Arithmetic, Set Theory, Reduction and Explanation

William D'Alessandro

May 20, 2017

Forthcoming in *Synthese*.

Abstract

Philosophers of science since Nagel have been interested in the links between intertheoretic reduction and explanation, understanding and other forms of epistemic progress. Although intertheoretic reduction is widely agreed to occur in pure mathematics as well as empirical science, the relationship between reduction and explanation in the mathematical setting has rarely been investigated in a similarly serious way. This paper examines an important and well-known case: the reduction of arithmetic to set theory. I claim that the reduction is unexplanatory. In defense of this claim, I offer some evidence from mathematical practice, and I respond to contrary suggestions due to Steinhart, Maddy, Kitcher and Quine. I then show how, even if set-theoretic reductions are generally not explanatory, set theory can nevertheless serve as a legitimate and successful foundation for mathematics. Finally, some implications of my thesis for philosophy of mathematics and philosophy of science are discussed. In particular, I suggest that some reductions in mathematics are probably explanatory, and I propose that differing standards of theory acceptance might account for the apparent lack of unexplanatory reductions in the empirical sciences.

> [I]n the philosophy of science the notions of explanation and reduction have been extensively discussed, even in formal frameworks, but there exist few successful and exact applications of the notions to actual theories, and, furthermore, any two philosophers of science seem to think differently about the question of how the notions should be reconstructed. On the other hand, philosophers of mathematics and mathematicians have been successful in defining and applying various exact notions of reduction (or interpretation), but they have not seriously studied the questions of explanation and understanding.[1]

[1] [Rantala 1992], 47.

Rantala's observation, though now almost a quarter-century old, remains pretty much on target. In spite of the recent surge of interest in various aspects of mathematical explanation[2], philosophers have yet to turn their attention to the explanatory (and more broadly epistemological) dimensions of intertheoretic reduction in mathematics.

I hope the rest of the paper will show that this neglect is unwarranted. As philosophers of science have long known in the empirical context, understanding the link between reduction and explanation is a vital part of understanding theory succession, mathematical progress, and the nature and purpose of foundations of mathematics. I aim to illustrate these points by examining a particularly important single case: the reduction of arithmetic to set theory. Although a number of authors have expressed views about the explanatory significance of this reduction, the issues involved haven't yet been weighed as carefully as they deserve to be. This paper sets out to do so, and to draw some considered conclusions.

My thesis is that the reduction of arithmetic to set theory is unexplanatory. After clarifying this claim and giving some positive reasons to believe it, I devote the larger part of the paper to showing that the most serious arguments to the contrary—due to Steinhart, Maddy, Kitcher, and Quine—are unsuccessful. Finally, I discuss some consequences of the view for philosophy of mathematics, philosophy of science and the theory of reduction.

1 Intertheoretic reduction and explanation in mathematics

We might start by asking what it means for one theory to be reducible to another (either in general, or in the mathematical context specifically). This is an important question, but I won't say very much about it here. For present purposes I take it as a datum that arithmetic is reducible to set theory; I'm more confident in the correctness of this claim than I am about the details of any particular theory of reducibility.

Nevertheless, I'm inclined to think that the concept of reduction is at least closely related to the model-theoretic notion of (*relative*) *interpretability*. Informally, a theory T_1 is interpretable in a theory T_2 if there's a theoremhood-preserving translation function from the language of T_1 to the language of T_2. (The rather unwieldy formal definition won't be necessary here; see e.g. [Feferman 1960], 49 or [Niebergall 2000], 30-31.)

Identifying reducibility with interpretability has several advantages. For one, interpretability is quite similar to the classic Nagel-style understanding of reduction in philosophy of science[3]—an approach that

[2]See for instance [Steiner 1978], [Mancosu 2001], [Weber & Verhoeven 2002], [Hafner & Mancosu 2005], [Tappenden 2005], [Avigad 2008], [Lange 2009], [Baker 2010], [Lange 2010], [Lange 2014], [Pincock 2015].

[3]See [Nagel 1961].

continues to find plenty of adherents. Note that for Nagel, however, "Reduction.... is the explanation of a theory or a set of experimental laws established in one area of inquiry, by a theory usually though not invariably formulated for some other domain."[4] I don't follow Nagel in taking explanatoriness to be partly definitional of reduction. But it seems likely that Nagel thought this only because he was a deductivist about explanation (in the manner of [Hempel & Oppenheim 1948])[5], and pretty much everyone now agrees that deductivism is false. So the disagreement doesn't trouble me much. (In any case, those who insist on viewing intertheoretic reduction as necessarily involving explanation can simply interpret me as denying that arithmetic is reducible to set theory.)

The interpretability criterion also gives the right results about central cases, including the case of Peano arithmetic and ZF set theory. Finally, one can find in [Niebergall 2000] a sophisticated defense of relative interpretability as superior to other proposed criteria for reduction.

In any case, nothing much about the issues to be discussed below hangs on details about the nature of reducibility. I assume only the uncontroversial premise that arithmetic is reducible to set theory in some interesting sense.

Another point worth addressing: if, as I've claimed, intertheoretic reduction isn't by definition an explanatory relation, why think the two notions are connected at all? To start with, note that the claim in question is quite weak. It leaves open the possibility, for instance, that all cases of reduction are *actually* explanatory, as a matter of empirical fact. (Or, alternatively, that this is true of all sufficiently natural reductions, or of some other relatively large and interesting class of examples.) And indeed, the evidence from philosophy of science strongly suggests the existence of a close connection of some such kind. Consider the (purported) reductions of classical thermodynamics to statistical mechanics, of Mendelian inheritance theory to biochemistry, of Kepler's theory of planetary motion to Newtonian mechanics, and so on.[6] In each such case, the viewpoint associated with the reducing theory yields substantial explanatory insights about the subject matter of the reduced theory. These insights can take the form of clarifications of the nature and properties of the entities postulated by the reduced theory (e.g., the identification of genes with regions of DNA), of the reasons why the laws of the reduced theory hold and the circumstances under which they do (e.g., the explanation of the Second Law of Thermodynamics in terms of the improbability of entropy-lowering microstate evolutions), of the place of the reduced theory in a larger conceptual or nomic framework, and so on.

Examples from empirical science thus suggest that explanatory reductions are the norm, if not

[4][Nagel 1961], 338.

[5]See [van Riel 2011] for a discussion of Nagel's deductivism and its influence on his account of reduction.

[6]Whether these cases count as bona fide intertheoretic reductions has been the subject of debate. I don't want or need to take a stand on this issue here—letting go of some (or even many) individual examples is no problem for my purposes, as long as there are some genuine instances of reduction that look something like the ones philosophers have discussed.

in fact the rule. A priori, there seems to be no obvious reason to expect this pattern to fail in the mathematical setting. It would be an interesting result, then, if a central and uncontroversial case of intertheoretic reduction in mathematics turned out to lack explanatory value. The next section argues that the reduction of arithmetic to set theory is such a case.

2 Arithmetic and set theory

2.1 The view to be defended

I claim that the reduction of arithmetic to set theory is unexplanatory.[7] I'll have much to say in defense of this claim below, but it's worth noting immediately that the thesis is highly plausible as a claim about other bits of mathematics, too. Perhaps the best example is the set-theoretic reduction of the ordered pair. Recall that, according to the now-standard definition (due to Kuratowski), the ordered pair (x, y) is identified with the set $\{\{x\}, \{x, y\}\}$. It's easy to see that the set in question possesses "the characteristic property of ordered pairs", namely the feature that $(x_1, y_1) = (x_2, y_2)$ just in case $x_1 = y_1$ and $x_2 = y_2$. So the reduction of the ordered pair is unobjectionable from a logical point of view. Nevertheless, it seems clear that the reduction gives us no new insight or understanding about the nature of order. Rather than being an explanatory achievement, we want to say, the identification $(x, y) = \{\{x\}, \{x, y\}\}$ is a stipulation designed to solve a technical problem—a clever and convenient stipulation that does its job well, granted, but not the sort of thing that leaves us epistemically better off than before. As Michael Potter's set theory textbook puts it,

> $\{\{x\}, \{x, y\}\}$ is a single set that codes the identities of the two objects x and y, and it is for that purpose that we use it; as long as we do not confuse it with the genuine ordered pair (if such there is), no harm is done. In other words, the ordered pair as it is used here is to be thought of only as a technical tool to be used within the theory of sets and not as genuinely explanatory of whatever prior concept of ordered pair we may have had.[8]

Randall Dipert comments similarly that

> the defining quality of an ordered pair is quite clear without a set-theoretic formulation. Namely, two ordered pairs are identical just when their respective members are identical. It *was* quite clear before the advent of sets and set-talk. ...In terms of increased precision and

[7]For the sake of variety or readability, I'll sometimes use expressions like "the reduction of numbers to sets" below. This is just loose talk. Unless the context indicates otherwise, such expressions always mean the same as "the reduction of arithmetic to set theory"—I don't mean to invoke some alternative, ontological notion of reduction by speaking of objects instead of theories.

[8][Potter 2004], 65.

> increased clarity, the set-theoretic reduction of ordered pair added nothing, historically and philosophically.[9]

Analogous remarks may apply to other familiar set-theoretic reductions. Take functions, for example. In set theory a function f is identified with the set of ordered pairs $\{(a_1, b_1), (a_2, b_2), \ldots\}$, where a_i and b_i are the elements of the domain and image of f, respectively, and where $f(a_i) = b_i$. Again, nobody disputes the usefulness of this maneuver. But its usefulness arguably lies in its adequacy as a technical device rather than its power to explain anything about functions.[10] The set theorist Yiannis Moschovakis notes in this vein that

> [t]his "identification" of a function [with a set of ordered pairs] has generated some controversy, because we have natural "operational" intuitions about functions and by "function" we often mean a formula or a rule of computation. ...There is no problem with this if we keep clear in our minds that the "definition" [just mentioned] does not replace the intuitive notion of function but only represents it within *set theory*, faithfully for the uses to which we put this notion *within set theory*.[11]

What's going on in such cases seems to be roughly this. At the beginning of the story, before the question of reduction enters the picture, we have some mathematical concept C with which we're epistemically satisfied in important respects—that is, a concept whose conditions of application aren't in dispute (at least in ordinary cases), and whose basic properties we feel we understand more or less fully.[12] We become convinced at some point of the desirability of set-theoretic foundations (either for mathematics as a whole, or for the particular subject matter to which C belongs). The goal is then to find a set-theoretic construction that can stand in for objects of type C. This is done by identifying the characteristic logical properties of C-type objects and then producing a set (or collection of sets) with the desired properties. If there are many different candidates, then judgments about relative simplicity, convenience, and so on are used to narrow the field down to a single structure S (or at most a small handful of such structures). Given a sufficiently robust consensus, S then comes to be viewed as the canonical representation of C-type objects within set theory. Whatever its virtues *qua* set-theoretic surrogate, though, the purpose of S isn't to explain anything or generate new understanding about C-type objects. Rather, S does its

[9] [Dipert 1982], 366-367. Emphasis in original.

[10] The case of functions is more complicated and controversial than that of the ordered pair. Identifying a function with its set-theoretic graph, for instance, allows one to prove nontrivial theorems that weren't available to earlier mathematicians operating with less clear or rigorous conceptions. (E.g., the theorem that there exist uncomputable functions.) The thesis that the reduction of functions is unexplanatory is probably most plausible when restricted to some class of "classical" functions, for instance differentiable or continuous functions of a real or complex variable.

