R on N is recursively enumerable iff R is definable by a formula of the form $\exists x_1, \ldots, x_n \phi$, where ϕ is $\mathcal{N}$-absolute.*

See Avron [2008] for details and for further connections between the notions of absoluteness and decidability.

3.4. **The logical complexity of the definitions.** Next we show that the basic *predicate* N of being a natural number, as well as the less-than relation $<$, can be defined using Δ_0-formulæ in case von Neumann's definition is used. (Recall that Δ_0 is the lowest set-theoretical degree of complexity of formulæ). Let $\emptyset$ be a defined constant for the empty

[18] In set-theoretical texts they are indeed used synonymously with respect to ordinals.

[19] An anonymous referee worries that we might merely shift the mystery, since we did not explain why the properties of the hierarchy in the universe of hereditarily finite sets is mirrored when we move to the transfinite. (Note that this complaint applies only to the second part of our description of the mystery.) We do not fully agree, but it does not matter much even if this is indeed the case, since reducing a problem (in this case the external relations between number theory and set theory) to another one which might be easier to tackle (here the internal relations between parts of set theory) is usually considered as a step forward and as shedding new light on the problem.

[20] D'Alessandro [2018] argues that the reduction of arithmetic to set-theory is unexplanatory. The above examples challenge this view.

[21] The notion of an initial segment of N is the arithmetical counterpart of the set-theoretical notion of a transitive subclass of the set-theoretical universe V.

set and let ; be the defined binary function symbol from Previale [1994] and Kirby [2009], where the meaning of $x; y$ is $x \cup \{y\}$.[22] Consider:

- $N(x) := \forall y \in x; x(y = \emptyset \vee \exists z \in x.y = z; z)$
- $x < y := x \in y$

As far as we can see, this is impossible with Zermelo's definition. In any case, the *natural* definitions of N and $<$ based on Zermelo's account are not Δ_0, but Π_1-formulæ (in the above-mentioned hierarchy):[23]

- $N_Z(z) := \forall Z((\emptyset \in Z \wedge \forall y \in Z.\emptyset; y \in Z) \rightarrow z \in Z)$
- $x <_Z z := x \neq z \wedge \forall Z((x \in Z \wedge \forall y \in Z.\emptyset; y \in Z) \rightarrow z \in Z)$

Here it should be admitted that there are ways of using Δ_0-formulæ for defining other notions of "natural numbers" which would satisfy the very weak criteria given by Benacerraf.[24] Nevertheless, using von Neumann's approach, the definitions of N and $<$ are the simplest and shortest possible.[25]

3.5. **The strength of the background theory.** As noted by Benacerraf [1998] (see also condition (a) in Section 2), an adequate analysis of the concept of natural number does not merely consist of some definition of the natural numbers. In addition, one should be able to prove in some designated background theory that all the basic properties that we expect of the natural numbers follow from the definition. Here it is crucial to employ a set theory which is as weak as possible, since the definition (and the proofs) should be sufficiently robust to be acceptable not only to platonists, but also, e.g., to predicativists and finitists. Thus, the natural definitions given above of "number" and "$<$" in the case of Zermelo's "natural numbers" require the infinity axiom, and are impredicative. Hence they are inadmissible for finitists or predicativists. Things are different with von Neumann's account. It permits the development of number theory in a very weak subtheory of **ZF**, which we call **VBS** (for Very Basic System[26]). See Appendix A for details. **VBS** is equivalent to **PRA** — which following Tait [1981] is usually taken as the number-theoretic theory which represents finitism. Therefore **VBS** is finitistically (and so also

[22] This notation was introduced in order to provide a simple, purely functional language and axiomatization for $\mathcal{HF}$. Accordingly, these formulations are not strictly in the basic language of $=$ and $\in$, but only abbreviate formulæ in that language. However, it is easy to see that what they abbreviate are Δ_0-formulæ in the basic language which are not much longer than their abbreviations. It is interesting to note that in the language which is used in Previale [1994] and Kirby [2009], the definitions of Zermelo and von Neumann are the simplest possible ways of defining a successor function with the basic required properties. (The successor of x according to Zermelo is $\emptyset; x$, while it is $x; x$ according to von Neumann.)

[23] Using the content of Footnote 24, it is possible to give also a Σ_1 definition of these relations, which can be shown in **ZF** to be equivalent to the standard ones given below. (Note that the infinity axiom is needed in order to prove this equivalence.) Hence the notion of being a Zermelo number belongs in **ZF** to Δ_1. (This is not true in the weaker set theories considered below.)

