Title
Understanding Arithmetic through Definitions

Author
Nutting, Eileen Susanna

Publication Date
2013

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA
Los Angeles

Understanding Arithmetic through Definitions

A dissertation submitted in partial satisfaction of the
requirements for the degree Doctor of Philosophy
in Philosophy

by

Eileen Susanna Nutting

2013

ABSTRACT OF THE DISSERTATION

Understanding Arithmetic through Definitions

by

Eileen Susanna Nutting

Doctor of Philosophy in Philosophy

University of California, Los Angeles, 2013

Professor Donald A. Martin, Chair

Identifying the source of our mathematical knowledge is an old and challenging problem. In order to avoid postulating cognitive faculties that seem entirely mysterious, such as mathematical intuition, some philosophers have attempted to explain our mathematical knowledge by claiming that it is grounded in definitions, stipulations, or postulations. In this dissertation, I argue against a range of views of this sort in the domain of arithmetic. I argue that none of these views can account for our mathematical knowledge unless we already understand arithmetical structure when we introduce the relevant definitions, stipulations, or postulations. An adequate account of mathematical knowledge must explain how we have this antecedent cognitive grip on the structure of arithmetic. The views under consideration, I argue, are inadequate because they fail to explain this antecedent cognitive grip.

The dissertation begins with a discussion of the problem of mathematical knowledge. In the process, I introduce several desiderata for theories of mathematical knowledge. Definition-based accounts are attractive because they fare well on many of these desiderata.

In the rest of the dissertation, I argue against three such accounts. First, it could not be the case that all of our mathematical knowledge is grounded in axiomatic definitions; we have arithmetical knowledge that could not emerge from axioms. Second, it could not be the case that our ability to think about the natural numbers is grounded in a stipulation of Hume's Principle as an implicit definition; such a stipulation could not uniquely fix referents for number-phrases, and ancient mathematicians could not have stipulated it in a suitable way. And third, our knowledge of basic arithmetical truths is not entirely grounded in the standard simple proofs of those truths, which rely on explicit definitions of the form '4=3+1'. While these proofs surely contribute to the security of our arithmetical knowledge, we can only understand these proofs if we already understand the structure of arithmetic. So, our most basic arithmetical knowledge is cognitively prior to these explicit definitions.

The dissertation of Eileen Susanna Nutting is approved.

John P. Carriero

Yiannis N. Moschovakis

Terence Dwight Parsons

Donald A. Martin, Committee Chair

University of California, Los Angeles

2013

In memory of Matt Garber,

who first encouraged me to pursue philosophy

Contents

ACKNOWLEDGMENTS

I did not write this dissertation in isolation; I had support from both people and institutions. Since I cannot possibly thank them all here, I will limit my specific thanks to those whose contributions most directly aided me in the writing and filing of this dissertation. I hope my family and friends already know how much I appreciate their support in other arenas.

In addition to stable funding, the UCLA Department of Philosophy has provided me with a friendly, encouraging, and invigorating work and learning environment over the past seven years. I thank the department for the financial and philosophical support. I also thank it for the administrative support, primarily in the forms of Kristin Olson and Rachel Lee, who have helped me to navigate UCLA's bureaucratic channels.

I also thank UCLA's Office for Instructional Development, which was my primary source of funding during two academic terms, and UCLA's Graduate Division, which has given me small amounts of supplementary funding on several occasions.

Tony Martin chaired my dissertation committee. His gentle guidance kept me honest and encouraged me to be more charitable in my interpretations of others. His influence has made this dissertation a better project, and it has made me a better philosopher. Quite possibly, it has also made me a better human being. I have trouble imagining a better or more supportive dissertation advisor.

John Carriero was the first to read the early versions of most of what you will find in this dissertation. Countless conversations with him have led me to improve considerably upon those early versions. As my first stop for advice and feedback, John has been an invaluable mentor to me throughout graduate school. He quite probably has become my most significant philosophical influence.

Terry Parsons and Yiannis Moschovakis rounded out my dissertation committee.

Terry's straightforward corrections and good-natured advice have saved me from many a mishap. Yiannis's comments during my preliminary oral exam led me to change my basic angle and approach, making this dissertation a much more interesting project. All four members of my committee are impressive scholars, attentive and helpful advisors, and genuinely good people. I have been privileged to work with them. With much gratitude, I thank them.

Philosophical discussion with people who were not on my dissertation committee also helped me to improve and develop the ideas in this dissertation. In particular, I thank Calvin Normore, Joseph Almog, Keith McPartland, Greg Antill, and Howie Wettstein for the many interesting and fruitful conversations.

I wrote much of this dissertation in coffee shops, while sitting across the table from one or another of a few hard-working friends. For the companionship and studious motivation, I thank Elizabeth Adams, Lee Ann Chae, and Joe Trombello.

Last but not least, I wish to thank Yiannis Moschovakis, Judy Martin, David Kaplan, Gavin Lawrence, John Carriero, and Barbara Herman, for their efforts to ensure that I would be able to file this dissertation in difficult circumstances. And with great admiration, I thank Tony Martin for his signature.

Vita

Education

M.A. Philosophy, University of California, Los Angeles, 2008

B.A. *magna cum laude* Honors Mathematics, Honors Philosophy,
University of Notre Dame, 2006

Presentations

"A Criterion for Theories of Mathematical Knowledge"

Albritton Society, UCLA, November 2012.

"Explicit Definition and Arithmetical Proof"

Philosophy of Language Workshop, UCLA, November 2012.

"The Naturals Amongst the Reals"

Albritton Society, UCLA, March 2012.

"Picking Axioms: A Set Theory Primer"

Albritton Society, UCLA, November 2009.

"Understanding and World-Tracking"

Southern California Epistemology Workshop, July 2009.

"The Nature of Rational Intuition"

Final Honors Research Symposium, Notre Dame, April 2006.

Fellowships, Honors, and Awards

Collegium of University Teaching Fellows. UCLA, 2011-2012.

Philosophy Department Fellowship, UCLA, 2010-2011.

Robert M. Yost Prize for Excellence in Teaching. Philosophy Department, UCLA 2010.

Graduate Summer Research Mentorship. Graduate Division, UCLA, 2007.

President's Fellowship. Graduate Division, UCLA, 2006-2007.

Dockweiler Medal for Philosophy. Notre Dame, 2006.

Senior Thesis Recognition. Arts and Letters/Science Honors Program, Notre Dame, 2006.

Research Experience for Undergraduates in Mathematics (Dynamical Systems). National Science Foundation, University of Wisconsin-Madison, 2004.

R. Catesby Taliaferro Prize. Notre Dame, 2004.

Chapter 1

The Problem of Mathematical Knowledge

Despite widespread agreement about the certainty and security of mathematical knowledge, there is little agreement about how we come to acquire that knowledge. This dissertation will begin with an exploration of the old and challenging problem of identifying the source of our mathematical knowledge. In the first chapter I will gradually unveil six expectations for accounts of mathematical knowledge. In the remaining chapters, I will consider a class of approaches to resolving the problem of mathematical knowledge that appear to fare well with respect to some of these expectations. Approaches of the relevant class attempt to ground all of our mathematical knowledge in some kind of definition, stipulation, or postulation.