[11] [Moschovakis 2006], 40. Italics in original.

[12] This isn't to suggest that the correct understanding of the concept in question need have been easy to determine. The modern notion of function, for instance, evolved slowly and rather tortuously from its roots in 17th-century geometry and analysis to its current general form. For a brief overview of this history, see [Kleiner 1989].

job by hewing faithfully to the defining C-properties. (And it's precisely because we already had an adequate epistemic handle on these properties that we were able to specify them in advance, and to use them to guide our choice of S.)

This, I think, is pretty clearly the correct story about the reduction of ordered pairs and functions to sets. I maintain that it's also the correct story about the reduction of arithmetic. Viewing the natural numbers and arithmetical operations as set-theoretic constructs is a useful expedient for certain purposes, but doing so has no particular explanatory value.

Before saying more, we might as well recall how the reduction is supposed to go. As is well known, there are many set-theoretic structures that model the Peano axioms, and hence many ways to reduce arithmetic to set theory. Historically, the systems of Zermelo and von Neumann have been the most influential. Of the two, the von Neumann approach is usually thought to have a preponderance of nice properties, and hence has become the default choice of most mathematicians.

A model of Peano arithmetic is a structure $(S, f, e, \prec)$, where S is a set (interpreted as the collection of *natural numbers*), f is a unary function (interpreted as *successor*), e is a constant (interpreted as *zero*), and $\prec$ is a binary relation (interpreted as *less than*). Hence the first step in specifying a reduction of arithmetic to set theory is to choose a collection of sets S to serve as the natural numbers. In the von Neumann system, these sets are the finite von Neumann ordinals, and the correspondence is

$$\begin{aligned} 0 &= \emptyset, \\ 1 &= \{\emptyset\}, \\ 2 &= \{\emptyset, \{\emptyset\}\}, \\ 3 &= \{\emptyset, \{\emptyset\}, \{\emptyset, \{\emptyset\}\}\}, \end{aligned}$$

and in general $n + 1 = n \cup \{n\}$. This also supplies the definitions of the successor function and the zero constant. For less than, $\prec$ is identified with the proper subset relation $\subset$. It isn't hard to show that the finite von Neumann ordinals, equipped with these definitions, satisfy the Peano axioms. (The Zermelo system uses the correspondence $0 = \emptyset$, $n + 1 = \{n\}$, and identifies $\prec$ with the ancestral of set membership. Zermelo's approach is rarely used today, though, so I'll focus on the von Neumann ordinals from now on.)

My view, then, is that correspondences of this sort between arithmetic and set theory are bona fide intertheoretic reductions which are nevertheless unexplanatory. By this I mean that they don't explain (or substantially contribute to explaining) anything about the natural numbers and their arithmetic.[13]

[13]Anything except, perhaps, certain kinds of possible explanations that are too trivial, insubstantial, and *pro forma*

Put another way, there's no explanatory benefit to *viewing numbers as sets* or *supposing that numbers are sets*—in the way that there is such a benefit to viewing a sample of radium as an assemblage of Ra atoms, say, or temperature as mean molecular kinetic energy. In a moment I'll say why I think so. For now, though, let me make a clarification and fend off some likely misunderstandings about the view.

First the clarification. I've been making claims about the explanatoriness of various bits of science and mathematics, but I haven't yet said exactly what I take such claims to mean. Just what notion of explanation is at stake here? To some extent I'd like to be noncommittal about this issue. Experience suggests that we—that is, mathematicians, philosophers of mathematics and other knowledgeable observers—are usually good at agreeing about what counts as explanatory in specific cases, once all the relevant facts are on the table. So I take considered and informed judgments of this sort, including my own, to be generally reliable. On the other hand, there's very little agreement about which theory of explanation is correct (including, and perhaps especially, in the mathematical setting), and so it seems best not to rely overmuch on any such theory.

Nevertheless, I think one can safely say at least a few things about the properties of (mathematical) explanations. To start with, there are at least two kinds of phenomenon that arguably warrant the name. One familiar type is an *answer to a why-question*.[14] Another type might be termed *rendering intelligible* (or *elucidating* or *clarifying*). Whereas answers to why-questions explain facts, elucidations explain things: they illuminate the nature of an object or concept, where our previous understanding was obscure, incomplete, inconsistent or otherwise problematic.[15] The discussion below will involve claims about both types of explanation. (For instance, section §2.4 considers Maddy's claim that the reductionist viewpoint explains why multiplication is commutative, while §2.6 deals with Quine's remarks about set-theoretic "explications" of arithmetical notions.)

What's more, explanations stand in a characteristic pattern of relations to other epistemic and cognitive goods. Possessing an explanation typically imparts understanding, for example, and often improves learning, inference and problem-solving in the relevant domain.[16] Hence it's reasonable to take the absence of epistemic and cognitive benefits to reliably indicate a lack of explanatoriness. Finally,

to be worth caring about. For instance, if one subscribes to the view that facts of the form $\exists x Fx$ are explained by their instances, then one might think that the reduction of arithmetic to set theory explains why there exists a theory to which arithmetic is reducible. Maybe so. But my interest is in explanations that make a substantive contribution to mathematical understanding, and such cases clearly don't fit the bill.

[14]As philosophers have taken pains to show over the past couple decades, this type of explanation is encountered in pure mathematics no less than in the empirical sciences; working mathematicians often ask why theorems are true, and they often go to lengths in search of answers. (See the references in footnote 2 above.) To take a random recent example, [Tao 2015] proposes a large-scale collaborative effort to find an explanation for certain surprising polynomial identities.

[15]Some philosophers may want to deny that this second phenomenon is a genuine kind of explanation. If they were right, then my main thesis would be that much easier to defend, since there would then be fewer ways for the reduction of arithmetic to set theory to be prospectively explanatory.

[16]See, e.g., [Lombrozo 2006] and [Lombrozo 2012] for references and discussion.

we know what sorts of things can serve as mathematical explananda and explanantia. Theorems often stand in need of explanation, as do (the natures of) kinds of mathematical entities. The former can be explained by appropriately enlightening proofs, or by other theorems or bodies of theory. The latter can be explained by suitably clear definitions or identifications.

I hope the above gives some reassurance that the question I've posed makes sense, and that we have some resources for trying to answer it. If this isn't yet entirely clear, perhaps the details that emerge in the rest of the paper will help make it so.

I now want to mention and try to dissolve a few possible confusions about what my view entails. First, in denying that the reduction of arithmetic to set theory is explanatory, I don't mean to presuppose that it *ought* to have been explanatory. So I don't claim that the lack of explanatory power means that the reduction lacks value, or that it failed to achieve its intended purpose. Although some people have demanded that set-theoretic foundations be explanatory—as I'll discuss below—my view is that such reductions can be (and are) successful without contributing to explanation or understanding. (See §2.7 for more on these issues.)

Second, it's no part of my view that set theory contributes nothing of explanatory value to *mathematics generally*. This claim is pretty clearly wrong. To take just one example, there were important explanatory issues at stake in the questions about infinite sets of real numbers that originally motivated Cantorian set theory. Consider in this regard Joseph Dauben's comments on Cantor's early work:

> [In 1874 Cantor] had established that $\mathbb{R}$ was nondenumerable. Following this remarkable discovery, certain new avenues of inquiry must have appeared simultaneously. If, in terms of cardinality, there were more real numbers than natural numbers, were there further quantitative distinctions to be made between and beyond them? How did the new discovery contribute to the explanation and understanding of continuity? ...As Cantor raised the question of mapping lines and planes, such questions seem to have been uppermost in his mind.[17]

Not only did Cantor puzzle over the correct explanation of continuity, infinite size, and other notions, his work also yielded illuminating answers to many such questions. The thesis I defend about the reduction of arithmetic isn't intended as a denial of set theory's explanatory achievements in these and other parts of mathematics.

Finally, and perhaps most importantly, my view should also be distinguished from the claim that set theory fails to explain anything about the natural numbers in particular. This last claim sounds

[17] [Dauben 1979], 58-59.

much like the one I want to defend. But the two aren't equivalent, and failing to properly distinguish between them is likely to cause confusion and prompt misplaced objections. In fact the view just mentioned is substantially stronger than my own view. Moreover, I think this stronger view is probably false. Arguably the notion of set is needed to properly characterize the natural numbers, since the (second-order) axiom of induction quantifies over sets of numbers.[18] And the fact that the collection of natural numbers has the smallest possible infinite cardinality is an interesting result that's plausibly explained by set theory. But these explanations in no way involve *viewing numbers as sets*. (Rather, they involve taking the collection of natural numbers and its subcollections as sets, which is something else altogether. Numbers need not *be* sets to occur as *elements* of sets.) So it's important to note that cases involving this sort of explanatory relationship pose no threat to my view.[19]

The thesis I'm defending, then, is more modest than the above claims. The view, to reiterate, is just that *the reduction of arithmetic to set theory*—not set theory itself, or the relationship between sets and numbers on the whole—is unexplanatory. There are both positive and negative reasons to think so. The positive reasons are considerations that directly cast doubt on the explanatoriness of the reduction. The negative reasons consist in the lack of evidence for the contrary claim (that the reduction is explanatory after all). As is often the case when one tries to establish a nonexistence claim, the issues surrounding the negative reasons are arguably more interesting and weighty than those surrounding the positive reasons—after all, once we've shown that the alleged evidence for the existence claim is unconvincing and that the claim isn't needed to account for the phenomena, considerations of parsimony give us good reason to reject it. So most of the work below will focus on answering various arguments purporting to show that the reduction of arithmetic to set theory is explanatory.

[18]Thanks to both Kenny Easwaran and an anonymous referee for *Synthese* for suggesting this example.

[19]An anonymous referee suggests that results like Goodstein's theorem—which are unprovable in first-order Peano arithmetic, but provable in the second-order setting using set-theoretic methods—might count as set-theoretic explanations of arithmetical facts, and hence might seem problematic for my view. I'm inclined to group this sort of case, together with the others mentioned in this paragraph, under the heading of "ways of applying set theory to arithmetic that may be explanatory, but which don't depend on viewing numbers as sets". So I don't find such cases worrying.

To see why, it's helpful to briefly describe the set-theoretic proof of Goodstein's theorem. One starts by considering an arbitrary Goodstein sequence $G(m)$, which is a certain sequence of natural numbers. One wants to show that $G(m)$ eventually terminates, i.e. that it takes the value 0 at some point. (Goodstein's theorem is the statement that all Goodstein sequences terminate.) To show this, one constructs a sequence of ordinal numbers $O(m)$ with the properties that (1) $G(m)$ can be shown to terminate if $O(m)$ terminates, and (2) $O(m)$ does in fact terminate. This is a sparse sketch of the proof, of course, but hopefully it's clear from the sketch that the proof doesn't exploit the set-theoretic representation of the natural numbers in any way. One could view the numbers as any sort of object whatsoever and the proof would still go through.

In any case, I think it's far from clear that the set-theoretic proof of Goodstein's theorem is explanatory in the first place. The fact that a certain sequence of ordinals terminates, it seems to me, surely isn't *the reason why* the associated Goodstein sequence terminates. The behavior of the two sequences is correlated, but what happens with the sequence $O(m)$ can hardly be said to "ground" or "determine" what happens with the sequence $G(m)$. So I'm doubtful that we're even dealing with a case of mathematical explanation here.