[24] An interesting case in point, which might be the strongest competitor of von Neumann's definition from the complexity point of view, is the use of the finite initial segments of the set of Zermelo's natural numbers (i.e. what should be taken as Zermelo's finite cardinals):

- $N^c_Z(x) := \forall y \in x(y = \emptyset \vee \exists z \in x(y = \emptyset; z)) \wedge \exists z \in x.\emptyset; z \notin x$
- $x <^c_Z z := x \subseteq z \wedge x \neq z$

Steinhart [2002] calls these the "van Zermano" numbers (p. 349). For further discussion see also Ginammi [2019, pp. 285–286] and Ebels-Duggan [2022, p. 238].

[25] This is trivial in the case of $<$. We do not have a mathematical proof of this claim in the case of N, but it is very difficult to imagine that a simpler possible definition of N exists in the first-order language of $\{\in, =\}$; and it seems impossible to find one for which the *combined* lengths of the definitions of N and $<$ will not be bigger than that of the definitions given above.

[26] **VBS** is a rather weak subtheory of the basic system **BS** used in Devlin [1984].

predicatively) acceptable according to the criterion given in Feferman [1974].[27] (Actually, the predicativity of **VBS** is practically self-evident according to the very meaning of this notion, as it was originally introduced *in the context of set theory*.[28]) The fact that von Neumann's definition permits the derivation of basic arithmetic properties within a very weak set-theoretical background is yet another reason to prefer it over the original account of Zermelo.[29]

4. Objections: Metaphysical Irrelevance

In the previous section we have shown that von Neumann's account outrivals its most attractive set-theoretic competitor in the possession of desirable properties. At this point, the advocate of Benacerraf's argument might object that these properties are not arithmetical and thus extraneous to our number concept. Yet, so the objection continues, only arithmetical reasons are relevant for the metaphysical problem of determining the correct set-theoretic reduction of arithmetic. Hence, according to this objection the success of Benacerraf's argument is not affected by the reasons given in the previous section.[30]

This objection employs a specific interpretation of Benacerraf's argument which can be made precise by replacing the premises (R3) and (R4) in our presentation of the argument on page 3 by

- (R3*) For no $R \in \mathcal{A}$ do we have *arithmetical* reasons to believe that it is the correct set-theoretic reduction of arithmetic;
- (R4*) If $R \in \mathcal{A}$ is the correct set-theoretic reduction of arithmetic, then we have *arithmetical* reasons to believe that it is the correct set-theoretic reduction of arithmetic.

According to this line of defense, Benacerraf's argument is not affected by a rejection of premise (R3), but rather requires a rebuttal of its weaker variant (R3*).

Before we engage closely with this objection, we point to two immediate shortcomings of the resulting interpretation of Benacerraf's argument. Firstly, it relies on a further strengthening of its most contested premise, namely, (R4), which has been frequently rejected in the literature (see Section 2). Secondly, it is not at all clear how to distinguish arithmetical from non-arithmetical properties of reductions—let alone how to distinguish arithmetical from non-arithmetical reasons to prefer one reduction over another. This difficulty is due to the fact that the properties of our number concept and of set-theoretic reductions of arithmetic are commonly described in a set-theoretic language. Consider, for instance, the central claim that the natural numbers are well-ordered. Clearly, this claim is formulated set-theoretically. In what sense then does it capture an *arithmetical* property of the number concept? Even Fermat's method of infinite descent, a historical

---

[27] In more detail: **VBS** and **PRA** are easily seen to be reducible to each other in the sense of Feferman [1974]. The reductions use the well-known fact that $\mathcal{HF}$ and $\mathcal{N}$ can be turned into isomorphic structures, in the following sense: using bounded formulæ, one can define in the first-order order language of $\mathcal{N}$ a relation $\tilde{\in}$ and in the first-order order language of $\mathcal{HF}$ operations $\tilde{+}$, $\tilde{\times}$ and $\tilde{S}$ so that the structures $\langle N, S, +, \times, \tilde{\in}\rangle$ and $\langle HF, \tilde{S}, \tilde{+}, \tilde{\times}, \in\rangle$ are isomorphic to each other. See e.g. Fitting [2007].

[28] See e.g. Feferman [2005] for the history of predicativism and Avron [2010] for the principles underlying predicative (in the sense of Poincaré, Weyl and Feferman) set theory and for several corresponding systems. For a good survey of what axioms of **ZF** are usually taken as impredicative we refer the reader to the nLab entry on predicative mathematics: https://ncatlab.org/nlab/show/predicative+mathematics#illfounded_structures

[29] We do not know whether there are other reasonable set-theoretical definitions of the natural numbers for which very weak systems like **VBS** suffice.