Our route through the problem of mathematical knowledge is largely historical. We will begin this chapter with issues that surface in the *Meno*. Next, we will progress to considerations about the role of compass-and-straightedge constructions in geometric reasoning. From there, we will segue into a discussion of an account of mathematical truth that David Hilbert proposed around the year 1900. This view seems to qualify as a 'combinatorial'—and hence a 'knowledge-motivated'—account of mathematical truth, according to a classification introduced by Paul Benacerraf. We will end the chapter with the question of whether or not views like Hilbert's are capable of adequately accounting for all of our mathematical knowledge. In the rest of the dissertation, I will argue that they are not.

1.1 Lessons from the *Meno*

One way to come to appreciate the challenge of accounting for mathematical knowledge is to consider lessons learned from past theories. Recognizing historical problems and appreciating historical insights will also help us to piece together some expectations for an adequate theory of mathematical knowledge.

1.1.1 The Obvious Problems

Accounts of how we come to have mathematical knowledge go back at least as far as Plato's *Meno*. There, Plato presents a theory according to which we acquire mathematical knowledge by recollecting it. There seems to be something right about his analysis; something is apt about the analogy between the way we figure out mathematical truths and the way we piece together things we have forgotten. But as a rule, philosophers reject Plato's account. Those who claim to agree with Plato agree with a very loose interpretation of the view presented in the text.[1]

The main reason that philosophers reject the account of mathematical knowledge in the *Meno* is that it is inextricably intertwined with a view about the soul that strikes most philosophers as far-fetched. According to Plato's view, the soul has always existed and it has knowledge before being united with the body. It then forgets its knowledge in the trauma of birth, but it can recover some of that former knowledge through recollection. Even philosophers who are sympathetic to Plato's account of mathematical knowledge typically

[1] In the *Fifth Meditation*, Descartes suggests that his innatism is in the same vein as the idea of learning as recollection: "And the truth of these matters is so open and so much in harmony with my nature, that on first discovering them it seems that I am not so much learning something new as remembering what I knew before; or it seem like noticing for the first time things that were long present within me although I had never turned my gaze on them before."

reject the pre-existing soul part of the story. Rather than thinking that we acquire knowledge by recollecting something that the soul has forgotten, they think that 'recollection' is a placeholder for some other cognitive process through which we discover a priori knowledge. It is not obvious, however, what that cognitive process is, and what recollection is a placeholder for.

Less commonly noted, but more important for our purposes, if we interpret Plato more straightforwardly, and if we grant that the soul existed before birth, his account of knowledge as recollection nonetheless remains incomplete. His account requires that the disembodied soul have a wealth of knowledge that it forgets at birth. But Plato never explains how the soul comes to acquire that knowledge in the first place. He does gesture at an idea, viz. that because the soul has always existed, it already has seen and learned everything there is to see or know. But this much gives us no explanation of how the soul acquired *that* knowledge; it only explains that the soul did acquire knowledge. Without explaining how the soul initially acquires the knowledge that it forgets and recollects, Plato's account is incomplete.

This kind of reaction to the *Meno* helps us to recognize that we hold theories of mathematical knowledge to the following standard:

(1) If an account posits or otherwise requires some initial cognition of mathematics, the account is incomplete until it identifies how we come to have that initial cognition.

On the grounds of (1), Plato's account in the *Meno* is incomplete, and hence inadequate. It does not explain the source of certain antecedent mathematical knowledge, viz. the mathematical knowledge that the soul has before being embodied and that the soul must forget in order to recollect.

There is little reason to believe that Plato had a well-developed account of how we initially acquire our mathematical knowledge. That said, he does hint at such an account in other dialogues. For example, in the *Phaedo*, Plato (through Socrates) uses perception-like language to describe how we must acquire knowledge: "if we are ever to have pure knowledge, we must escape from the body and observe things in themselves with the soul by itself."[2] Plato writes as though he has some kind of non-perceptual observation in mind. He seems to think that, when it is not constrained by the body, the soul can acquire knowledge by immediately cognizing things like the objects of mathematics.

If Plato does indeed have such an account in mind, then he is positing something like what we now call 'mathematical intuition'. But even positing something like this does not make Plato's account much more satisfying than it was before. Many of us find a faculty of mathematical intuition *spooky*. There seems to be something inscrutable about the cognitive mechanism posited to give the soul immediate access to the objects of mathematics; we do not understand the inner workings of an account that relies on mathematical intuition. The fact that this strikes us as unsatisfying suggests that we expect an account of mathematical knowledge to do more than merely identify the initial source of our mathematical knowledge. We also evaluate accounts according to the following standard:

(2) A theory of mathematical knowledge is unsatisfying if it posits an inscrutable cognitive mechanism.

The role of the pre-existing soul is not the only aspect of Plato's account of mathematical knowledge that we find wanting. It is not clear that Plato has an account of how we initially acquire the knowledge that we forget and then recollect. Even if he does have such an account, it relies on a deeply mysterious cognitive mechanism.

1.1.2 The Insights

It is unsurprising that philosophers reject Plato's account of mathematical knowledge. But if we all take the view to be wrong, why do we still read the *Meno*, and why do many of us think it gets something right? I think the appeal of Plato's account is in its respect for the methods of mathematical discovery. We develop at least some of our mathematical

[2] *Phaedo* 66d-e.

knowledge through reasoning, without relying on empirical observations or the testimony of others. Since all of the resources we use in such reasoning come from within, at least in some sense, it almost seems as if the resulting knowledge is also already there within us, waiting to be (re-)discovered. Part of the appeal of Plato's account is just this: it captures a fairly intuitive idea that knowledge acquired in an a priori way is somehow available to us internally.

But there is more to the appeal of Plato's account. It also captures the process of reasoning in mathematics, which progresses in small steps. While each step seems clearly and naturally correct to us based on the steps we have already taken, some steps are not entirely transparent; sometimes it takes effort to figure out which step to take next, or whether the next idea in the sequence is correct, even though its correctness seems obvious upon reflection. Plato's account in the *Meno* is sensitive to this in two ways. First, through the example of a slave-boy, the *Meno* gives a realistic description of the cycles of confusion, hesitance, and eventual confidence that a person undergoes while working through steps of mathematical reasoning. And second, Plato recognizes that the step-by-step effort that goes into mentally constructing a piece of mathematical reasoning is phenomenologically similar to the step-by-step effort that goes into retracing your actions to remember where you must have left your keys.

If my diagnosis is correct, Plato's account is appealing both because it respects the a priori nature of mathematical reasoning and because it captures the phenomenology of step-by-step mathematical reasoning. All this suggests another feature of a good theory of mathematical knowledge:

(3) A theory of mathematical knowledge should respect the methods of mathematical reasoning, especially mathematical proof, as a source of secure knowledge.