2.2 Some positive considerations

First, however, the positive considerations. One serious reason for doubt, alluded to above, is the fact that certain other set-theoretic reductions—e.g. that of the ordered pair—are pretty clearly unexplanatory. If it turns out that the reduction of arithmetic is similar to these cases in important respects, then we have good inductive grounds to expect that the reduction of arithmetic isn't explanatory either.

Are the two kinds of case similar in ways that matter? I think so. In particular, the account sketched above about the translation of an antecedently well-understood concept C into set-theoretic language seems to fit the case of arithmetic nearly as well as it fits the case of the ordered pair. In both cases, a specification of the logical properties of the structure preceded the attempt to model the the structure in set theory. Also in both cases, several such surrogates were proposed[20]—each of which satisfied the relevant postulates, but which differed along various other dimensions—before the mathematical community eventually settled on a favorite. And the consensus in each instance was based on practical considerations of simplicity, well-behavedness and so on, rather than epistemic claims about explanatoriness or understanding.

In the case of the ordered pair, this process yielded a construction that pretty obviously has no claim to explanatory value. Since the (von Neumann) reduction of arithmetic to set theory was the result of much the same type of process, it's reasonable to expect much the same outcome.

Another reason for doubt is based on number theorists' own attitudes and behavior. It's evident that mathematicians care a great deal about obtaining understanding, and that this value substantially influences mathematical practice and pedagogy.[21] It's also a commonplace that explanation and understanding are closely related[22], insofar as good explanations typically produce better understanding of their *explananda* (and the associated subject matter). Thus, if viewing numbers as sets had appreciable explanatory value, we should expect number theorists to adopt this vantage point for the sake of exploiting its epistemic and cognitive resources.

But this isn't in fact what happens. Although number theorists draw extensively on other areas of mathematics to gain insight and solve problems—think of Wiles's use of algebraic geometry in the proof of Fermat's Last Theorem, Hrushovski's model-theoretic approach to the Mordell-Lang conjecture, the application of Fourier-type methods in studying prime distributions, and so on—one doesn't find the

[20]The 1921 Kuratowski definition of the ordered pair was preceded by attempts by Wiener (1914) and Hausdorff (1914), who identified (a, b) with $\{\{\{a\}, \emptyset\}, \{\{b\}\}\}$ and $\{\{a, 1\}, \{b, 2\}\}$, respectively. (The 1 and 2 in Hausdorff's definition are arbitrarily chosen objects distinct from a, b, and each other.)

[21]See for instance [Avigad 2008], [Carter 2008], [Lipton 2011], [Sierpinska 1994], [Tappenden 2005].

[22]It's sometimes argued (e.g. by [Khalifa 2012]) that understanding is nothing other than having an explanation. At other times (e.g. by [Strevens 2013]) the two states are treated as distinct, but explanation is viewed a necessary component or precondition of understanding.

identification of numbers with sets being deployed for such purposes.[23] Viewing the natural numbers as, say, the von Neumann ordinals has led to no breakthroughs on longstanding problems, insightful new ways of organizing and teaching number theory, or the like. The fact that number-theoretic practice has so little use for the set-theoretic viewpoint suggests that the latter doesn't contribute significantly to understanding. Hence the reduction of arithmetic to set theory is probably not explanatory.

I turn next to negative considerations. In the parts to follow, I'll consider four arguments purporting to show that the reduction of arithmetic to set theory is explanatory after all. The first, due to Zermelo and his followers, centers on the ontological thesis that numbers just *are* sets. The second, due to Maddy, claims that there are explanatory set-theoretic proofs of some arithmetical facts. The third, due to Kitcher, suggests that the reduction of arithmetic to set theory sheds light on some high-level features of the natural number system. The last, due to Quine, maintains that the reductionist viewpoint renders the concept of number more intelligible. These arguments raise many interesting issues, but none, I think, are ultimately successful. Hence we've been given no good reason to think that the reduction of arithmetic to set theory is explanatory.

2.3 A reductionist argument

Reductionism about objects of a given type, say A, is the view that there exist objects of some notionally different type B such that the As are in fact identical to some of the Bs. *Reductionism* about As is thus a stronger thesis than the claim that the A-theory is *reducible* to the B-theory, since reducibility (on the Nagel-style view that I endorse) is a formal rather than metaphysical affair.[24]

It's often tempting to make the leap from reducibility to reductionism—if the A-theoretic truths correspond systematically to B-theoretic truths (with nothing left over), why not try to keep ontological costs down by doing away with the As entirely? This maneuver has made plenty of appearances in philosophy of science, philosophy of mind and elsewhere. It's not too surprising, then, that some mathematicians and philosophers have embraced *set-theoretic reductionism* of one sort or other.

The version of reductionism that's relevant for present purposes is the claim that the natural numbers in particular are nothing but sets. (This thesis is arguably traceable to, and most famously associated with, the early work of Ernst Zermelo. See [Zermelo 1909a] and [Zermelo 1909b] for statements of Zermelo's reductionism, and [Taylor 1993] and [Hallett 1984] for analysis and commentary.) If reductionism about the natural numbers is true, then the line I'm defending about set theory and arithmetic is in trouble. After all, if numbers are identical to sets, it's hard to see how the reduction of arithmetic to set

[23] As I pointed out before, care should be taken not to confuse this point with the claim that set theory *in general* has nothing of explanatory value to contribute to number theory. I take the latter claim to be false, or at least highly dubious.

[24] See [Klein 2009] for an extended (and convincing) defense of this claim.

theory could fail to be explanatory. Ascertaining what numbers really are surely counts as explaining something important about them, to start with. And presumably this identification will have all sorts of explanatorily valuable further consequences. Compare the identification of water with H_2O, of genes with regions of DNA, and so on. Such identifications are always explanatory achievements in their own right, it seems, and they invariably give rise to explanations involving other properties of the reduced entity—why salt but not sand is soluble in water, or how allelic dominance and recession are implemented at the molecular level, for instance. So it should go with numbers and sets, if reductionism is true.

The most influential argument against reductionism is due, of course, to Paul Benacerraf ([Benacerraf 1965]). Benacerraf's reasoning is roughly as follows. There are several alternative reductions of arithmetic to set theory, each giving a different correspondence between numbers and sets. If numbers are identical to sets, as the reductionist claims, then any particular number must be determinately identical to some particular set. So at most one of the alternative reductions gives the right identification of numbers with sets. Suppose that R is this reduction. Then R must have some special feature in virtue of which it stands out as uniquely correct. But, in fact, none of the alternative reductions has any such special feature. (There seems to be nothing outstanding in the relevant way about the von Neumann ordinals $0 = \emptyset,\ 1 = \{\emptyset\}\,,\ 2 = \{\emptyset, \{\emptyset\}\}\,, \ldots$ as compared to the Zermelo ordinals $0 = \emptyset,\ 1 = \{\emptyset\}\,,\ 2 = \{\{\emptyset\}\}\,, \ldots$, for instance.) Therefore reductionism is false.

The broad consensus in philosophy of mathematics has been that Benacerraf was right, and hence that reductionism is untenable. Even philosophers who are sympathetic to some version of mathematical realism tend to think that numbers are obviously not sets: the platonist Mark Balaguer, for instance, writes that "it is more or less beyond doubt that no sequence of sets stands out as the sequence of natural numbers."[25] Still, there are occasional voices of dissent. The most extensive recent defense of reductionism is [Steinhart 2002], which argues that the natural numbers are identical to the finite von Neumann ordinals.[26]

[25] [Balaguer 1998], 64.

[26] Of course, Benacerraf's argument has generated a great deal of discussion over the years, and some post-Benacerrafian philosophers have continued to hold views that ascribe some special metaphysical status to sets vis-à-vis the natural numbers. One might think that a proper examination of the reductionism issue would include some mention of these views. In fact, though, the views in question—or at least the ones I'm aware of—uniformly concede Benacerraf's point that numbers can't be uniquely identified with any particular sets in a principled way. Hence they're not directly relevant to the line of thought I take up here. Some noteworthy examples of what I have in mind are the views of Penelope Maddy, W.V. Quine and Nicholas White.

Maddy's early work (e.g. [Maddy 1990]) argues that numbers are (equinumerosity) *properties* of sets. This is indeed reductionism of a certain sort. But Maddy explicitly says that, for Benacerrafian reasons, there's no hope of identifying numbers with sets themselves. (Whether numbers can or should be identified with properties of sets, and whether such an identification would have explanatory value, is an interesting question. But it's not quite the question this paper is trying to answer.)

Quine (e.g. [Quine 1960]) held a view that can be described as claiming that "numbers are sets", which sounds anti-Benacerrafian. But his version of reductionism basically amounts to the thesis that it's convenient to identify numbers with sets, together with a pragmatic approach to ontology. On Quine's view, we're free to make any identification of

Steinhart's argument has two parts. The first part consists of an enumeration of some nice properties of the finite von Neumann ordinals (FVNOs) as compared to the Zermelo ordinals and other candidates. (The list includes the naturalness of the extension of the FVNOs to the transfinite ordinals, the elegance of identifying the numerical ordering $<$ with set membership, and so on. Other candidate reductions also have some of these features, but the FVNOs arguably boast the most impressive assortment.) As Steinhart points out, mathematicians standardly represent the natural numbers with the FVNOs on account of their having these convenient features. This fact, Steinhart thinks, amounts to "an argument from mathematical practice" that the natural numbers are the FVNOs.

There are at least two ways to understand the argument in question. One possible claim is that mathematicians (or set theorists, at any rate) are themselves committed to the view that the FVNOs are the natural numbers—and so, since we should defer to epistemic authorities about questions within their domain of expertise, we should follow them in being so committed. A second possible claim is that, irrespective of what mathematicians actually happen to believe, their reasons for preferring the FVNOs are in fact compelling grounds to identify the FVNOs with the natural numbers.

Neither version of the argument is very convincing. Starting with the first version, it's doubtful that mathematicians (or set theorists in particular) are generally committed to identifying the natural numbers with the FVNOs. For one, some prominent set theorists are quite clear about rejecting this sort of reductionism. For instance, as we saw above, Moschovakis takes pains to warn his readers against supposing that functions and other familiar objects are identical to the sets used to represent them. Even setting this point aside, however, there's no good reason to think that mathematicians' preference for the FVNOs should be understood as an ontological prescription rather than a practical one. Even if some set theorists have suggested that "the natural numbers are (or ought to be viewed as) the FVNOs", the context in which such claims are made is that of ascertaining which practices are most amenable to

---

numbers with sets that serves our purposes; if there are several equally handy choices, then each identification counts as equally correct, as long as we stick with the one we've chosen. So there's no notion here of getting at a deep truth about what numbers "really were all along". In particular, Quine agrees with Benacerraf that there's nothing metaphysically special about either the Zermelo ordinals or the von Neumann ordinals. (Nevertheless, one might think that Quine's view still imputes a type of explanatory value to set-theoretic reductions, even if it isn't for the metaphysical reasons discussed in this section. See §2.6 below for more on this issue.)