[30] We are grateful to an anonymous referee for raising this worry and for helping us to clarify the material of this section.

precursor of mathematical induction, involves talk not only about numbers but also about infinite sequences or chains of numbers. Also Benacerraf [1998] uses set-theoretical vocabulary ("ω-sequence", "ancestral", etc.) when he describes the properties any account of arithmetic ought to satisfy (see the quotation on page 2). Recall that from Ernie's perspective, arithmetic is simply the study of those initial segments of On which satisfy certain set-theoretical conditions (see Section 3.1). Hence, for Ernie the distinction between arithmetical and set-theoretical properties is highly arbitrary if not outright incomprehensible. This all goes to say that at least at first glance, the boundary between arithmetical and set-theoretical properties appears to be rather murky. We now turn to Benacerraf's writings and its commentators for cues of how arithmetical and non-arithmetical properties may be distinguished.

An anticipation of the raised objection can be traced back to Benacerraf himself. For example, on one plausible reading of Benacerraf's argument, a property of a set-theoretic reduction of arithmetic is superfluous if and only if it is not grounded in our concept of number (see clause (5) on page 2). This position, dubbed the "deferential view" by Ebels-Duggan [2022], holds that what numbers are is exclusively determined by some pre-theoretic number concept of the folk, which is captured by the way we use numbers in our reasoning and practical life. According to Ebels-Duggan's (2022) reconstruction of Benacerraf's position, a set-theoretic reduction of arithmetic captures the folk number concept if it replicates the folk's arithmetical reasoning in a lossless, i.e., structure-preserving way (p. 227). According to the deferential view, everything else is extraneous. That is, *set-theoretic reductions of numbers are metaphysically on par as long as they preserve all relevant features of our number concept.* In particular, since von Neumann and Zermelo's accounts only "differ at places where there is no connection whatever between features of the accounts and our uses of the words in questions" [Benacerraf, 1965, p. 62], they do not differ in any metaphysically relevant sense. Thus, according to this sharpening of the raised objection, in Section 3 we merely showed that von Neumann's account is in *some* sense more adequate than its competitors. However, we failed to single out von Neumann's account *in terms of features which are grounded in our concept of number*, i.e., in any metaphysically relevant way.

We believe that this line of defense on behalf of Benacerraf's argument is unconvincing, for three reasons. Firstly, we reject the assumption that deference to the folk number concept outweighs all other theoretical virtues of a set-theoretic reduction (in doing so, we follow Paseau [2009]). Surely, the role of a set-theoretic reduction of numbers is to preserve all features relevant to arithmetic. But contra the deferential view, we belief that the success of a reduction should be also judged by the way it integrates arithmetic into our larger theoretical framework. Since currently the most overarching mathematical theory is provided by set theory, the adequacy of a reduction also depends on the way it coordinates arithmetical and set-theoretical concepts. Hence, according to our view, certain set-theoretic features of reductions which relate numbers and sets are metaphysically relevant. For example, von Neumann's reduction of arithmetic is more adequate than its alternatives, since it relates the number concept with the ordinal concept in the most natural way (see Section 3.1). Moreover, recall that we have shown in Section 3.3 that von Neumann's account explains certain connections between the fundamental notions of absoluteness in set theory and decidability in arithmetic. Since explanatoriness is an important desideratum for theory choice, this feature renders von Neumann's account metaphysically superior over other reductions which lack this explanatory power.

The advocate on behalf of Benacerraf might rejoin that our reply misconstrues reductions as mere explications. While von Neumann's reduction might provide the best way to mimic number talk in set-theoretic terms, set-theoretic concepts are entirely foreign to what numbers *really are.* As suggested by an anonymous referee, consider Fermat, who successfully referred to the natural numbers when he formulated his arithmetical theorems. Since set theory was only introduced more than 200 years later, the number concept of Fermat and his contemporaries arguably did not include any set-theoretical features. Hence, so the objection continues, it is very hard to believe that any set-theoretic facts or concepts determined the referents of Fermat's number terms. But then, the deferentialist concludes, set-theoretical features should be irrelevant to what the numbers are.

To begin with, this rejoinder seems to undermine the structuralist view about arithmetic, a philosophical position which solves Benacerraf's (1965) identification problem and which (to some extent) has been endorsed by Benacerraf himself. According to this view, numbers are positions in a number structure. Now, it is safe to assume that Fermat was completely unaware of the notion of a number structure (at least as it is used in the philosophical literature nowadays). But then, in analogy to the argument given above, talk about positions in a number structure should be irrelevant to what numbers really are. Hence, arithmetical structuralism is wrong.

What is more, we believe that this rejoinder disregards a crucial detail, namely, that the folk might have impoverished knowledge about the properties of their subjects of discourse. Hence, what numbers are might be determined by facts unknown even to competent participants of a number-discourse.[31] Fermat not only successfully referred to numbers without knowing set theory, but presumably he also successfully referred to water, without knowing that each of its molecules contains one oxygen and two hydrogen atoms. Yet, the analogous conclusion that water is not H_2O—given that the folk water concept included the belief that water is a chemical element and Fermat successfully referred to it while being completely unaware of the concepts of oxygen theory—is absurd.