Presumably, respecting the methods of mathematical reasoning also requires an account to be reasonably compatible with the psychology of using reason to extend mathematical knowledge, as Plato's account is.

1.2 Mathematical Reasoning and Hilbert's Geometry

As part of the *Meno* argument that we acquire knowledge through recollection, Plato works through a geometric example. Socrates prods an uneducated slave-boy to reason from a square to the length required to form the base of another square with twice the area of the original. In his questioning, Socrates suggests particular geometric constructions to the boy. For example, Socrates extends the line segment at the base of the original square. He also constructs diagonals between the opposite corners of squares. The boy's reasoning makes use of these constructions.

1.2.1 Constructive and Non-Constructive Geometry

The geometric constructions that Socrates suggests involve two of three constructive procedures that appear in the postulates of Euclid's *Elements*: (i) extending a line segment, (ii) constructing a line between two points, and (iii) constructing a circle from a center point with a given length for the radius.[3] These physical constructions all can be performed with a compass or a straightedge. It is no accident that Plato's example of geometric reasoning is one that relies on such constructions. All of the proofs of non-trivial geometrical theorems that were available to him use at least one constructive procedure, and nearly all of them use all three construction procedures. The geometry available at the time—Euclid's geometry—was essentially constructive.

Not all versions of geometry are constructive.[4] In 1899, Hilbert published a version of geometry in which axioms and definitions replaced all of the constructive postulates.[5]

[3] Actually, Socrates extends line segments by given lengths. This ultimately requires constructing a circle as well, in order to ensure that the extended segment has the proper length.

[4] I am using the term 'constructive' to describe a reliance on straightedge and compass constructions in gometrical reasoning. This is not the standard use of the term in the philosophy of mathematics. The standard use is tied to the interpretation of the existential quantifier: in order to prove that something exists, it is necessary to construct or identify it. I use the term 'constructive' to talk about reasoning that relies on physical constructions, which is rather different.

[5] Hilbert does this in his *Foundations of Geometry*.

For example, Euclid's first postulate is: "To draw a straight line from any point to any point." Hilbert's first axiom is an analog of this postulate: "Two distinct points A and B always completely determine a straight line a." In Euclid's version of geometry, figures come to exist through their constructions. In Hilbert's version, the figures already exist, and await determination; they just need to be picked out. Instead of relying on constructions, Hilbert gives a characterization of Euclidean geometry that is fully axiomatic. Every theorem provable in Euclid's geometry is provable without constructions from Hilbert's axioms.

Giving a fully axiomatized version of Euclid's constructive geometry is not simply a matter of replacing constructive postulates with existential axioms. Euclid's geometry is *essentially* constructive; his method of reasoning only works on the assumption that figures have certain features in virtue of their physical constructions. For example, Euclid's demonstrations depend on the fact that constructed lines and circles that appear to intersect actually do intersect, and that they intersect at a point. But this feature does not appear in the definitions or common notions (axioms) that Euclid sets out. Euclid was able to use this feature because it emerges from the physical constructions; in physical space, the constructed lines or circles actually do intersect.

Accordingly, to give a fully axiomatized theory of Euclidean geometry, Hilbert had to do more than simply substitute axioms for constructive procedures. For the system to work, the axioms must account for all the features that emerge from the construction in Euclid's version. Successfully eschewing Euclid's constructive procedures required Hilbert to otherwise ensure the theoretically significant assumptions that Euclid left unstated.

Furthermore, since a fully axiomatized theory does not rely on constructions, it must be strictly deductive; its theorems must be derivable from its axioms. Consider, then, the kind of logic required to derive theorems from the axioms of Hilbert's fully axiomatized geometry. We have already seen the axiom that replaces the postulate for constructing a line between two points. It states that, for any points A and B, there exists a line a that goes through both A and B. This statement has the form of an $\forall\exists$ statement in our contemporary notation; it requires nested quantifiers. And this is typical of Hilbert's geometric axioms; nearly all of them involve nested quantifiers. Accordingly, it is unsurprising that Hilbert's work axiomatizing geometry emerged in the midst of major advancements in logic in the second half of the 19th century. In particular, these advancements included the development of the earliest systems of logic to handle nested quantification in a systematic way—that is, the earliest systems of polyadic logic. Such developments license the kinds of logical inferences that make it possible to derive theorems from Hilbert's axioms without relying on physical constructions.

1.2.2 Consistency

Hilbert's primary goal in his work on the foundations of geometry was to prove that both Euclidean and non-Euclidean geometries are consistent. The first step in his strategy was fully axiomatizing Euclidean geometry. The second step was providing models of those axioms.[6] To demonstrate the consistency of Euclidean geometry, it suffices to give a consistent model that satisfies all of its axioms. Similarly, to demonstrate the consistency of non-Euclidean geometry, it suffices to give a consistent model that violates the Axiom of Parallels, but satisfies all of the other axioms of Euclidean geometry.

Hilbert builds his models out of pairs and triples of algebraic numbers.[7] Accordingly, his result is one of *relative* consistency. His demonstration of the consistency of Euclidean and non-Euclidean geometries depends on the consistency of his models for them. Those depend on the consistency of the system of algebraic numbers, which in turn depends on

---

[6] Hilbert only provided models in two dimensions, but the generalization into three dimensions is straightforward.

[7] More precisely, the elements of the model are ordered pairs and ordered triples of numbers in a set Ω that Hilbert defines. Ω is the smallest class such that:

- $1 \in \Omega$
- Ω is closed under addition, subtraction, multiplication, and division
- If $\omega \in \Omega$, then $\sqrt{1+\omega^2} \in \Omega$.

Clearly, Ω is a subset of the algebraic numbers.

the consistency of the system of natural numbers. Nonetheless, it seems fairly clear that the natural numbers are consistent, and the algebraic numbers can be modeled in the natural numbers. So it also seems fairly clear that Hilbert's models are consistent.

Intuitively, when Hilbert builds his model of (two-dimensional) Euclidean geometry, he does so using the Cartesian coordinates of points on a Euclidean plane. He defines points to be pairs of algebraic numbers, and he defines lines to be triples of those numbers. Again intuitively, we can think of each line as the unique triple $< a, b, c >$ such that the Cartesian coordinates $< x, y >$ of each point on the line satisfy the linear equation $ax + by + c = 0$.

But it is important to appreciate that our intuitive idea of the ordered pairs as providing locations for physical points is no more than an intuitive idea; in Hilbert's models, the pairs of numbers *are* the points, as opposed to representations of them. If the axioms are strictly about physical points, and pairs of numbers merely represent those physical points, then the axioms are only satisfied if there are points at all of those Cartesian coordinates—i.e. they are only satisfied if space actually satisfies the axioms of Euclidean geometry. In that case, accepting the model would require us to presuppose that Euclidean geometry correctly describes physical space; it would not help to demonstrate the existence of a consistent model. Hilbert makes no assumptions about the nature of physical space. His idea is to demonstrate that the axioms can be mutually satisfied if the terms of those axioms—terms like 'point' and 'line'—are interpreted on a domain of non-spatial entities. He accomplishes this using a domain whose elements are pairs and triples of algebraic numbers—elements that decidedly are not found in physical space.