A less well-known but notable reaction to Benacerraf is [White 1974]. White agrees with the anti-uniqueness part of Benacerraf's argument, but from here he takes the unusual line that "the existence of multiple set-theoretic models of arithmetic should prompt us, not to say with Benacerraf that numbers cannot be sets, but rather to suggest that there are multiple full-blown series of natural numbers. Thus, for example, instead of there being only one three, there are after all many threes, and many thirty-sevens, and so on" (112). White's view sounds at first like a version of reductionism, but it later becomes clear that this isn't what he has in mind. His view is rather that objects of *any* kind count as numbers, insofar as they can be placed in an $\mathbb{N}$-like progression. So White doesn't, after all, identify numbers with set-theoretic finite ordinals in particular (although these are *among* the objects that count as numbers for him). As he points out, the view is better described as a sort of Pythagoreanism than a type of set-theoretic reductionism.

Finally, it's worth mentioning that another prominent claim from "What Numbers Could Not Be" is Benacerraf's thesis that numbers aren't objects *of any sort at all*. This is a claim with which many people *have* directly disagreed. I find Benacerraf's argument for this view unconvincing myself, but its truth or falsity doesn't directly bear on anything I say below, so there's no need to canvass responses to it here.

doing set theory, not that of answering philosophical questions about identity.

As for the second version, suppose we grant for the sake of argument that the FVNOs edge out other candidates in the possession of nice properties. Does any interesting metaphysical conclusion follow? It's hard to see how it could. It's not clear what sort of plausible principle would let us infer "the Xs are the Ys" from "representing the Xs as the Ys within set theory is moderately more convenient than doing otherwise". (Analogously: there are many possible ways of representing the natural numbers by arrays of dots. Some such representations are more simple, elegant and practical than some others; perhaps there's some reasonable way of assigning each one an overall niceness score, and perhaps if we did so there'd be a convincing overall winner. In itself, though, none of this seems seems to convey any information about whether numbers really are arrays of dots, or which arrays of dots they'd be if they were.) In particular, if this line of reasoning is to be understood as an inference to the best explanation, it's unclear how to make it work: the proffered *explanans* hardly seems to count as any sort of explanation of the *explanandum*, let alone as a uniquely good such explanation.

In any case, Steinhart himself concedes that the "argument from mathematical practice" is inconclusive. But he also offers a second, and supposedly decisive, argument for reductionism. (Indeed, he refers to the second argument as "a precise mathematical demonstration that the natural numbers are the finite von Neumann ordinals."[27]) The reasoning is as follows. To start with, Steinhart borrows from Benacerraf two sets of conditions that a structure S must satisfy in order to serve as the set of natural numbers. The first set of conditions is the Peano axioms for arithmetic. The second set consists of two "cardinality conditions":

1. For all $n \in S$—i.e., for all putative natural numbers n—there exists a set $n^* = \{m \in S : m < n\}$ of all numbers less than n.

2. The cardinality of any set A is n if and only if there's a bijection between A and n^*.

Steinhart argues that, *pace* Benacerraf and his followers, accepting these two sets of conditions ("the NN-conditions") commits us to identifying the natural numbers with the FVNOs.

Steinhart's argumentative maneuvers aren't always easy to follow, but the general shape of the reasoning seems to be this. First, suppose that reductionism is true, so that the natural numbers and arithmetical operations are uniquely identifiable with a particular set-theoretic structure $(S, f, e, \prec)$, or S for short. (As above, S is a set, f is a one-place function to be interpreted as *successor*, e is an element of S to be interpreted as 0, and $\prec$ is a two-place relation to be interpreted as *less than.*) Then, whatever structure S might be, the cardinality conditions above "[compel] us to form certain definite

[27][Steinhart 2002], 355.

sets" of the elements of S, namely the sets n^* for each $n \in S$.[28] With appropriately chosen definitions for zero, successor, and less than, however, the sets n^* themselves constitute a structure S^* that satisfies the Peano axioms and the cardinality conditions. This structure S^* is in fact nothing other than the FVNOs. By uniqueness, therefore, $S = S^*$, and so the natural numbers are uniquely identifiable with the FVNOs.

Perhaps the most important problem with Steinhart's argument is that it's not clear why the structure S^* is the only candidate we're permitted to consider in determining the identity of S (and hence that of the natural numbers). That is, it's unclear why the following reasoning is apparently not allowed: "Suppose that the natural numbers and arithmetical operations are uniquely identifiable with a particular structure S. As is well known, the finite Zermelo ordinals with the usual operations constitute a structure Z that satisfies the Peano axioms and the cardinality conditions. By uniqueness, therefore, $S = Z$, and so the natural numbers are uniquely identifiable with the finite Zermelo ordinals."

Steinhart's answer to this challenge seems to involve the idea that "the cardinality [condition] compels us to form certain definite *sets*... Specifically: we have to form, for each [$n \in S$], the set of all numbers less than n. For every number n, we must form the set [$n^* = \{m \in S : m \prec n\}$]. ...We cannot avoid forming all these sets of numbers."[29] Later, he adds that the conditions mentioned above "*do not* allow you to form any other sets of [numbers from S]. They do not assert or imply that the NN-universe contains any other sets built from [such] numbers."[30] ("The NN-universe" is defined to be "the universe of objects to which we may ontologically reduce the natural numbers".[31])

This is rather puzzling. Steinhart's thought seems to be that we can only consider S and S^* (and not any other structures to which we might reduce the natural numbers) because the existence of S^*, and only that of S^*, is entailed by the existence of S together with the cardinality conditions. But why should this matter? Given that we accept ZFC (or some other system of set theory), we have a universe's worth of sets already at our disposal. It makes little sense to forgo the use of these other sets, or to pretend they don't exist, for the reasons Steinhart gives. Whether or not we can "avoid forming" other sets on the basis of the existence of S and the cardinality conditions seems hardly relevant, and perhaps isn't even a coherent question—the other sets are in some sense already there, and hence not in need of being "formed" by us at all. Thus, as long as we're committed on good independent grounds to the existence of the Zermelo ordinals (say), they ought to be fair game in our deliberations about the identity of the natural numbers. And I take it we are so committed. Hence I fail to see how Steinhart

[28][Steinhart 2002], 351.
[29][Steinhart 2002], 351.
[30][Steinhart 2002], 353.
[31][Steinhart 2002], 352.

can block the $S = Z$ argument sketched above in a motivated way. (Of course, if he can't do so, then nothing prevents Benacerraf's reductio from arising again: in that case we'll have $S = Z$ and also $S =$the FVNOs, and hence we'll have to reject the assumption that the numbers are uniquely identifiable with a particular series of sets.)

In light of these issues, I conclude that Steinhart's defense of reductionism is unsuccessful. There remains no compelling reason to identify the natural numbers with any particular set-theoretic structure. Hence we needn't think that the reduction of arithmetic to set theory is explanatory in virtue of its revealing the true nature of the numbers.

2.4 A Maddian argument

Here's another worry. It might be that general, a priori reflection about set theory and arithmetic is liable to paint an incomplete picture of the relationship between the two. Set-theoretic language and reasoning have a distinctly different flavor from their number-theoretic counterparts, after all, and it's not easy to imagine how the theorems and proofs of arithmetic will look when viewed through the lens of set theory. If we pay close attention to details, the reduction of numbers to sets might yet yield an unexpectedly insightful proof or piece of reasoning.

What might a case of this sort look like? Here's a suggestion from an early paper of Penelope Maddy:

> [L]et me give an elementary example of what I take to be a demonstration of the explanatory power of set theory. Suppose you wonder why multiplication is commutative. You could prove this by induction from the Peano postulates by showing:
>
> (1) $n \cdot 0 = 0 \cdot n$
>
> (2) if $n \cdot m = m \cdot n$, then $(n + 1)\, m = m\, (n + 1)$
>
> ...If you were convinced of the truth of the Peano postulates and of the soundness theorem, this exercise should convince you of the commutativity of multiplication, but you were convinced of this in the first place; your question was why, and the proof did little towards answering it. Now suppose you take a set theoretic perspective and again ask why multiplication is commutative. Here an answer is forthcoming: because if A and B are sets, then there is a one-to-one correspondence between the cartesian products $A \times B$ and $B \times A$. The central idea in the proof of this fact is the old observation that a rectangle of n rows of m dots contains $n \cdot m$ dots, but turned on its side it contains $m \cdot n$ dots. I take it that this explains why multiplication is commutative, and their ability to provide such an explanation

provides some theoretical support for the set theoretic axioms involved.[32]

This is an interesting idea, but I don't think it works. The main problem is that the envisaged proof doesn't look like proper set-theoretic reasoning at all, since rectangles made of rows of dots aren't objects of pure set theory. What Maddy is describing is instead a (partly) *diagrammatic* or *geometric* proof—i.e., something like the following.

1. If a and b are natural numbers, then the product $a \cdot b$ is equal to $|A \times B|$ (i.e., the number of elements in the Cartesian product of A and B, where these sets are the ordinals representing a and b respectively). [Fact]

2. $|A \times B|$ is equal to the number of dots in a rectangle consisting of $|A|$ rows of $|B|$ dots. ["Bridge premise", established by a separate argument]

3. The number of dots in a rectangle consisting of $|A|$ rows of $|B|$ dots is equal to the number of dots in a rectangle consisting of $|B|$ rows of $|A|$ dots. [Geometric premise, evident from an appropriately constructed diagram]

4. So $|A \times B| = |B \times A|$. [From 2 and 3]

5. Therefore $a \cdot b = b \cdot a$. Hence multiplication is commutative. [From 1 and 4]

The geometric premise 3 is clearly essential to the argument. I don't think it has any place in a supposedly pure set-theoretic proof, and so I don't think Maddy can claim the proof she describes as an explanatory victory for set theory. But suppose one is convinced for some reason that premise 3 is fair game for the set theorist. Then, presumably, it should also be fair game for the number theorist. In that case, though, nothing prevents the number theorist from constructing a similar proof of her own:

1. The product $a \cdot b$ is equal to the number of dots in a rectangle consisting of a rows of b dots. [Bridge premise]

2. The number of dots in a rectangle consisting of a rows of b dots is equal to the number of dots in a rectangle consisting of b rows of a dots. [Geometric premise]

3. Therefore $a \cdot b = b \cdot a$. Hence multiplication is commutative. [From 1 and 2]

The second proof is simpler, more transparent, and arguably more explanatory than the first, since it avoids the extra detour through set theory. So even if we allow proofs of the commutativity of

[32] [Maddy 1981], 498-499.

multiplication that appeal to dot diagrams, the set-theoretic viewpoint still has no explanatory advantage over the "naïve" number-theoretic approach.

In response to this criticism, one might claim that the so-called geometric premise should in fact be viewed as a set-theoretic proposition, so that Maddy's proof is really a purely set-theoretic argument. The idea could be cashed out in a number of different ways. One could hold, for instance, that the notions of dots and rectangles are somehow *dependent* upon, or *grounded* in, or *best understood* in terms of the primitive notions of set theory, e.g. the concept of Cartesian product.