Finally, the folk concept might not only be incomplete, but the folk may even hold false beliefs about the relevant subject matter. For example, Fermat's used real numbers without knowing what they are and without any justification of that usage. He certainly did not think of them in terms of their current rigorous definition(s), and there was no clear "folk notion" of a real number at his time. (Actually, there is no such "folk notion" today either.) Even worse, Fermat's analytic practice made essential reference to infinitesimally small entities and contained highly dubious derivations [Edwards, 2012]. Therefore there is no question that from a modern perspective, Fermat's real number concept was defective. Yet, Fermat successfully participated in the analytic discourse, correctly calculated tangent lines of curves, etc. Hence, according to the rejoinder's approach, what real number *really are* is determined by the conceptual resources which were available to Fermat. But clearly it would be absurd to retain an ontological commitment to infinitesimals on these grounds. The upshot is that metaphysical accounts of water and arithmetic should not try to preserve the folk concept at all costs. Rather, the questions of what water and numbers are should be treated within the context of our comprehensive scientific and mathematical theories respectively. By rejecting the deferentialist view along these lines, we are no longer exclusively committed to a faithful interpretation of a potentially misinformed or even defective folk number concept. As a result, also theoretical features of

[31]Here we follow Paseau [2009] in his response to a related objection due to Field [1974].

reductions which are not grounded in the folk number concept can provide metaphysically relevant reasons to single out these reductions as correct.

The second rejoinder to the above objection is dialectical. Recall that according to the deferential position, set-theoretic reductions of numbers are metaphysically on par as long as they preserve all relevant features of our number concept. So far, we have not said what the "relevant features of our number concept" are. A common approach is to take all relevant features of our number concept to be codified in the consequences of second-order Peano Arithmetic $\mathbf{PA}_2$.[32] Structure-preservation thus amounts to consequence-preservation (see, e.g., [Ebels-Duggan, 2022, p. 227]). On this view all models of $\mathbf{PA}_2$ are metaphysically on par, thus barring any further escape from Benacerraf's identification argument.

While $\mathbf{PA}_2$ may be indeed our best arithmetical theory, there is no reason to believe that its consequences capture all aspects of the folk number concept. For instance, $\mathbf{PA}_2$ does not prove that numbers are not chairs and do not contain elements. Yet, we pre-theoretically believe that numbers are neither chairs nor contain elements and these beliefs may inform our use of number words. In the words of Balaguer [1998], "it is built into our conception of the natural numbers that they do not have members [and] that they cannot be sat on" (p. 66). Moreover, even though $\mathbf{PA}_2$'s consistency statement is not among its consequences (if $\mathbf{PA}_2$ is treated as a formal system), number theorists surely believe that their number concept is consistent. As we see it, the deferentialist has two options to respond. The first is to take serious her commitment to a lossless interpretation of the folk's number talk. But then no set-theoretic reduction of arithmetic captures the folk number concept, since the feature of not having members cannot be preserved by any set-theoretic reduction. Hence, the first option immediately yields the conclusion that numbers are not sets, even without the need of pressing Benacerraf's identification argument. The second option is to relax this commitment by dismissing *being a chair* or *containing an element* as descriptions of the "internal" nature of objects and thus irrelevant to arithmetic (see, e.g., Resnik [1981]). But by dismissing "internal" features of numbers as metaphysically irrelevant, the deferentialist adopts at the outset a structuralist conception of arithmetic. Hence, both options are heavily biased, if not entirely question-begging, against the target of Benacerraf's argument, to wit, reductionism. In particular, there is no reason for a realist to hold a structuralist view about arithmetic.[33] Hence, construed in this way, Benacerraf's argument against reductionism loses much of its force.

We now turn to our third and most important rejoinder to the above objection. For argument's sake we grant that the deferential position is overall correct, i.e., that set-theoretic reductions of numbers are metaphysically on par as long as they preserve all relevant features of our number concept. However, we argue that there are plausible explications of the "relevant features of our number concept" such that the opponent's challenge can be met. That is, von Neumann's account can be singled out as more adequate than its alternatives in terms of features which are grounded in our concept of number.

We extract two reasons to reject the claim that all relevant features of our number concept are codified in the consequences of $\mathbf{PA}_2$ from Benacerraf's own writings. Firstly,

[32]Most authors, including Benacerraf, do not specify whether the number concept should be explained in terms of first-order or second-order Peano arithmetic. Moreover, $\mathbf{PA}_2$ can be associated either with standard or Henkin semantics respectively. However, the subsequent arguments are not affected by these choices.