Hilbert's consistency proofs, then, rely on a somewhat peculiar interpretation of the geometric axioms. His models only satisfy the axioms of Euclidean (or non-Euclidean)[8]

[8]More specifically, Hilbert's non-Euclidean model is a model of hyperbolic geometry. The model takes the pairs and triples corresponding with the inner part of a circular region of the non-Euclidean model, and changes the metric so that lengths closer to the boundaries of the circle are intuitively shorter than congruent lengths nearer the center. Given a line a and a point A that does not lie on a, many lines through A will intersect a outside the boundaries of the circle; since such lines do not intersect a in the model, they are parallel to a in the model.

geometry if non-spatial interpretations of key terms in the axioms are admissible. Such permissiveness about the interpretations of these geometric terms was not entirely uncontroversial at the time; in letters to Hilbert, Frege objected to the consistency proofs on the grounds that, e.g., pairs of numbers are not points.[9]

Setting objections aside, it is clear that Hilbert's approach only yields consistency results for rather abstract geometric axioms. If the meanings of terms like 'point' and 'line' in the axioms are fixed in advance and are constrained to have spatial significance, Hilbert's proof does not work. To demonstrate the consistency of the axioms in this way, Hilbert had to abstract the meanings of axioms away from physical space. The physical constructions of Euclid's geometry served to enforce constraints on the meanings of the key terms in these axioms; in Euclid's geometry, the meaning of 'line' is tied to a straightedge construction, and the meaning of 'circle' is tied to a compass construction. Because he presented a full, non-constructive axiomatization of geometry, Hilbert was able to abstract the meanings of axioms and their key terms from any grounding in physical constructions or physical space.[10]

1.2.3 Implicit Definitions

Though Hilbert presented his axioms in an attempt to characterize the structural features of Euclid's geometry, his axioms are more abstract than Euclid's. In Euclid's geometry, terms have privileged spatial interpretations; this applies both to singular terms like 'point' and 'line' and to relational phrases like 'goes through' (a line goes through a point) and 'lies between' (a point lies between two other points on a line). But in Hilbert's geometry, these terms have no such privileged interpretations; they must refer to pairs of algebraic numbers and to algebraic relations just as well as they refer to spatial elements and to spatial relations.

---

[9] See the Frege-Hilbert correspondence in Frege's *On the Foundations of Geometry and Formal Theories of Arithmetic*.

[10] Hilbert was actually skeptical about the idea that geometric terms could have precise, antecedently-intended interpretations in physical space. In a Dec. 29, 1899 letter to Frege, Hilbert writes, "If one is looking for other definitions of 'point', e.g. through paraphrase in terms of extensionless, etc., then I must indeed oppose such attempts in the most decisive way; one is looking for something one can never find because there is nothing there."

Because these terms lack privileged interpretations, they play a role somewhat like that of variables. They merely require consistent interpretation within and across the axioms, e.g., so that all instances of 'point' get the same interpretation, and similarly for instances of 'lies between'.

Hilbert's mathematically-motivated abstraction, then, changes the semantics of geometric terms; it changes the meaning or reference of these terms. And when semantics change in this way, truth-conditions also must change. Claims like the Axiom of Parallels may well have straightforward truth-values when their terms have intended interpretations. But just as a statement containing unbounded variables might lack a straightforward truth-value, the abstract, uninterpreted version of the Axiom of Parallels seems to lack a straightforward truth-value; after all, its terms lack privileged meanings.[11] And if axioms lack straightforward truth-values, it is natural to think that the theorems that follow from them also lack straightforward truth-values. If the notion of truth is to apply to Hilbert's abstract axioms, or to anything else in abstract mathematics, there must be a way to introduce meaning and truth without requiring that abstracted terms have privileged interpretations.

Hilbert's theory of truth and semantics in geometry, and presumably in other areas of mathematics as well, fits with these considerations. On his account, a system of axioms implicitly defines its central concepts by setting out relations between those concepts. For example, the meaning of the concept 'point' is determined by its relations to other core geometric concepts, both concepts of kinds (e.g. 'line') and relational concepts (e.g. 'lies between'). The axioms fix what these relations are; so, by identifying a relation between

---

[11] The Axiom of Parallels states that, for any line a and any point A, if A does not lie on a, then there is a unique line b through A that does not intersect with a. If the terms 'point' and 'line', and the relations 'lie on' and 'intersects' have antecedently fixed meanings, then this is a straightforward claim, the truth of which depends on the way points and lines are. But if these terms lack fixed meanings, then it is not a precise enough claim to have a straightforward truth-value. The claim looks like:

$$(\forall x)(\forall y)((Lx \wedge Py \wedge \neg yLOx) \rightarrow (\exists! z)(Lz \wedge yLOz \wedge \neg xIz)),$$

where 'Lx' means 'x is a line', 'Py' means 'y is a point', '$yLOx$' means 'y lies on x', and 'xIz' means 'x intersects z'. Without fixing an interpretation for the property and relation variables, the statement simply lacks a truth-value.

points and lines, Hilbert's axiom that any two points determine a line in part fixes the meanings of both of the concept 'point' and the concept 'line'.[12]

One feature of this view is that concepts like 'point' and 'line' are defined within a specific system of axioms. Part of what fixes the meaning of the concept 'point' in Euclidean geometry is the Axiom of Parallels, and part of what fixes the meaning of the concept 'point' in non-Euclidean geometry is the negation of the Axiom of Parallels. Accordingly, different kinds of geometry have different 'point' concepts (and 'line' concepts, and 'lies between' concepts). Thus, since things in Hilbert's Euclidean model—including pairs and triples of algebraic numbers, and algebraic relations—mutually satisfy the axioms of Euclidean geometry, the pairs of algebraic numbers in those models *are* (Euclidean) points; they fully satisfy the implicitly-defined concept of (Euclidean) 'point'. Notice, then, that because of the way in which the concepts of Euclidean geometry are fixed, every two instances of Euclidean points will determine a Euclidean line. Hilbert's first axiom is true.

Ultimately, then, all of the axioms of Euclidean geometry are true. And so are the axioms of non-Euclidean geometry. Since the two systems of axioms involve different concepts, they do not contradict each other. The axioms are all made true by the definitions of the concepts that appear within them. Those definitions, in turn, are true because they are set out as implicit definitions. In Hilbert's words, "as soon as I have laid down an axiom, it exists and is 'true' ..."[13] A system of axioms is laid down, or stipulated, to implicitly define the concepts of the theory the system of axioms describes. As long as the axioms laid down are consistent, they delineate coherent concepts and the axioms are true.