I'm not convinced by this line of thought. The correct response will depend to some extent on the way the details are spelled out, of course. Nevertheless, I see no reason to accept that set theory is epistemically or metaphysically prior to geometry in any interesting sense. As with any mathematical theory, it's certainly possible to model arrangements of dots or whatever with sets. But this doesn't show that such objects are identical to sets, or metaphysically grounded in sets, or rendered more intelligible by sets. And there's plenty of reason to doubt such claims. (Indeed, the idea that geometry depends in some such way on set theory may be more dubious than the corresponding claim about numbers and sets.) Compare these remarks of Moschovakis on the relation between lines, points and sets of numbers:

> What is the precise meaning of this "identification" [of a line with a set of real numbers]? *Certainly not that points are real numbers.* Men have always had direct geometric intuitions about points which have nothing to do with their coordinates... Every Athenian of the classical period understood the meaning of the sentence 'Phaliron is between Piraeus and Sounion along the Saronic coast' even though he was (by necessity) ignorant of analytic geometry. In fact, many educated ancient Athenians had an excellent understanding of the Pythagorean theorem, without knowing how to coordinatize the plane.[33]

If this is true for lines and points, as I think it is, then something analogous is surely also true for arrangements of dots. We can "identify" the content of such a picture with the Cartesian product of two sets if we want. But to suggest that this content depends in some way on products of sets, or that it can only be properly understood in light of set theory, is a different (and much less plausible) story. On the contrary, it's exactly the visuospatial character of the diagram that makes premise 3 of the Maddian argument obvious, and in turn makes the whole proof explanatory. And this visuospatial character is lost when we boil the diagram down to a proposition about sets.

In any case, Maddy's framing of the situation is misleading in a more basic way. To explain why, let's say that a proof of an arithmetical theorem is *reductive* if the proof involves viewing numbers as

[33] [Moschovakis 2006], 33. The italics are Moschovakis's, but the sentence in single quotes is displayed on a separate line in the original.

sets and *non-reductive* otherwise. Maddy's argument compares a reductive proof of the commutativity of multiplication with one particular non-reductive proof, which starts from the Peano axioms and proceeds by induction. Maddy may well be right that this proof is unexplanatory. (Cf. [Lange 2009], which argues that proofs by induction are unexplanatory as a rule.) But the more interesting question here is whether some reductive proof is more explanatory than *any* non-reductive proof. It could be, after all, that Maddy's non-reductive proof is particularly unexplanatory, but that better alternatives exist. Even if Maddy's claims are all true, their force would be less clear in this sort of situation.

So are there better non-reductive proofs of commutativity? I think so. One such proof, for instance, is a simple and compelling argument based on the Principle of Recursive Definition. The idea here is to show that the functions $m \times n$ and $n \times m$ both satisfy the definition of multiplication

$$\forall m, n \in \mathbb{N} : \begin{cases} m \times 0 & = 0 \\ m \times (n+1) & = m \times n + m. \end{cases}$$

Since the Principle of Recursive Definition says that there exists exactly one function satisfying such a recursion formula, it follows that $m \times n = n \times m$, so multiplication is commutative. (See Theorem 16.10 and its proof in [Warner 1990].) I don't want to claim too much on behalf of this proof, but it at least doesn't strike me as totally lacking in explanatory value. Note that it allows a straightforward and fairly satisfying answer to the commutativity question: multiplication is commutative because, if it weren't, there would be two distinct functions satisfying the recursive definition of multiplication, which is impossible.

I conclude that Maddy is wrong to claim that set theory explains the commutativity of multiplication. The proof she offers, though plausibly explanatory, is essentially geometric or diagrammatic, and this geometric element can't be harmlessly replaced with more set theory. And the purely set-theoretic, "reductive" version of Maddy's proof is no more explanatory than the non-reductive proof just mentioned.

Obviously Maddy's example represents just one way that set theory could be claimed to explain a fact about numbers. I don't claim to have ruled out all the other possible candidates. On the other hand, few philosophers have thought about set theory as much and as well as Maddy, so the failure of her preferred example is a discouraging sign for this line of thought in general.

2.5 A Kitcherian argument

The objection in the last section involved the possibility that, once we've reduced numbers to sets, some number-theoretic facts might turn out to be explained by set-theoretic considerations. The focus there was on individual theorems, like the commutativity of multiplication, and their proofs. Here I'll consider a different type of objection. Instead of (or in addition to) explaining particular results, the thought goes, it might be that set-theoretic reduction makes its explanatory contribution by shedding light on higher-level, relational features of the natural numbers and their arithmetic. Such features might include, for instance, revealing analogies or unexpected logical connections involving arithmetic and other branches of mathematics. A defense of this line of thought comes from Philip Kitcher:

> Mathematical proofs can serve other functions besides that of increasing the certainty of our knowledge. In some cases, as in the case at hand, they can be part of a scheme of ontological reduction in which our justifications for accepting the theorems proved remain unaffected. Moreover, without making us any more certain that a theorem is true, a proof can show us why it is true: proofs may yield explanatory dividends. ...Reducing arithmetic to set theory has explanatory, as well as ontological, value. For, in the light of the reduction, our understanding is advanced through exhibition of the kinship between theorems of arithmetic and theorems in other developments of set theory (in particular, branches of abstract algebra).[34]

Kitcher doesn't explain what sort of kinship he has in mind, but I think the general idea is clear enough. When Kitcher speaks of abstract algebra and its influence on our understanding of arithmetic, he's presumably referring to the developments surrounding the birth of "modern algebra" around the beginning of the 20th century. This period—which saw the first general definitions and systematic study of objects like groups, rings and fields, owing to the work of Noether, Hilbert, Artin, Steinitz, van der Waerden and others—had major consequences for almost every area of mathematics, including our understanding of the natural numbers and related number systems.

One type of progress was classificatory. By categorizing number systems according to their abstract algebraic properties, modern algebraists got a better fix on the essential features of particular structures, and a better understanding of genus-species relations between structures. For instance, the natural numbers with addition $(\mathbb{N}, +)$ were seen to be an example of a *commutative monoid*, the integers with addition $(\mathbb{Z}, +)$ an *abelian group*, the integers with addition and multiplication $(\mathbb{Z}, +, \cdot)$ a *commutative ring*, the rationals $(\mathbb{Q}, +, \cdot)$ a *field*, and so on. Along with these taxonomic advances, algebraists figured out how to pass from the natural numbers to other number structures by way of rigorous algebraic

[34] [Kitcher 1978], 123.

constructions. The integers, for example, can be obtained as the Grothendieck group of the natural numbers (as the construction is now called). Similarly, the rationals arise as the fraction field of the integers, and the complex numbers as the algebraic closure of the reals.

Finally, algebraists in the modern era discovered general theorems about algebraic structures that shed new light on the properties of the familiar number systems. For instance, the Fundamental Theorem of Arithmetic—the fact that every natural number greater than 1 has a unique decomposition as a product of primes—had been known at least since Euclid. But only with the help of abstract algebra did the FTA clearly emerge as an instance of a more general phenomenon, involving notions like *commutative ring* and *irreducible element*[35]. The picture is as follows. A commutative ring in which every element factors uniquely as a product of irreducibles (and which lacks zero divisors) is called a "unique factorization domain". The integers $\mathbb{Z}$, of course, are a unique factorization domain, and it turns out that the irreducible elements of $\mathbb{Z}$ are exactly the primes. Taken together, these facts imply the FTA. So abstract algebra gives us a nice story about why the FTA holds for the natural numbers, where before this looked like a more or less brute fact. (It also reveals how FTA-like results hold, or fail to hold, in other structures.)

In light of these kinds of advances, it's reasonable to think that developments in algebra have explained some things about the natural numbers and their place in the mathematical universe. In particular, I take it that the example of the previous paragraph is the sort of thing Kitcher has in mind when he talks about "the kinship between theorems of arithmetic and theorems in... abstract algebra". And I agree that understanding why and to what extent results like the FTA hold in various number systems can improve our understanding of arithmetic, as this line of thought has it.

So where does the reduction of arithmetic to set theory enter this picture? According to Kitcher, both arithmetic and algebra count as "developments of set theory" in some sense, and this is supposed to explain why the reductionist viewpoint deserves the credit for the illuminating parallels between the two subjects. How exactly this argument is supposed to work is a little unclear. But perhaps what Kitcher has in mind is something like this. As a matter of historical fact, it's true that abstract algebra reached maturity around the same time as set theory, and set-theoretic language and tools exerted a notable influence on algebra during this period (as did as the axiomatic method pioneered by Hilbert, along with other trappings of the emerging "modern" style of mathematics). For instance, van der Waerden's *Moderne Algebra*—the first definitive modern algebra textbook, published in 1930—defines a group as a set equipped with a certain type of binary operation. And a great deal of progress during this

[35]An element of an integral domain is irreducible if it's neither zero nor a unit, and if it isn't expressible as a product of two non-units. (An integral domain is a commutative ring with no zero divisors, meaning that the product of nonzero elements is always nonzero. A unit is an element with a multiplicative inverse.)

period was due to Dedekind, who was extremely interested in rigorous set-theoretic foundations, and who used "set-theoretic tools and techniques... to construct new mathematical objects (the natural and real numbers, ideals, modules, etc.) or whole classes of such objects (various algebraic number fields, rings, lattices, etc.)."[36]

With these facts in mind, one might be tempted to make the following sort of argument:

1. Abstract algebra is fundamentally set-theoretic in nature, and so anything explained by algebra is ultimately explained by set theory.

2. Algebra provides explanations of some arithmetical facts (e.g., the explanation of the FTA mentioned earlier).

3. Therefore, set theory provides explanations of some arithmetical facts.

I don't know whether this is exactly the argument Kitcher had in mind, but it's the most plausible way I can see to make the general idea more precise. Still, I think the argument is unsuccessful for at least two reasons.

For one, the first premise is doubtful, at least under any interpretation that's interesting for present purposes. To start with, it's certainly not the case that algebra developed *from* set theory, or that algebra couldn't possibly be done outside a set-theoretic framework. The roots of modern algebra lie with Lagrange, Ruffini, Galois and others in the 18th and early 19th centuries, well before Cantor was born. Their successes, and those of their pre-Cantorian followers, show that much of abstract algebra could—and surely would—have been worked out without the backing of a formalized set theory. The kinds of algebraic insights about number theory I've mentioned, in particular, mostly work just fine without a rigorous notion of set. There's nothing particularly set-theoretic about the "algebraic explanation" of the Fundamental Theorem of Arithmetic outlined above, for instance.

Actually, though, the truth or falsity of the first premise doesn't matter much in the end. That's because, even if the above argument is sound, it nevertheless misses its mark. The view we wanted to investigate—and the one Kitcher claimed to be defending—is that the reduction of arithmetic to set theory is explanatory. But if the above argument shows anything, it's that *algebra* is reducible to set theory, and that *algebra* explains some things about arithmetic. Nothing about this conclusion depends in any way on viewing *numbers* as sets.

Thus I disagree with Kitcher's assessment of the reduction of number theory to set theory. On the one hand, Kitcher is right that one theory can advance our understanding of another by showing

[36] [Reck 2016].

how the latter is situated in the broader mathematical universe—by "exhibiting its kinship" with other areas of mathematics, showing its theorems to be special cases of more general phenomena, and so on. What's more, Kitcher is also right to think that abstract algebra has advanced our understanding of arithmetic in these ways. But he's mistaken to conclude that the explanatory virtues of algebra accrue to set theory. Still less do they accrue to the *reduction* of arithmetic to set theory.

2.6 A Quinean argument

I conclude this section with a line of thought due to Quine, tracing to his well-known discussion of set-theoretic reductions in §§53-54 of *Word and Object* ([Quine 1960]). Quine's view is that such reductions constitute major epistemic and theoretical achievements, in that they rendered more intelligible various notions which are extremely useful but which were previously unclear. Indeed, he takes the reduction of the ordered pair to be exemplary of the best way to solve a certain kind of philosophical problem, and he holds the reduction of arithmetic in similarly high esteem. (This discussion is the source of Quine's famous dictum "explication is elimination".)