[33]Similar points have been made by Ruffino [2001], Clarke-Doane [2008] and Ebels-Duggan [2022, footnote 19].

Benacerraf [1965] argues that a relevant feature of our number concept is the recursiveness of the less-than relation $<$ on numbers, which enables us "to tell in a finite number of steps which of two numbers is the greater" (p. 52). According to Benacerraf, any set-theoretic reduction of arithmetic which captures the folk number concept thus has to preserve this computational feature, i.e., has to result in a model of $\mathbf{PA}_2$ whose less-than relation is recursive [Benacerraf, 1965, p. 53].[34] This shows that even by Benacerraf's own standards we can privilege some models of $\mathbf{PA}_2$ over others in terms of features which are grounded in our number concept. The analysis of our number concept can be pursued further along similar lines. For instance, we may observe that our arithmetical practice not only relies on the computational simplicity of basic arithmetical notions such as $<$, but also relies on the simplicity of our basic arithmetical vocabulary. Hence, in complete analogy to Benacerraf's reasoning above, we may require that set-theoretic reductions of arithmetic which capture the folk number concept also preserve complexity-theoretic simplicity. As we have shown in Section 3.4, von Neumann's account fares better in this respect than its competitors. Once again, this reasoning only involves features which are grounded in our number concept.

The second reason to reject the claim that all relevant features of our number concept are codified in the consequences of $\mathbf{PA}_2$ is that it does not account for arithmetical applications. For example, Benacerraf [1965] takes the cardinality relation between numbers and sets as an essential feature of the number concept: "[t]o count the members of a set is to determine the cardinality of the set. It is to establish that a particular relation C obtains between the set and one of the numbers" (p. 50, see also condition (c) on page 2). Hence, the number concept includes features which relate numbers and sets in specific ways. Once again, this shows that $\mathbf{PA}_2$ is insufficient to capture all relevant features of our number concept. Rather, an apparatus is required which additionally incorporates a theory of sets together with certain facts about cross-sortal relations between numbers and sets. Moreover, we may require that set-theoretic reductions of arithmetic which capture the number concept also preserve principles which govern the specific ways cardinal numbers are correlated to sets. For example, following Wright [1983], Ebels-Duggan argues that reductions of arithmetic should preserve certain identity criteria specific to cardinal numbers:

> Part of the number concept includes how to distinguish or identify numbers as applied in cardinality judgments—identity criteria specific to numbers. And we have the same for the application of sets. What is needed to determine number-set identity ... is something that coordinates the individual conditions of identity for each kind of object. Identity criteria for the application of sets which are also numbers must be the same as that for the application of numbers, full stop. [Ebels-Duggan, 2022, p. 237]

In other words, a set-theoretic reduction of arithmetic only captures our number concept, if different numbers-as-sets are not equipollent. Using this reasoning we can privilege some models of $\mathbf{PA}_2$ over others in terms of features which are grounded in our number concept. In particular, von Neumann's account captures our number concept construed in this way, while Zermelo's does not. By drawing on further aspects which are central to the cardinal number concept, analogous reasoning yields even more narrow classes of models of $\mathbf{PA}_2$. In particular, the reasons we provide in Section 3.2 single out von

[34] This condition is slightly problematic, since recursiveness only applies directly to numbers, while domains of models of $\mathbf{PA}_2$ may contain any kind of objects (see Halbach and Horsten [2005]). Benacerraf [1996] later drops this condition, but for very different reasons.

Neumann's account merely in terms of features which are grounded in our concept of number—suitably construed.

To sum up, the objection contesting the metaphysical significance of our reasons is blocked. In particular, we have shown that the deferential position is not very stable as there are different plausible ways to explicate the folk number concept. Moreover, the view that all relevant features of our number concept are codified in the consequences of $\mathbf{PA}_2$ is not compatible with Benacerraf's own claims. Finally, there are plausible ways to construe our number concept such that von Neumann's account can be singled out only in virtue of features which are grounded in the concept of number. Note that even if the reasons in Section 3 do not provide *compelling* grounds to privilege von Neumann's account, they afford at least some grounds to believe it more likely to be correct than its competitors. This is all that is needed to reject Premiss (R3) and thus to rebut Benacerraf's anti-reductionist argument.

Appendix A. Ordinals and Natural Numbers in Minimal Set Theories

A.1. The Very Basic System VBS.