Hilbert's interest in consistency led him to a new theory of semantics for geometry, and a new theory of geometrical truth. Geometric truths are conceptual truths, true in virtue of the meanings of the concepts of geometry. Those meanings are introduced by

[12] "In my opinion, a concept can be fixed logically only by its relations to other concepts. These relations, formulated in certain statements, I call axioms, thus arriving at the view that axioms (perhaps together with propositions assigning names to concepts) are the definitions of the concepts." - Letter from Hilbert to Frege, Sept. 22, 1900.

[13] Hilbert to Frege, Dec. 29, 1899.

implicit definitions, and those implicit definitions are fixed by the stipulation of a system of axioms. Ultimately, that is, on Hilbert's view the truths of geometry are true in virtue of stipulations. But now that we have set out his view on semantics and truth, we steer our discussion back towards knowledge. We will return to Hilbert shortly.

1.3 Benacerraf Problems: Truth and Knowledge

Our discussion began with the problem of mathematical knowledge, transitioned to a discussion of mathematical reasoning, continued to considerations about mathematical truth and the semantics of mathematical terms like 'point' and 'line', and now gradually returns to mathematical knowledge. This is fairly natural; in the philosophy of mathematics, concerns about mathematical knowledge and concerns about mathematical truth tend to bleed into one another. The inter-connectedness of theories of truth and theories of knowledge in this area, however, can raise challenges for theories of either type. In "Mathematical Truth," Paul Benacerraf suggests that there are fundamental obstacles to reconciling a plausible account of mathematical knowledge with a plausible account of mathematical truth.

According to Benacerraf's diagnosis, views in the philosophy of mathematics fit into one of two categories. Some are truth-motivated, presenting accounts of mathematical truth and semantics that are meant to cohere with more general theories of truth and semantics. Others are knowledge-motivated, presenting accounts of mathematical knowledge that are meant to cohere with more general theories of knowledge. Benacerraf argues that truth-motivated views inevitably sacrifice a plausible epistemology, and that knowledge-motivated views inevitably sacrifice a plausible semantics and/or a plausible theory of truth. If he is right, developing a general theory that is satisfying in both domains is a rather dim prospect.

1.3.1 Sacrificing Knowledge

Benacerraf starts setting up his problem by identifying two minimal standards that he thinks any adequate theory of mathematical truth must satisfy. One is that of having "the account of mathematical truth mesh with a reasonable epistemology."[14] Since we think it should be possible to account for our mathematical knowledge, and since we want to avoid restricting ourselves to unreasonable epistemic theories, this standard seems to be eminently reasonable. In fact, its dual seems like an important addition to our list of expectations for theories of mathematical knowledge:

(4) A theory of mathematical knowledge should be compatible with a plausible account of mathematical truth.

Notice that we have not yet suggested substantive restrictions on how a plausible theory of mathematical truth or a plausible theory of mathematical knowledge ought to go. Benacerraf will have ideas on both of these fronts.

The second minimal standard Benacerraf introduces at the beginning of his paper is that of "having a homogenous semantical theory in which semantics for the propositions of mathematics parallel the semantics for the rest of language."[15] Though this seems like a fairly natural standard for *semantical* theories, it strikes me as somewhat less transparent as a minimal standard for a theory of mathematical *truth*. A semantical theory might primarily be a theory about meaning and/or reference, rather than a theory about truth or truth-conditions. A more obvious standard would be that of having a theory of mathematical truth cohere with a more general theory of truth that applies to non-mathematical cases. As it turns out, Benacerraf's ideas about what makes a theory of truth plausible ultimately emerge from his ideas about what makes a good semantical theory—that is, a good theory of meaning and reference for mathematical terms. This is not an unreasonable approach; the connections between truth, meaning, and reference are deep and fundamental.

[14] p.661

[15] Ibid.

According to Benacerraf, then, there are two minimal standards for a theory of mathematical truth: (i) that it mesh with a reasonable epistemology, and (ii) that the semantics of mathematical claims parallel the semantics of non-mathematical claims. On their own, it is not obvious how these two standards give rise to a problem for the possibility of a comprehensive philosophy of mathematics. To motivate the problem, Benacerraf relies on his division of accounts in the philosophy of mathematics into those motivated by knowledge and those motivated by truth. He argues that truth-motivated views must violate (i)—they cannot mesh with a reasonable epistemology—while knowledge-motivated views must violate (ii)—they require a semantics for mathematical language that deviates wildly from semantics for non-mathematical language. Assuming that any respectable account in the philosophy of mathematics must fit into one of these categories, then, no respectable account can satisfy both minimal standards.

Truth-motivated views, on Benacerraf's classification, are those that are meant to cohere with a more general theory of truth and semantics. Benacerraf also thinks that the best general theory of truth must be essentially Tarskian; he tells us that Tarski "has given us the only viable systematic general account of truth."[16] Since he thinks this is the only plausible general theory of truth, the views that he classifies as 'truth-motivated' are those that have Tarskian truth-conditions. Benacerraf goes on to identify Gödel as "one of the most explicit and lucid proponents" of such an account.[17] In Gödel's ontology, there are mathematical objects like sets and numbers, and these objects "clearly do not belong to the physical world."[18] Though Benacerraf does not explain how exactly this ontology yields a theory of mathematical truth, he presumably thinks that, for Gödel, mathematical truths are true in virtue of these objects and the relations that hold between them.[19] Such an

[16]p.670

[17]p.674

[18]Ibid, quoted from Gödel's "What is Cantor's Continuum Problem?" p.271 in Benacerraf and Putnam.

[19]It is not obvious to me that this is Gödel's theory of mathematical truth. On his view, there are both mathematical *concepts* and mathematical *objects*. The objects bear all the relations to each other that are required by the concepts. Since Benacerraf never mentions the concepts, he seems to assume mathematical

account would fit a generally Tarskian mold because, e.g., '$2+3=5$' is true if and only if $2+3=5$. That is, '$2+3=5$' is true if and only if the objects 2, 3, and 5 bear the proper relations to one another.

It may seem odd that Benacerraf offers Gödel's account as a paradigmatic example of a truth-motivated view, given that Gödel describes a view about ontology instead of a theory of truth. It also may seem odd that Benacerraf terms truth-motivated accounts 'platonistic', when mathematical platonism is primarily an ontological theory. Hilbert's account, for example, seemed to be motivated by concerns about mathematical truth, but it does not seem to require a platonistic ontology.[20]

I believe Benacerraf takes this ontological angle because he has a rather robust conception of what it is for a theory of truth to be Tarskian. The most conspicuous feature of Tarski's account of truth is that it is *disquotational*; Tarski tells us that 'p' is true if and only if p.[21] But the mere fact that truth is disquotational does not entail anything about ontology. In order for a disquotational theory of truth to have ontological implications, it must also include a theory of semantics—a theory of meaning or reference—for the right-hand side of the biconditional. A merely disquotational theory need not include such a theory of semantics.