Although Quine rarely uses the language of explanation explicitly, his arguments can nevertheless be taken to pose a challenge to the thesis I've been defending. For, as I noted above, rendering a given type of entity more intelligible can be viewed as a sort of objectual explanation. And Quine certainly thinks that set theory plays this clarifying role.

I'll start by focusing on Quine's discussion of the ordered pair in §53, since it's in this section that he lays out his view most clearly. Briefly, Quine's argument is that the ordinary conception of the ordered pair is "defective", "perplexing", "inadequately formulated", and "problematic", whereas the set-theoretic conception is "clear and couched in terms to our liking".[37] The argument for this weighty claim, unfortunately, is brief and unconvincing. As evidence for the defectiveness of the pre-set-theoretic conception, Quine cites Peirce's definition of the ordered pair (or "Dyad"), which Quine considers confused and obscure. I won't go into the details of Peirce's account here. Even granting the unsatisfactoriness of the Peircean definition, however, it's not clear how Quine's conclusion is supposed to follow. For we're given no reason to think that Peirce's remarks are representative of the ordinary conception in any way, or that his account is the best sense we can hope to make of the ordered pair without the help of set theory.

On the contrary, understanding the ordered pair from the pre-set-theoretic viewpoint presents no particular problem. The characteristic property of pairs, $(x_1, y_1) = (x_2, y_2) \leftrightarrow x_1 = y_1 \wedge x_2 = y_2$, is perfectly clear, and it makes no mention of sets. (Quine himself seems to agree; he praises the ordered

[37] [Quine 1960], 258-260.

pair postulate as "preternaturally succinct and explicit"[38].) Moreover, the postulate was formulated already by Schröder in 1895 ([Schröder 1895]), some twenty years before the first set-theoretic definition due to Wiener ([Wiener 1914]).[39] It's not at all obvious what's supposed to be so problematic about conceiving of an ordered pair as a *sui generis* object satisfying this axiom. Mathematicians like Schröder apparently did so without suffering any unease or confusion. So Quine's complaints about the epistemic inadequacies of the pre-set-theoretic conception aren't very compelling.

What about the supposed epistemic advantages of identifying pairs with sets? Even if the ordinary conception is clear enough on its own, after all, it's possible that the set-theoretic conception is better yet in some interesting way. But this is also doubtful. Although the naïve notion of set may have had some claim to special clarity, simplicity, and the like, modern axiomatized set theories are far more complex, and the nature of the sets they describe remains mysterious in many ways. It's hard to see what gains are to be made by trading the *sui generis* ordered pair for a construction that carries so much epistemological baggage. Indeed, I'm inclined to agree with Dipert's opinion that

> the pre-Wiener, primitive notion of ordered pair in Peirce and Schröder is actually clearer than its alleged set-theoretic representations... since the notion of set is itself so cumbersome and obscure. And a New Foundations set is still more obscure.[40]

Michael Hallett notes likewise that

> 'set' (unlike 'aggregate' perhaps) is not an ancient, well-understood concept which can easily be taken as an axiomatic primitive in the knowledge that it can be supported by extra-axiomatic explanation. (Unlike the case of natural number, this is largely *why* set theory is axiomatized, because we do *not* understand the set concept well.)[41]

There seems to be little reason, then, to accept either component of Quine's view. The pre-set-theoretic conception of the ordered pair isn't unclear or mysterious, except perhaps to the extent that one finds it hard to wrap one's mind around ordered pairs as *sui generis* objects. But the set-theoretic conception is hardly an improvement on this score. If anything, the sets postulated by ZFC, NF and similar systems are more problematic and harder to understand than ordered pairs. So, contra Quine, there's no obvious epistemic advantage to viewing pairs as sets.

So much for ordered pairs—what about the reduction of arithmetic to set theory? Quine gives essentially the same arguments here as in the previous case, and he draws a similar conclusion. ("But for

[38][Quine 1960], 259.

[39]See [Dipert 1982], fn. 23 for more on the early history of the ordered pair in modern logic.

[40][Dipert 1982], 367-8. New Foundations is Quine's system of set theory, which admits non-well-founded sets and has various other unusual features.

[41][Hallett 1984], 300. Emphasis in original.

its greater antiquity and its concern with a more venerable notion," he says, "the philosophical question 'What is a number?' is on a par with the corresponding question about ordered pairs."[42]) Quine's charge is that the reductions of Zermelo and von Neumann brought clarity to the pre-set-theoretic conception of the natural numbers, which on its own was muddled and inadequate. In response, we can take essentially the previous line—exchange the ordered-pair postulate for the second-order PA axioms, and as before we have a satisfactory characterization of the relevant domain which doesn't involve viewing numbers as sets, and which set-theoretic reduction does nothing to improve on. Hence there are some plausible epistemic disadvantages, but no apparent advantages, to treating numbers as sets.

Although Quine was wrong to take set-theoretic reductions as his main examples of epistemically salutary explications, I'm inclined to think there are better ones. Consider, for instance, the replacement of the classical notion (or notions) of infinitesimals with the modern construction due to Robinson.[43] Unlike the ordinary conception of ordered pairs or numbers, the old notion of infinitesimals really was defective, perplexing, and inadequately formulated. In certain circumstances, for instance, the classical analysts had to "neglect" infinitesimal quantities by treating them as equal to zero, while at other times infinitesimals were assumed positive in order to function as divisors. These practices were decried by Berkeley and other critics, and only uneasily tolerated by many working mathematicians. Even Newton and Leibniz themselves made various (unsuccessful) attempts to remove infinitesimals from the calculus, or at least to explain them away as mere heuristic devices with no official standing. Later, the 19th-century "arithmetization of analysis" replaced infinitesimals with the now-standard epsilons and deltas. But Abraham Robinson's work on nonstandard analysis in the 1960s[44] led to a sort of redemption. Robinson identified infinitesimals with certain elements of $^*\mathbb{R}$, the non-Archimedean field of hyperreal numbers, and showed how to do analysis rigorously on the basis of this construction.

The history of the infinitesimal concept furnishes what I take to be a more plausible case of epistemic progress through Quinean explication. The story seems to comfortably fit the Quinean mold: it involves the replacement of a useful but problematic concept with a relatively clear and concrete alternative. The alternative—Robinson's nonstandard reals—may not exactly match what the early analysts had in mind, and it may not have any special claim to being the "one true notion of infinitesimal". But this is no objection, for Quine. What matters is just that the Robinsonian infinitesimals can do the useful work done by the original notion without partaking in its defects. ("We do not claim synonymy. We do not expose hidden meanings... we supply lacks."[45])

[42][Quine 1960], 262.
[43]Quine can't be faulted for overlooking this example; *Word and Object* was published in 1960, while Robinson's work on nonstandard analysis didn't appear until later in the decade.
[44]See [Robinson 1974].
[45][Quine 1960], 258.

Importantly, the Robinsonian explication of the infinitesimal concept is more than just a technical fix. It also represents an epistemic advance over earlier ways of thinking, in that it renders the notion of infinitesimal more intelligible and paves the way for new explanatory proofs in analysis. As F.A. Medvedev says,

> Nonstandard analysis makes it possible to answer a delicate question bound up with earlier approaches to the history of classical analysis. If infinitely small and infinitely large magnitudes are regarded as inconsistent notions, how could they serve as a basis for the construction of so grandiose an edifice of one of the most important mathematical disciplines?[46]

Similarly, according to the historian of analysis Henk Bos, "A preliminary explanation of why the calculus could develop on the insecure foundation of [infinitesimals] is provided by the recently developed nonstandard analysis, which shows that it is possible to remove the inconsistencies without removing the infinitesimals themselves."[47]

To sum up, the problem with Quine's discussion of explication isn't that it describes a nonexistent phenomenon. In the history of mathematics, at least, epistemic progress does sometimes happen in much the way Quine describes. Only his examples are off the mark. Contrary to Quine's claims, arithmetic and the ordered pair were never in need of set-theoretic explication, for two reasons. First, the relevant concepts aren't unclear or defective in the way Quine suggests. Second, set theory is in no position to offer help at any rate, since the sets of modern axiomatized systems like ZFC are if anything more obscure than numbers and ordered pairs. The contrast with the case of infinitesimals—a genuine example of better understanding achieved through explication—makes these points especially clear.

2.7 If not explanation, then what?

A final question is worth addressing before this section ends. If the reduction of arithmetic to set theory has no explanatory value (or other epistemic benefits), as I've argued, then what sort of value does it have? Why bother reducing numbers to sets at all?

One possible answer is that the reduction has no (or very little) value. Someone might think this, perhaps, because they're convinced that a legitimate foundation for mathematics must be a sort of Cartesian-style *epistemic* foundation—i.e., that it should offer greater certainty, clarity, insight, explanation, understanding, or whatever package of epistemic goods one takes to be necessary. This type of view is a recurring theme, for instance, in Hallett's *Cantorian Set Theory and Limitation of Size.* I quote a particularly forceful passage at length:

[46][Medvedev 1998], 664.
[47][Bos 1974], 13.

> Because of the reductionist ambition, we demand that set theory genuinely explain all other mathematical concepts. When I say 'we demand', I mean that is what we should demand. Useful reductionism (philosophically speaking) cannot be just successful theoretical translation, though of course this would suffice in a relative consistency exercise, or as an original part of Hilbert's programme, say. Rather reduction must (in the best instance) be accompanied by a gain in conceptual clarity. Where set theory suffers as a foundation framework is that in general it does *not* bring this conceptual clarity with it. It is no good, philosophically speaking, reducing to set theory something so basic to human thought as the elementary theory of natural number if you cannot also explain why numbers are sets, and why the set concept is even more fundamental. But the set concept is too unclear for any such explanation to be given.[48]

As should be clear by now, I largely agree with Hallett's pessimistic assessment of the epistemology of set theory, and of the reduction of numbers to sets. And Hallett may even be right that an ideal foundation for mathematics would be epistemically illuminating and metaphysically revealing in the ways that set theory often fails to be. If so, we may be justified in continuing to look for an alternative framework—one based on categories or homotopy types[49] rather than sets, maybe.

Still, I want to resist the conclusion that set-theoretic foundations are "no good" or not "useful"—either "philosophically speaking", whatever exactly that means, or otherwise. There are plenty of ways for a foundational theory to earn its keep, and a failure along the dimensions Hallett mentions needn't be a failure across the board. Hence we don't have to put ourselves in the awkward position of claiming, *pace* long-standing mathematical practice, that set-theoretic reductions are a useless boondoggle.

Many people have written about set theory and its role in foundations, but I find Maddy's recent work on this issue particularly enlightening.[50] [Maddy 2011] and [Maddy 2017] take up more or less the question just raised—given that set theory can't serve as a Hallett-style epistemic and metaphysical basis for the rest of mathematics, in what way can it serve as a legitimate foundational theory?[51] In Maddy's view, set theory succeeds in playing a number of distinctively foundational roles, including the following:

[48] [Hallett 1984], 300-301.

[49] See [Linnebo & Pettigrew 2011] and [Ladyman & Presnell 2014], respectively, for discussion of these frameworks as possible foundations for mathematics.

[50] See also [Shapiro 2000], which defends a view similar to Maddy's. For a contrary picture, see [Mayberry 1994].

[51] Maddy's early work, including the 1981 paper discussed above, espoused "set-theoretic realism"—a sort of reductionist view according to which numbers are properties of sets. Maddy has since abandoned set-theoretic realism and its metaphysical commitments, which explains why the recent papers mentioned here strike a quite different tone.