Extensionality (Ex):

$$\forall z(z \in x \leftrightarrow z \in y) \to x = y$$

Empty Set (Em):

$$\exists Z \forall x(x \notin Z)$$

Singleton (Si):

$$\forall y \exists Z \forall x(x \in Z \leftrightarrow x = y)$$

Binary Union (BU):

$$\forall x \forall y \exists Z \forall w(w \in Z \leftrightarrow w \in x \vee w \in y)$$

Bounded $\in$-induction ($B\!-\in\!-ind$):

$$\forall x((\forall y \in x \varphi\{y/x\}) \to \varphi) \to \forall x \varphi \quad \varphi \text{ bounded}$$

Note. The first four axioms of **VBS** provide the minimal set theory which allows to define every element of $\mathcal{HF}$ (the collection of hereditarily finite sets). They also allow to introduce into the language of set theory the constant $\emptyset$ (for the empty set), the unary function symbol $\{-\}$ (for singletons) and the binary function symbol $\cup$ (for binary union). The scheme of $B\!-\in\!-ind$ is the minimal principle which is sufficient for fruitful work within $\mathcal{HF}$, but is valid in the whole universe of **ZF**.

Proposition.

(1) $\vdash_{\mathbf{VBS}} \forall x_0 \forall x_1 \in x_0 \forall x_2 \in x_1 \cdots \forall x_n \in x_{n-1}.x_0 \notin x_n$ $(n \geq 0)$
(2) $\vdash_{\mathbf{VBS}} \forall x(\exists y(y \in x) \to \exists y(y \in x \wedge \forall z(z \in y \to z \notin x))$ (Foundation axiom)

A.2. von Neumann's Ordinals in VBS.

Definition.

- $\mathit{Transitive}(x) := \forall y \in x \forall z \in y.z \in x$
- $\mathit{Linear}(x) := \forall y \in x \forall z \in x.y \in z \vee y = z \vee z \in y$
- $\mathit{Ordinal}(x) := \mathit{Transitive}(x) \wedge \mathit{Linear}(x)$
- $S(x) = x \cup \{x\}$. Hence:
  - $y \in S(x) := y = x \vee y \in x$
  - $y = S(z) := \forall u(u \in y \leftrightarrow u = z \vee u \in z)$
- $x \leq y := x \in y \vee x = y$

Proposition. The following are provable in **VBS**:

(1) $Ordinal(x) \wedge y \in x \to Ordinal(y)$
(2) $Ordinal(\emptyset)$
(3) $Ordinal(x) \to Ordinal(S(x))$
(4) $S(x) = S(y) \to x = y$
(5) $Ordinal(x) \wedge Ordinal(y) \to (x \leq y \leftrightarrow x \subseteq y)$
(6) $Ordinal(x) \wedge Ordinal(y) \to (x \leq y \leftrightarrow x \in S(y))$
(7) $Ordinal(x) \to x = \emptyset \vee \emptyset \in x$
(8) $Ordinal(x) \wedge y \in x \to x = S(y) \vee S(y) \in x$
(9) $Ordinal(x) \wedge Ordinal(y) \to y \in x \vee x = y \vee x \in y$

Theorem. The following are provable in **VBS**:

(1) The class of ordinals is well-ordered by $\in$.
(2) Every ordinal is well-ordered by $\in$.

Theorem. Given a bounded formula φ, $\mathbf{VBS}$ proves that if the class of ordinals that satisfy φ is not empty, then it has a minimal element.

A.3. von Neumann's Natural Numbers in VBS.

Definition.

- $N(x) := \forall y \in S(x)(y = \emptyset \vee \exists z \in x.y = S(z))$
- $N^\star(x) := Ordinal(x) \wedge \forall y \in S(x)(y = \emptyset \vee \exists z \in y.y = S(z))$

Note. $N^\star$ is the standard definition by a bounded formula of the property of being a natural number (without an appeal to the axiom of infinity). Its advantage is that it directly defines the natural numbers as a special sort of ordinals. N, on the other hand, is the simplest such definition we are aware of. It is obtained by a single small change in the second conjunct of the definition of $N^\star(x)$. As noted below, the two definitions are equivalent in $\mathbf{VBS}$.

Proposition. The following axioms of Peano are theorems of **VBS**:

(1) $N(\emptyset)$
(2) $\forall x.N(x) \to N(S(x))$
(3) $\forall x.S(x) \neq \emptyset$
(4) $\forall x \forall y.S(x) = S(y) \to x = y$
(5) $Ind_N[\varphi]$ for every bounded formula φ, where $Ind_N[\varphi]$ is the following formula:

$$\varphi\{\emptyset/x\} \wedge \forall x(\varphi \to \varphi\{S(x)/x\}) \to \forall x(N(x) \to \varphi)$$

Corollaries.