Benacerraf clearly includes more than mere disquotation in his interpretation of what qualifies a theory of truth as Tarskian. He casts Tarskian views as ones that analyze propositions "in terms of the names and quantifiers they might contain and in terms of the properties

truth is grounded in the objects, rather than the concepts. It is not clear to me that this is right.

[20] Benacerraf offers some discussion of Hilbert in this paper. But the Hilbert described by Benacerraf seems rather different from the Hilbert described here; the views and motivations are not the same. The view I discuss is Hilbert's view in 1899-1900. The view Benacerraf discusses appears 25 years later in his 1925 address "On the Infinite." (Republished in Benacerraf and Putnam's *Philosophy of Mathematics: Selected Readings*.) Hilbert's views changed in those 25 years. In the earlier period, his philosophical views seem to have been primarily motivated by mathematical concerns about things like consistency. In the later period, his philosophical views seem to have been motivated by more philosophical concerns about knowledge beyond the limits of observation.

[21] More precisely, Tarski describes truth in terms of a schema of such sentences, called 'T-sentences'. Among other places, this can be found in his 1944 paper "The Semantic Conception of Truth: And the Foundations of Semantics," published in *Philosophy and Phenomenological Research.*

they assign to the objects within their domains of discourse."[22] This suggests that Benacerraf thinks the right-hand side of Tarski's disquotational biconditionals must be interpreted in such a way that singular terms refer to objects. Though controversial, and though Benacerraf might have a rather ontologically-loaded idea of what reference to objects involves, it is fairly standard to interpret Tarski along these lines.

If Benacerraf is right and the right-hand side of Tarskian biconditionals must be interpreted according to a theory of reference like the one suggested above, and if truth-motivated accounts must be Tarskian, then truth-motivated accounts must be ones that involve reference to mathematical objects of one sort or another. On this construal, Gödel's platonistic account is indeed a paradigmatic example of a truth-motivated account, but only because it is an account on which mathematical singular terms (e.g. 'the empty set') refer to mathematical objects.

Truth-motivated views in the philosophy of mathematics, on Benacerraf's construal, are views that provide plausible accounts of mathematical truth. Regardless of whether we are inclined to agree with Benacerraf about what makes an account of truth plausible, it does seem correct to think that a plausible account of semantics and truth in the philosophy of mathematics must be in keeping with a good account of semantics and truth in general. Similarly, a plausible account of knowledge in the philosophy of mathematics must be in keeping with a good account of knowledge in general. Benacerraf describes it this way: a plausible theory of knowledge is one in which "it must be possible to link up what it is for p to be true with my belief that p;" such a theory will provide "an account of the link between our cognitive faculties and the objects [or truths] known."[23]

These expectations should seem familiar to us. The possibility of linking up the truth of p with my belief that p is essential to satisfying (4)—the expectation that a theory of mathematical knowledge cohere with a plausible theory of mathematical truth. The

[22] p.665

[23] p.667; p.674.

expectation that there be an account of the link between our cognitive faculties and the objects or truths known is tied in with (2)—the expectation that an account of mathematical knowledge posit no inscrutable cognitive mechanisms.

In addition, Benacerraf's presentation draws attention to an important point that we have not yet addressed directly. In considering what qualifies as a good theory of mathematical knowledge, we ought to pay heed to theories of knowledge in areas outside mathematics. After all, mathematical knowledge is a species of knowledge. Accordingly, as Benacerraf makes clear, we have another basic expectation for a theory of mathematical knowledge:

(5) A theory of mathematical knowledge should be contiguous with a plausible general theory of knowledge.

Benacerraf seems to be right about this much. But recall that he has rather restrictive ideas about what is required for a general theory of truth to be plausible; he assumes that a good theory of truth must be a Tarskian theory interpreted according to an ontologically robust semantical theory. As it turns out, he has equally restrictive ideas about what is required for a general theory of knowledge to be plausible.

Benacerraf's positive sketch of an account of knowledge is not terribly popular. In his words, he "favor[s] a causal account of knowledge on which for X to know that S is true requires some causal relation to obtain between X and the referents of the names, predicates, and quantifiers of S."[24] This suggests that Benacerraf thinks a plausible account of knowledge in general ought to be modeled closely on a standard kind of account of perceptual knowledge of medium-sized physical objects. Presumably, such a model also motivates his favored theory of reference; on a fairly standard theory of reference for medium-sized physical objects, singular terms refer to objects.

Aspects of Benacerraf's characterization of truth-motivated accounts in the philosophy of mathematics strike me as potentially objectionable. First, it is not clear that accounts motivated by concerns about truth must be Tarskian. For example, Hilbert's view appears

[24] p.671

to be motivated by concerns about truth that arise out of mathematics, rather than concerns about truth that arise out of a more strictly philosophical interest in maintaining coherence with a general theory of truth. It is wholly unclear to me whether or not his view must be interpreted as Tarskian. Second, as both a matter of interpretation and a matter of truth, it is not clear that Tarskian biconditionals must be interpreted according to a semantical theory on which singular terms refer to objects.

It is clear, however, that Benacerraf's overall characterization is overly restrictive. There are respectable views in the philosophy of mathematics, such as Hilbert's, that are motivated by concerns about truth but nonetheless are excluded from Benacerraf's 'truth-motivated' classification because of the substantive restrictions Benacerraf sets on theories of truth. Truth-motivated accounts in the philosophy of mathematics need not be platonistic.

At some level, however, I am merely quibbling about terminology. Benacerraf presents a classification of 'truth-motivated' accounts in the philosophy of mathematics. Whether or not it is misleadingly-termed, the accounts that fit the classification are platonistic; they involve reference to mathematical objects and their properties.[25] Furthermore, Benacerraf thinks that a plausible account of mathematical knowledge should be causal. If we grant him this framework, there is some intuitive appeal to the thought that truth-motivated accounts must make epistemic sacrifices.

The example of Gödel's platonistic view might help us appreciate the sorts of epistemic sacrifices at issue. Gödel posits a perception-like faculty of mathematical intuition through which we are able to interact with the mathematical abstracta that ground mathematical truth. But, as Benacerraf notes, such a story about mathematical knowledge lacks "an account of the link between our cognitive faculties and the objects known."[26] This point echoes a concern about mathematical intuition raised in our discussion of the *Meno*. One of our standards for an adequate account of mathematical knowledge is (2)—it should

[25]The mathematical objects on truth-motivated accounts need not be abstract, however. Benacerraf's classification includes views on which mathematical objects are inscriptions or strings of symbols. See p.665.

[26]p.674

not involve inscrutable cognitive mechanisms. And Gödelian intuition does appear to be inscrutable. Even if we set aside Benacerraf's causal requirement and the fact that the link between our belief that p and the truth of p seems to involve no causal relation, Gödel's account of mathematical knowledge does not seem to mesh with a plausible epistemology.