- **Meta-mathematical Corral**: By translating all of mathematics into set theory, it becomes possible, or at least much easier, to obtain general results about consistency and provability for mathematics as a whole. (Cf. Hilbert's program and Gödel's theorems.)
- **Elucidation**: Some set-theoretic realizations of classical mathematical notions introduce a degree of clarity and precision that was lacking in the original concepts (e.g., Dedekind's construction of the real numbers as cuts of the rationals).[52]
- **Risk Assessment**: A proof that a given theory is consistent relative to ZFC (or a lack of such a proof) provides an important measure of the theory's logical "safety" or "riskiness".
- **Shared Standard**: The ZFC axioms function as a shared standard of existence and provability for mathematicians working in diverse specialty areas.
- **Generous Arena**: The set-theoretic universe serves as a single forum in which "all the various structures studied in all the various branches [of mathematics] can co-exist side-by-side, where their interrelations can be studied, shared fundamentals isolated and exploited, effective methods exported and imported from one to another, and so on."[53]

Clearly these functions are important ones, and each of them plausibly pertains to foundations. Moreover, nothing requires that sets be metaphysically or epistemically fundamental in order to play these roles well. So we have a good start on an answer to the question posed at the beginning of this part. Even if the reduction of arithmetic to set theory has no explanatory benefits, this doesn't entail that it and other such reductions are without value. Translating arithmetic and other theories into the language of sets contributes in an important way to a worthwhile end, namely the orderliness, systematicity and communality of mathematics generally.

[52] Maddy's "elucidation" is apparently much the same sort of thing as Quine's "explication". For the record, I have some reservations about the extent to which Dedekind's construction of the continuum is a good example of this phenomenon. One reason to be careful here is because, as Solomon Feferman has argued, there may in fact be several different, and equally worthwhile, conceptions of the continuum. So it's not even obvious what would count as a successful explanation (or explication or elucidation) of continuity and its properties. Moreover, the device of Dedekind cuts does a poor job capturing at least one such conception, namely the one that's operative in Euclidean geometry. As Feferman writes:

> The main thing to be emphasized about the conception of the continuum as it appears in Euclidean geometry is that the general concept of set is not part of the basic picture, and that Dedekind style continuity considerations... are at odds with that picture. It does not make sense, for example, to think of deleting a point from a line, or to remove the end point of a line segment. Given two line segments L and L', we can form a right triangle with legs L_1 and L'_1 congruent to the given segments, respectively; but these share a vertex as a common point, each an end point. Thought of as a set, L is transformed into L' by a rigid motion, and the same for L_1 and L'_1. Thought of in that way, the vertex of the right triangle has displaced one of the end points, but which one? There are many similar thought experiments which dictate that lines, line segments and other figures in Euclidean geometry are not to be identified with their sets of points. ([Feferman 2009], 174-5)

Of course, "Dedekind's continuum" was quite useful for the foundations of analysis and for other purposes. In any case, the issue is complex, and I can't pretend to do it full justice here. (Thanks to the two anonymous referees from *Synthese* for prompting me to say more about this example.)

[53] [Maddy 2017], 305.

3 Implications and further questions

The contention of the previous section, if correct, raises a number of interesting issues for philosophy of science and philosophy of mathematics. I won't try to discuss these issues in detail here—that's a task for future work– but a few are at least worth pointing out and briefly commenting on.

One obvious question to ask is whether or not *all* reductions in mathematics follow the arithmetic-set theory case in being unexplanatory. In my view, the answer is almost certainly no. On the contrary, there are pairs of mathematical theories such that one is plausibly reducible to the other, and also such that the reducing theory sheds significant explanatory light on the reduced theory.[54]

To take one example, consider the advances in algebraic geometry made possible by Alexander Grothendieck's theory of schemes. (Roughly speaking, Grothendieck's idea was to view the *varieties* studied by classical algebraic geometry as a species of *scheme*, a superficially quite different-looking type of object consisting of a topological space with a commutative ring paired to each open set.) It's not hard to see that, given suitably formalized versions of both theories, classical algebraic geometry should come out reducible to scheme theory. Moreover, there's broad agreement among mathematicians that the scheme-theoretic viewpoint provides important insights about various classical phenomena. Indeed, according to Vakil, "the wonderful machine of modern algebraic geometry [i.e., scheme theory and related ideas] was created to understand basic and naive questions about geometry."[55] (To take a particular example of Perrin's, "In the plane... we can always find lines which are not tangent to a given curve. The notion of a scheme is indispensable for understanding these phenomena".[56]) Eisenbud and Harris's textbook on schemes makes the broad scope of Grothendieck's achievements clear:

> The theory of schemes... is the basis for a grand unification of number theory and algebraic geometry, dreamt of by number theorists and geometers for over a century. It has strengthened classical algebraic geometry by allowing flexible geometric arguments about infinitesimals and limits in a way that the classic theory could not handle. In both these ways it has made possible astonishing solutions of many concrete problems.[57]

[54] In light of the infinitesimal example from §2.6 and the other cases considered so far, one might wonder, as an anonymous referee did, whether the difference between explanatory and non-explanatory reductions is just the difference between cases where the relevant concepts weren't or were originally "in good working order". It seems to me that the example given below challenges this idea. We seem to have an explanatory reduction of algebraic varieties to schemes, but there was nothing unclear or contradictory about the notion of variety before Grothendieck came along. And I don't think anyone would suggest that the notion of a scheme—which involves a lot of complex technical machinery with no immediately obvious geometric meaning—is simpler, clearer, more intuitive, or otherwise more epistemically or logically adequate than the notion of a variety. What's going on in this case seems to be, rather, that schemes are *richer in structure* and that they *carry more data* than varieties, and this means that the scheme-theoretic viewpoint allows one to see further into the phenomena than the classical approach allows.

[55] [Vakil 2015], 12.

[56] [Perrin 2008], 213.

[57] [Eisenbud & Harris 2000], 1.

(This case deserves a much more detailed examination, which I hope to give it in future work.) Other examples of explanatory mathematical reductions include, perhaps, the application of Galois theory to questions about the solvability of polynomial equations, and of matroid theory to the notion of "independence".

Supposing this is right, reducibility relations in mathematics evidently come in two different flavors—those that are substantially explanatory and those that aren't (by virtue, perhaps, of being in some sense merely "conventional"). A second question is then how to account for this difference. Is it, for instance, a matter of metaphysics? On this type of view, explanatory reductions are explanatory because the objects in the domain of the reduced theory are metaphysically related to those in the domain of the reducing theory in some appropriate way. This relation could simply be identity, for example. (One might think that the reduction of arithmetic fails to be explanatory precisely because numbers aren't identical to sets.) Or the relation could be something weaker—parthood, maybe, or even metaphysical dependence of a sort peculiar to mathematical objects.[58] This last idea, incidentally, suggests the need for a closer look at the relationship between mathematical explanation and the currently popular notion of metaphysical grounding.[59]

Alternatively, the difference between explanatory and unexplanatory reductions might lie in the logical or epistemological features of the theories themselves, rather than the properties of the objects they describe. On this type of view, explanatory reductions are explanatory because the reducing theory is logically stronger or conceptually richer than the reduced theory, or because it yields new knowledge of an appropriate kind about the reduced theory's subject matter. (I happen to think the second approach has the brighter prospects. But defending that claim is a task for another time.)

At any rate, answering this second question ought to help with a third, related one: why are reductions in empirical science—at least in natural, interesting, and relatively uncontroversial cases—apparently always explanatory, if the same isn't true in mathematics? In addition to the ideas sketched above, one might think that the difference is due to the relatively "empiricist" standards of theory acceptance in force in the natural sciences. Consider string theory, for instance. Contemporary versions of string theory (and its siblings, like M-theory) promise a sweeping reduction-cum-unification of classical particle physics, spacetime theory, and perhaps more besides. As its proponents have pointed out, such a string-theoretic foundation for physics would be formally elegant, conceptually neat, and institutionally convenient—that is, it would provide many of the same benefits that set-theoretic foundations do for mathematics. Unlike set theory, however, string theory has so far failed to win general acceptance.

[58] See [Pincock 2015] for a defense of a dependence-style account.

[59] [Correia & Schnieder 2012] collects some recent work on the subject.

And the proffered reduction of elementary particles to strings is still viewed with skepticism by many physicists.

As is well-known, a major reason for this hesitation is the lack of independent experimental evidence in support of the theory.[60] Without such evidence in its corner, it seems, even the most attractive empirical theory rarely if ever gets itself accepted by the scientific community—and one can't have a proper reduction without a widely accepted reducing theory. On the other hand, if and when we do have good independent evidence of the existence of strings, it seems we'll *ipso facto* have evidence for an explanatory connection between string theory and the theories it reduces. (Since we'll then have good reason to think, for instance, that classical particles are really strings, and that the laws governing the former are just a special case of the laws governing the latter. And such "interlevel" facts—about cross-theoretic identities, composition relations, and the like—are seemingly always explanatory.) So we seem to have the following situation. If an empirical theory is a candidate for involvement in a "purely conventional" reduction—because it has no suitably direct evidential support, but only considerations of formal niceness and unificatory power in its favor—then the theory and the associated reduction won't command general assent. On the other hand, if an empirical theory amasses enough experimental support to be widely endorsed, then this very evidence will underwrite acceptance of the entities and laws the theory describes. And the "interlevel" information typically derivable on this basis is explanatorily valuable by nature.

As we've seen, things are otherwise in pure mathematics. A mathematical theory, and the reduction relations in which it's involved, can win institutional acceptance on the grounds of elegance, neatness and convenience alone; there's no counterpart to the demand for (relatively) direct empirical evidence. Hence unexplanatory, merely "conventional" reductions are possible in mathematics, but typically not in the sciences. (Or so this line of thought would have it. I think the story has some plausibility, but I'm not certain it's entirely or uniquely the right one.)

This essay has covered a lot of ground; let me briefly summarize where we've been. My main goal has been to show that the reduction of arithmetic to set theory is unexplanatory. I've argued that there are both positive and negative reasons to accept this conclusion. On the positive side, I suggested that the reduction of arithmetic is similar in relevant ways to other paradigmatically unexplanatory set-theoretic reductions, and moreover that number theorists don't regard the reduction as a source of insight or understanding. On the negative side, I considered four arguments purporting to show that the reduction of arithmetic is explanatory after all—the first claiming that numbers are identical to sets as a matter of metaphysics, the second claiming that there are explanatory set-theoretic proofs

[60] For an extended critique of this sort, see [Woit 2006].

of arithmetical facts, the third claiming that the reductionist viewpoint explains high-level features of the natural numbers, and the fourth claiming that reductionism renders arithmetical concepts more intelligible. I've tried to show that none of these arguments are successful. This outcome needn't lead to a pessimistic view of set-theoretic foundations, though, since there remain a number of ways in which an unexplanatory reduction to a foundational theory might still be valuable. Finally, as I hope the last section has shown, accepting these conclusions raises a number of provocative questions for philosophers of science and mathematics interested in the epistemology and metaphysics of intertheory relations.

Acknowledgments

Many thanks to Mahrad Almotahari, Kenny Easwaran, Colin Klein, Marc Lange, Daniel Sutherland, Lauren Woomer, and two anonymous referees for conversations and comments that helped improve the paper in countless ways. Thanks also to Jeremy Avigad, Alan Baker and the other participants in the 2015 Mathematical Aims Beyond Justification workshop in Brussels, where I presented an early version of some of this material.