(1) $\vdash_{\mathbf{VBS}} \forall x.N(x) \to Ordinal(x)$
(2) $\vdash_{\mathbf{VBS}} \forall x \forall y.N(x) \wedge y \in x \to N(y)$
(3) $\vdash_{\mathbf{VBS}} \forall x.N(x) \leftrightarrow N^\star(x)$

Acknowledgements

We would like to thank an anonymous referee for valuable comments and suggestions. This work was supported by Len Blavatnik and the Blavatnik Family foundation. Balthasar Grabmayr would also like to thank the Azrieli Foundation for their support.

References

A. Avron. Constructibility and decidability versus domain independence and absoluteness. *Theoretical Computer Science*, 394(3):144–158, 2008. URL https://doi.org/10.1016/j.tcs.2007.12.008.

A. Avron. A new approach to predicative set theory. In R. Schindler, editor, *Ways of Proof Theory*, volume 2 of *Onto Series in Mathematical Logic*, pages 31–63. onto verlag, 2010. URL https://doi.org/10.1515/9783110324907.31.

M. Balaguer. Non-uniqueness as a non-problem. *Philosophia Mathematica*, 6(1):63–84, 1998. URL https://doi.org/10.1093/philmat/6.1.63.

P. Benacerraf. What numbers could not be. *The Philosophical Review*, 74(1):47–73, 1965. URL https://doi.org/10.2307/2183530.

P. Benacerraf. Recantation or any old ω-sequence would do after all. *Philosophia Mathematica*, 4(2):184–189, 1996. URL https://doi.org/10.1093/philmat/4.2.184.

P. Benacerraf. What mathematical truth could not be–I. In M. Schirn, editor, *The Philosophy of Mathematics Today*, pages 33–75. Clarendon Press, 1998.

Paul Bernays. A system of axiomatic set theory–part II. *The Journal of Symbolic Logic*, 6 (1):1–17, 1941. URL https://doi.org/10.2307/2267281.

J. Clarke-Doane. Multiple reductions revisited. *Philosophia Mathematica*, 16(2):244–255, 2008. URL https://doi.org/10.1093/philmat/nkm034.

I. M. Copi. The Burali-Forti paradox. *Philosophy of Science*, 25(4):281–286, 1958.

K. Devlin. *The Joy of Sets*. Undergraduate Texts in Mathematics. Springer, 1993. URL https://doi.org/10.1007/978-1-4612-0903-4.

K. J. Devlin. *Constructibility*. Perspectives in Mathematical Logic. Springer-Verlag, 1984. URL https://doi.org/10.1017/9781316717219.

W. D'Alessandro. Arithmetic, set theory, reduction and explanation. *Synthese*, 195(11): 5059–5089, 2018. URL https://doi.org/10.1007/s11229-017-1450-8.

S. C. Ebels-Duggan. On number-set identity: A study. *Philosophia Mathematica*, 30(2): 223–244, 2022. URL https://doi.org/10.1093/philmat/nkac008.

C.H.J. Edwards. *The Historical Development of the Calculus*. Springer Study Edition. Springer New York, 2012. URL https://doi.org/10.1007/978-1-4612-6230-5.

A. S. Esenin-Volpin. The ultra-intuitionistic criticism and the antitraditional program for foundations of mathematics. In A. Kino, J. Myhill, and R. Vesley, editors, *Intuitionism and Proof Theory*, pages 3–45. North-Holland, Amsterdam, 1970. URL https://doi.org/10.1016/S0049-237X(08)70738-9.

S. Feferman. Predicatively reducible systems of set theory. In D. S. Scott and T. Jech, editors, *Axiomatic Set Theory* (Proceedings of Symposia in Pure Mathematics XIII), volume 2, pages 11–32. American Mathematical Society (providence), 1974.

S. Feferman. Predicativity. In S. Shapiro, editor, *The Oxford Handbook of the Philosophy of Mathematics and Logic*, pages 590–624. Oxford University Press, 2005.

H. Field. Quine and the correspondence theory. *Philosophical Review*, 83(2):200–228, 1974. doi: 10.2307/2184138. URL https://doi.org/10.2307/2184138.

M. Fitting. *Incompleteness in the Land of Sets*. College Publications, 2007.

G. Frege. *Die Grundlagen der Arithmetik – Eine logisch mathematische Untersuchung über den Begriff der Zahl*. Verlag von Wilhelm Koebner, Breslau, 1884.

M. Ginammi. In defense of Benacerraf's multiple-reductions argument. *Philosophia Mathematica*, 27(2):276–288, 2019. URL https://doi.org/10.1093/philmat/nkz009.

V. Halbach and L. Horsten. Computational structuralism. *Philosophia Mathematica*, 13 (2):174–186, 2005. URL https://doi.org/10.1093/philmat/nki021.

M. Hallett. *Cantorian Set Theory and Limitation of Size*. Oxford logic guides. Clarendon Press, Oxford, 1986.

L. Kirby. Finitary set theory. *Notre Dame Journal of Formal Logic*, 50(3):227–244, 2009. URL https://doi.org/10.1215/00294527-2009-009.