It seems, then, that there is some intuitive appeal to the idea that a truth-motivated account, on Benacerraf's construal, must make epistemic sacrifices. It cannot mesh with a epistemology that Benacerraf considers plausible—i.e. one that requires causal relations for knowledge. But we have said enough about truth. Our main concerns here are about constraints on theories of mathematical knowledge. It is time that we returned to epistemology.

1.3.2 Benacerraf's 'Combinatorial' Classification

Benacerraf calls views that he takes to be motivated by epistemic considerations "'combinatorial' views of the determinants of mathematical truth."[27] He primarily characterizes these views in a negative way, contrasting them with 'truth-motivated' views. In his words, "truth is conspicuously not explained in terms of reference, denotation, or satisfaction" on combinatorial accounts.[28] Ultimately, such views entail that mathematical discourse is about something other than mind-independent objects and their properties. Mathematical truth is not tied to a semantics of reference and denotation. In short, mathematical truth is not Tarskian on combinatorial accounts. At least, it is not Tarskian according to Benacerraf's peculiar construal of what it is for an account to be Tarskian. On this characterization, all accounts of mathematical truth are indeed either truth-motivated (Tarskian) or knowledge-motivated (non-Tarskian). And surely, 'knowledge-motivated' accounts on such a classification must sacrifice a plausible theory of truth, if the only plausible theory of truth is Tarskian.

While he primarily characterizes combinatorial views as non-Tarskian, Benacerraf also closely identifies such views with more interesting positive characterizations that more

[27]p.665

[28]Ibid.

clearly connect combinatorial views with epistemic motivations. He seems to be targeting a class of views on which mathematical truth is constructed, by way of combinatorial manipulations, from cognitively accessible starting points like axioms. According to his description, these accounts take the fact that much mathematical knowledge is derived through proofs to suggest that truth itself is grounded in proof.[29] To help us identify this class of views, Benacerraf suggests three kinds of accounts of mathematical truth that qualify as combinatorial: (a) conventionalist accounts, (b) "certain views of truth in arithmetic on which the Peano axioms are claimed to be 'analytic' of the concept of number," and (c) views according to which mathematical truth is a matter of formal derivability from the axioms of a theory.[30]

Depending on how both Benacerraf and Hilbert are interpreted, Hilbert's view can plausibly be thought an example of any of these three categories of knowledge-motivated views; it is not, you may recall, a truth-motivated view on Benacerraf's classification. I believe it is an example of the first two categories, but not the third. Hilbert's view seems to be of type (a)—a conventionalist view—because it is a view on which the truths of geometry are fixed by stipulation, by the setting out of a system of claims. Admittedly, his view only requires *consistency* in the axioms set out for implicit definitions to be fixed; it does not depend on any *agreement* within a group. Hilbert's view most likely is the kind of conventionalist view Benacerraf has in mind. But if not, it is at least a precursor to the conventionalist views of truth that Benacerraf does have in mind.

Setting aside the fact that Hilbert gives his account in the context of geometry instead of arithmetic, his view also seems to be a likely example of type (b). Since his view is one on which the axioms of a theory implicitly define its central concepts, it is also a view on which the axioms are analytic of the relevant mathematical concepts. But this is not enough to determine whether or not Hilbert's view fits under this second kind of combinatorial accounts, which Benacerraf limits to 'certain views' on which axioms are analytic. After all,

[29]See p.675.

[30]p.665

Benacerraf doesn't say *which* views he has in mind. Presumably, given his general conception of combinatorial views, the views that qualify as examples of the second type of combinatorial view are not ones according to which analyticity emerges from robust pre-axiomatic concepts of the natural kinds of mathematics. Rather, the relevant views are ones on which the relevant mathematical concepts somehow emerge from combinatorial constructions. Since analyticity follows from axiomatic definition on Hilbert's view, his view seems to yield the analyticity of the axioms for combinatorial sorts of reasons. So, presumably the way his view renders the analyticity of axioms qualifies it as a view of type (b).

According to some interpretations of Hilbert, his view also matches Benacerraf's description of type (c) of combinatorial views—views according to which mathematical truth is grounded in formal derivability from the axioms of a theory.[31] I do not interpret Hilbert this way. Presumably he thinks that formal derivation is truth-preserving, and hence that theorems derived from axioms are true. But he never claims that truth is restricted to the axioms and what is formally derivable from them. All Hilbert claims is that the meanings of the relevant concepts are implicitly defined by the system of axioms in which they appear. He also seems to think that all geometric truths follow from the axioms. This much is compatible with assigning truth to satisfaction in all models of the axioms, instead of assigning truth to formal derivation from the axioms.

Admittedly, Hilbert seems to have assumed that formal derivability and satisfaction in all models are co-extensive. Nonetheless, he might not have thought that formal derivation grounds truth, properly speaking. He might instead have thought that a sentence is true in a theory (e.g. Euclidean geometry) in virtue of the fact that any model that satisfies the axioms of that theory must also satisfy that sentence. In this way, a true sentence would be necessitated by the structural (or conceptual) restrictions imposed by the axioms. Formal derivation, then, would be a way of tracking those sentences that are satisfied in all models of the theory. Whether or not this was Hilbert's view is an interpretive matter, and I do not

[31] Patricia Blanchette interprets Hilbert this way in "Frege and Hilbert on Consistency."

intend to focus on it here.[32] However, it is worth noting that Hilbert's view that the axioms of geometry implicitly define the core geometrical concepts appears to be compatible with a satisfaction-style account of mathematical truth, and not only a derivation-style account.

This point is important because Hilbert was mistaken in thinking that satisfaction in all models of a set of axioms is co-extensive with derivability from the axioms. Admittedly, these features are co-extensive if the axioms in question are axioms in a *first-order* language; together, the Soundness and Completeness Theorems for first-order logic yield precisely this result.[33] But in more robust languages, these features are not co-extensive. For certain sets of axioms, there are sentences that are not derivable, but that are satisfied by all models of those axioms. This is a consequence of Gödel's Incompleteness Theorems.

Gödel's First Incompleteness Theorem establishes that, if a system of axioms is both consistent and strong enough to characterize the structure of the natural numbers in a fairly robust way, it inevitably leaves some arithmetical sentences undecidable. That is, some arithmetical sentences are neither provable nor disprovable from the axioms by formal derivation.[34] This result is particularly striking given that it is possible to uniquely charac-

[32] I suspect that Hilbert most fundamentally thought of geometric truth as a matter of satisfaction in all models, at least around 1900. His interest in consistency seems to be an interest in whether or not a set of axioms has a model, rather than an interest in deductive consistency. Furthermore, in a Dec. 29, 1899, letter to Frege, when emphasizing the abstract nature of his axioms, he writes:

> But it is surely obvious that every theory is only a scaffolding (schema) of concepts together with their necessary connections, and that the basic elements can be thought of in any way one likes. E.g. instead of points, think of a system of love, law, chimney sweep . . . which satisfies all the axioms; then surely Pythagoras' theorem also applies to these things. (42)

I interpret the comment about Pythagoras's theorem as a comment about satisfaction, not derivation.