References

[Avigad 2008] Avigad, Jeremy. 2008. "Understanding proofs." In Paolo Mancosu (ed.), *The Philosophy of Mathematical Practice*. Oxford University Press: New York.

[Baker 2010] Baker, Alan. 2010. "Mathematical induction and explanation." *Analysis* 70, 681-689.

[Balaguer 1998] Balaguer, Mark. 1998. "Non-uniqueness as a non-problem." *Philosophia Mathematica* 6, 63-84.

[Benacerraf 1965] Benacerraf, Paul. 1965. "What numbers could not be." *Philosophical Review* 74, 47-73.

[Bos 1974] Bos, H.J.M. 1974. "Differentials, higher-order differentials and the derivative in the Leibnizian calculus." *Archive for the History of the Exact Sciences* 14, 1-90.

[Carter 2008] Carter, Jessica. 2008. "Visualization and understanding in mathematics." In Bharath Sriraman, Claus Michelsen, Astrid Beckmann and Viktor Freiman (eds.), *Interdisciplinary Educational Research in Mathematics and Its Connections to the Arts and Sciences*. Information Age Publishing: Charlotte, NC.

[Correia & Schnieder 2012] Correia, Fabrice and Benjamin Schnieder (eds.). 2012. *Metaphysical Grounding: Understanding the Structure of Reality*. Cambridge University Press: Cambridge.

[Dauben 1979] Dauben, Joseph. 1979. *Georg Cantor: His Mathematics and Philosophy of the Infinite*. Princeton University Press: Princeton.

[Dauben 1988] Dauben, Joseph. 1988. "Abraham Robinson and nonstandard analysis: History, philosophy, and foundations of mathematics." *Minnesota Studies in Philosophy of Science* 11, 177-200.

[Dipert 1982] Dipert, Randall R. 1982. "Set-theoretical representations of ordered pairs and their adequacy for the logic of relations." *Canadian Journal of Philosophy* 12, 353-374.

[Eisenbud & Harris 2000] Eisenbud, David and Joe Harris. 2000. *The Geometry of Schemes*. Springer-Verlag: New York.

[Feferman 1960] Feferman, Solomon. 1960. "Arithmetization of mathematics in a general setting." *Fundamenta Mathematicae* 49, 35–92.

[Feferman 2009] Feferman, Solomon. 2009. "Conceptions of the continuum." *Intellectica* 51, 169-189.

[Hafner & Mancosu 2005] Hafner, J. and P. Mancosu. 2005. "The varieties of mathematical explanation." In P. Mancosu, K. Jørgensen and S. Pedersen (eds.), *Visualization, Explanation and Reasoning Styles in Mathematics*, Springer: Berlin, 215-250.

[Hallett 1984] Hallett, Michael. 1984. *Cantorian Set Theory and Limitation of Size*. Oxford University Press: New York.

[Hempel & Oppenheim 1948] Hempel, Carl and Paul Oppenheim. 1948. "Studies in the logic of explanation." *Philosophy of Science* 15, 135–175.

[Khalifa 2012] Khalifa, Kareem. 2012. “Inaugurating understanding or repackaging explanation?” *Philosophy of Science* 79, 15-37.

[Kitcher 1978] Kitcher, Philip. 1978. “The plight of the Platonist.” *Noûs* 12, 119-136.

[Klein 2009] Klein, Colin. 2009. “Reduction without reductionism: A defence of Nagel on connectability.” *Philosophical Quarterly* 59, 39-53.

[Kleiner 1989] Kleiner, Israel. 1989. “Evolution of the function concept: A brief survey.” *The College Mathematics Journal* 20, 282-300.

[Ladyman & Presnell 2014] Ladyman, James and Stuart Presnell. 2014. “Does Homotopy Type Theory provide a foundation for mathematics?” Unpublished MS.

[Lange 2009] Lange, Marc. 2009. “Why proofs by mathematical induction are generally not explanatory.” *Analysis* 69, 203-211.

[Lange 2010] Lange, Marc. 2010. “What are mathematical coincidences (and why does it matter)?” *Mind* 119, 307-340.

[Lange 2014] Lange, Marc. 2014. “Aspects of mathematical explanation: Symmetry, unity, and salience.” *Philosophical Review* 123, 485-531.

[Linnebo & Pettigrew 2011] Linnebo, Øystein and Richard Pettigrew. 2011. “Category theory as an autonomous foundation.” *Philosophia Mathematica* 19, 227-254.

[Lipton 2011] Lipton, Peter. 2011. “Mathematical understanding.” In John Polkinghorne (ed.), *Meaning in Mathematics*. Oxford University Press: New York.

[Lombrozo 2006] Lombrozo, Tania. 2006. “The structure and function of explanations.” *Trends in Cognitive Sciences* 10, 464-470.

[Lombrozo 2012] Lombrozo, Tania. 2012. “Explanation and abductive inference.” In Keith J. Holyoak and Robert G. Morrison (eds.), *The Oxford Handbook of Thinking and Reasoning*. Oxford University Press: New York.

[Maddy 1981] Maddy, Penelope. 1981. “Sets and numbers.” *Noûs* 15, 495-511.

[Maddy 1990] Maddy, Penelope. 1990. *Realism in Mathematics*. Oxford University Press: New York.

[Maddy 2011] Maddy, Penelope. 2011. "Set theory as a foundation." In Giovanni Sommaruga (ed.), *Foundational Theories of Classical and Constructive Mathematics*. Springer: Dordrecht.

[Maddy 2017] Maddy, Penelope. 2017. "Set-theoretic foundations." In Andrés Eduardo Caicedo, James Cummings, Peter Koellner and Paul B. Larson (eds.), *Contemporary Mathematics, Volume 690: Foundations of Mathematics*. American Mathematical Society: Providence, RI.

[Mayberry 1994] Mayberry, John. 1994. "What is required of a foundation for mathematics?" *Philosophia Mathematica* 3, 16-35.

[Mancosu 2001] Mancosu, Paolo. 2001. "Mathematical explanation: Problems and prospects." *Topoi* 20, 97-117.

[Medvedev 1998] Medvedev, F.A. 1998. "Nonstandard analysis and the history of classical analysis." Translated by Abe Shenitzer. *American Mathematical Monthly* 105, 659-664.

[Moschovakis 2006] Moschovakis, Yiannis. 2006. *Notes on Set Theory*, 2nd edition. Springer: New York.

[Nagel 1961] Nagel, Ernest. 1961. *The Structure of Science*. Routledge and Kegan Paul: London.

[Niebergall 2000] Niebergall, Karl-Georg. 2000. "On the logic of reducibility: Axioms and examples." *Erkenntnis* 53, 27-61.

[Perrin 2008] Perrin, Daniel. 2008. *Algebraic Geometry: An Introduction*. Translated by Catriona Maclean. Springer-Verlag: London.

[Potter 2004] Potter, Michael D. 2004. *Set Theory and Its Philosophy: A Critical Introduction*. Oxford University Press: New York.

[Quine 1960] Quine, Willard Van Orman. 1960. *Word and Object*. MIT Press: Cambridge, MA.

[Pincock 2015] Pincock, Christopher. 2015. "The unsolvability of the quintic: A case study in abstract mathematical explanation." *Philosophers' Imprint* 15 (3), 1-19.

[Rantala 1992] Rantala, Veikko. 1992. "Reduction and explanation: Science vs. mathematics." In Javier Echeverria, Andoni Ibarra and Thomas Mormann (eds.), *The Space of Mathematics: Philosophical, Epistemological, and Historical Explorations*, Walter de Gruyter: New York, 47-59.

[Reck 2016] Reck, Erich. 2016. "Dedekind's Contributions to the Foundations of Mathematics." In Edward N. Zalta (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2016 Edition), URL = <https://plato.stanford.edu/archives/win2016/entries/dedekind-foundations/>.

[Resnik & Kushner 1987] Resnik, Michael D. and David Kushner. 1987. "Explanation, independence and realism in mathematics." *British Journal for the Philosophy of Science* 38, 141-158.

[Robinson 1974] Robinson, Abraham. 1974. *Non-standard Analysis.* Princeton University Press: Princeton.

[Schröder 1895] Schröder, Ernst. 1895. *Vorlesungen über die Algebra der Logik*, vol. 3. Teubner: Leipzig.

[Shapiro 2000] Shapiro, Stewart. 2000. "Set-theoretic foundations." *The Proceedings of the Twentieth World Congress of Philosophy* 6, 183-196

[Sierpinska 1994] Sierpinska, Anna. 1994. *Understanding in Mathematics.* Falmer Press: Washington, D.C.

[Steiner 1978] Steiner, Mark. 1978. "Mathematical explanation." *Philosophical Studies* 34, 135-151.

[Steinhart 2002] Steinhart, Eric. 2002. "Why numbers are sets." *Synthese* 133, 343-361.

[Strevens 2013] Strevens, Michael. 2013. "No understanding without explanation." *Studies in History and Philosophy of Science* 44, 510-515.

[Tao 2015] Tao, Terence. December 28, 2015. "Polymath proposal: explaining identities for irreducible polynomials." *The Polymath Blog.* Retrieved from <http://polymathprojects.org/2015/12/28/polymath-proposal-explaining-identities-for-irreducible-polynomials/>.

[Tappenden 2005] Tappenden, Jamie. 2005. "Proof style and understanding in mathematics I: Visualization, unification and axiom choice." In P. Mancosu, K. Jørgensen and S. Pedersen (eds.), *Visualization, Explanation and Reasoning Styles in Mathematics*, Springer: Berlin.

[Taylor 1993] Taylor, R. Gregory. 1998. "Zermelo, reductionism, and the philosophy of mathematics." *Notre Dame Journal of Formal Logic* 34, 539-563.

[Vakil 2015] Vakil, Ravi. 2015. *The Rising Sea: Foundations of Algebraic Geometry.* Retrieved from <http://math.stanford.edu/~vakil/216blog/FOAGoct2415public.pdf>.

[van Riel 2011] van Riel, Raphael. "Nagelian reduction beyond the Nagel model." *Philosophy of Science* 78, 353-375.

[Warner 1990] Warner, Seth. 1990. *Modern Algebra.* Dover: Mineola, NY.

[Weber & Verhoeven 2002] Weber, Erik and Liza Verhoeven. 2002. "Explanatory proofs in mathematics." *Logique & Analyse* 179-180, 299-307.

[White 1974] White, Nicholas P. 1974. "What numbers are." *Synthese* 27, 111-124.

[Wiener 1914] Wiener, Norbert. 1914. "A simplification of the logic of relations." *Proceedings of the Cambridge Philosophical Society* 17, 387-390.

[Woit 2006] Woit, Peter. 2006. *Not Even Wrong: The Failure of String Theory And the Search for Unity in Physical Law.* Basic Books: New York.

[Zermelo 1909a] Zermelo, Ernst. 1909. "Sur les ensembles finis et le principe de l'induction complète." *Acta Mathematica* 32, 185-193.

[Zermelo 1909b] Zermelo, Ernst. 1909. "Über die Grundlagen der Arithmetik." In G. Castelnuovo (ed.), *Atti del IV Congresso Internazionale dei Matematici* (vol. 2), 8-11. Academia dei Lincei: Rome.