K. Kunen. *The Foundations of Mathematics*. Number 19 in Studies in Logic. College Publications, 2009.

P. Maddy. *Realism in Mathematics*. Oxford University Press, Oxford, UK, 1990. URL https://doi.org/10.1093/019824035X.001.0001.

C. McLarty. Numbers can be just what they have to. *Noûs*, 27(4):487–498, 1993. URL https://doi.org/10.2307/2215789.

B. M. Mount. Antireductionism and ordinals. *Philosophia Mathematica*, 27(1):105–124, 2019. doi: 10.1093/philmat/nky025. URL https://doi.org/10.1093/philmat/nky025.

A. Paseau. Reducing arithmetic to set theory. In O. Bueno and Ø. Linnebo, editors, *New Waves in Philosophy of Mathematics*, pages 35–55. Palgrave-Macmillan, 2009. URL https://doi.org/10.1057/9780230245198_3.

F. Previale. Induction and foundation in the theory of hereditarily finite sets. *Archive for Mathematical Logic*, 33:213–241, 1994. URL https://doi.org/10.1007/BF01203033.

M. D. Resnik. Mathematics as a science of patterns: Ontology and reference. *Noûs*, 15(4): 529–550, 1981. doi: 10.2307/2214851. URL https://doi.org/10.2307/2214851.

M. D. Resnik. *Mathematics as a Science of Patterns*. Oxford University Press, 1997. URL https://doi.org/10.1093/0198250142.001.0001.

P. Rouilhan. Identification and transportability: Another moral for Benacerraf's parable of Ernie and Johnny. In F. Pataut, editor, *Truth, Objects, Infinity: New Perspectives on the Philosophy of Paul Benacerraf*, volume 28, pages 159–176. Springer, 2016. URL https://doi.org/10.1007/978-3-319-45980-6_8.

M. Ruffino. Numbers and sets. *Kriterion: Journal of Philosophy*, 42(104):130–146, 2001. URL https://doi.org/10.1590/S0100-512X2001000200007.

S. Shapiro. What numbers could be; what objects could be. In F. Pataut, editor, *Truth, Objects, Infinity: New Perspectives on the Philosophy of Paul Benacerraf*, volume 28, pages 177–191. Springer, 2016. URL http://doi.org/10.1007/978-3-319-45980-6_9.

M. Steiner. *Mathematical Knowledge*. Cornell University Press, 1975.

E. Steinhart. Why numbers are sets. *Synthese*, 133(3):343–361, 2002. URL https://doi.org/10.1023/A:1021266519848.

W. Tait. Finitism. *Journal of Philosophy*, 78:524–556, 1981. URL https://doi.org/10.2307/2026089.

D. Van Dantzig. Is $10^{10^{10}}$ a finite number? *Dialectica*, 9:273–277, 1956. URL https://doi.org/10.1111/j.1746-8361.1955.tb01332.x.

A. Visser. Can we make the Second Incompleteness Theorem coordinate free? *Journal of Logic and Computation*, 21(4):543–560, 2011. URL https://doi.org/10.1093/logcom/exp048.

A. Visser. The second incompleteness theorem: Reflections and ruminations. In L. Horsten and P. Welch, editors, *Gödel's Disjunction: The Scope and Limits of Mathematical Knowledge*, pages 67–91. Oxford University Press, 2016. URL https://doi.org/10.1093/acprof:oso/9780198759591.003.0003.

J. von Neumann. Zur Einführung der transfiniten Zahlen. *Acta Scientiarum Mathematicarum (Szeged)*, 1(4):199–208, 1923.

J. von Neumann. Über die Definition durch transfinite Induktion und verwandte Fragen der allgemeinen Mengenlehre. *Mathematische Annalen*, 99:373–391, 1928. URL https://doi.org/10.1007/BF01459102.

L. Wetzel. That numbers could be objects. *Philosophical Studies*, 56(3):273–92, 1989. URL https://doi.org/10.1007/BF00354365.

N. P. White. What numbers are. *Synthese*, 27(1-2):111–124, 1974. URL https://doi.org/10.1007/BF00660892.

C. Wright. *Frege's Conception of Numbers as Objects*. Aberdeen University Press, 1983.

E. Zermelo. Untersuchungen über die Grundlagen der Mengenlehre. i. *Mathematische Annalen*, 65:261–281, 1908. URL https://doi.org/10.1007/BF01449999.

Ernst Zermelo. Über Grenzzahlen und Mengenbereiche: Neue Untersuchungen über die Grundlagen der Mengenlehre. *Fundamenta Mathematicæ*, 16:29–47, 1930.