But others may disagree with my interpretation. For example, in "Frege and Hilbert on Consistency," Blanchette presents Hilbert's view as one on which what it is for a set of axioms to be consistent is for no contradiction to be logically deducible from the axioms. She supports this claim by pointing out something Hilbert says about the *consequences* of the axioms. It is not clear to me why she thinks the consequences Hilbert has in mind must be ones that can be logically deduced.

[33] For a discussion of what makes a language first-order or second-order, see Chapter 2.

[34] This is a very rough characterization of the theorem. The system of axioms must also be *recursively enumerable* (roughly, a computer must be able to generate the set of axioms), and the rules of inference must be *effective* (this excludes the use of an ω-rule, which requires infinitely many premises). Also, whether or not a sentence is 'provable by formal derivation' is about the existence of a proof—i.e. a derivation in finitely many steps. Thus, whether or not a sentence is 'provable' in this sense is relative to the structure of the model of the arithmetical axioms in which the sentence is being interpreted. Different models of the

terize the structure-type of the natural numbers using second-order axioms, which ensures that any arithmetical sentence true in one model of the second-order axioms is true in all models of those axioms.[35] But Gödel's First Incompleteness Theorem ensures that, since the second-order axioms are able to characterize the natural numbers in a sufficiently robust way, some second-order arithmetical sentences are undecidable on the basis of those axioms. That is, there is a gap between satisfaction and derivability. There is no general system of derivation for second-order statements, so formal derivation is inevitably limited to first-order statements. Accordingly, there are statements that cannot be derived from the system of second-order axioms, but that are nonetheless satisfied in every model of the system.

Because of the second-order implications, this theorem is commonly interpreted to state that, given a fixed system of axioms, not all arithmetical truths are provable. Such an interpretation relies on two assumptions: first, that there is a unique structure-type of the natural numbers in which every arithmetical sentence is either satisfied or not; and second, that arithmetical truth is a matter of satisfaction in that unique structure or structure-type. Given these assumptions (which I am happy to make), Gödel's result poses fundamental problems for views on which truth is limited to axioms and theorems that are formally derived from those axioms.

Accordingly, if Hilbert's view limits truth to formal derivability from axioms, it cannot be extended from geometry to arithmetic. Such a restricted view cannot work as a general theory of mathematical truth.[36] It would be, as Benacerraf says of derivation-based views, "torpedoed by the incompleteness theorems."[37] Benacerraf's concerns about knowledge-motivated (combinatorial) views seem to be well-placed in the case of derivation-based views.

---

arithmetical axioms sometimes lead to differences in what sentences are provable. The issue of different models of arithmetic will be explored at much greater length in Chapter 2.

[35]This unique characterization will also be discussed in Chapter 2.

[36]I do not intend to suggest here the inference that Hilbert specifically avoided a derivation-restricted view, nor that he did so because such a view would run into incompleteness problems. The former claim, I believe, is false. The latter claim would be anachronistic; Gödel's 1931 incompleteness result was not well anticipated by the mathematical community.

[37]p.665.

Combinatorial views of type (c) must sacrifice a plausible account of truth; such views must deny truth of any claims that are entailed by the (true) axioms, but are not formally derivable from them.

But it is not clear that Benacerraf's concerns properly extend to all combinatorial views. Hilbert's view seems to be categorized as a combinatorial, knowledge-motivated account on Benacerraf's classification, but it is not obvious that it must sacrifice a plausible theory of truth. After all, Hilbert seems to have been focused on rendering a plausible theory of truth capable of preserving the consistency of different systems of axioms. And what we have seen of his view—viz. that all geometrical truths are grounded in implicit definitions given through axioms—seems to be compatible with a satisfaction-style account that does not sacrifice truth in the same way that, in light of incompleteness, derivation-style accounts must.

Admittedly, one of Benacerraf's primary characterizations of the 'combinatorial' classification is that such views do not explain truth in terms of reference, denotation, *or satisfaction*. Perhaps Benacerraf's characterization excludes any view that assigns truth on the basis of satisfaction, even a view like Hilbert's that both grounds all mathematical truth in the setting out of a system of axioms and explicitly rejects the idea that any entities or relations that satisfy the axioms are privileged referents of the relevant terms.

To be honest, it is not clear to me how Benacerraf would classify a satisfaction-style view in Hilbert's vein. Either it is combinatorial, and hence knowledge-motivated according to his classification, or his classification neither includes it as truth-motivated nor includes it as knowledge-motivated. If the latter, Benacerraf is targeting a very limited set of views in the philosophy of mathematics. But instead of focusing on the most limited class of views that could be construed as combinatorial, let us focus on a broader class that could plausibly be characterized as knowledge-motivated. This broader class presumably includes views like the satisfaction-style interpretation of Hilbert's view.

1.3.3 The Epistemic Appeal

By Benacerraf's lights, the chief virtue of combinatorial accounts in the philosophy of mathematics is their friendliness to plausible epistemic theories. Hilbert's view—whether or not it qualifies as 'combinatorial', and whether or not it is satisfaction-based—is equally friendly to such theories. What makes all of these views epistemically promising is, first, that the kinds of things they identify as the grounds of mathematical truth (e.g. stipulations and definitions) are the kinds of things that are cognitively accessible in familiar ways, and second, that they draw tight connections between truth and proof in mathematics.

Benacerraf never ventures an explanation of how we come to know the axioms or starting-points of views in the combinatorial vein, but Hilbert's view provides a nice example of how such an explanation might go. On Hilbert's view, all geometrical truth is grounded in the truth of the axioms of geometry. The axioms, in turn, are true because they are stipulated as the implicit definitions of concepts. Accordingly, our knowledge of the grounds of geometrical truth—that is, the axioms—is no more complex than our knowledge of any other stipulated definitions. Such knowledge is almost trivial. We know these truths independently of our knowledge or examination of any independent subject matter; we know them because we consciously introduce them. It is easy for us to know the foundations of mathematical truth—in Hilbert's case, geometrical truth—because we ourselves generated those foundations by setting them out as truths.

The way in which we come to know further mathematical truths depends on how those truths are grounded in the known foundations of mathematics. In the case of Hilbert's view, the known foundations are axioms. If truth is identified with formal derivability, then in the words of Benacerraf, "We need only account for our ability to produce and survey formal proofs" to explain this further knowledge.[38] If truth is identified with logical derivation (which need not be formalized and written), then explaining our further mathematical knowledge is a matter of accounting for our ability to reason deductively. To be fair, it is

[38]p.668

not a simple task to account for either of these abilities. But once the problem of mathematical knowledge reduces to explaining these abilities, it no longer seems that the remaining task is to resolve the problem of *mathematical* knowledge. What remains to be explained is something much more basic than that—perhaps inferential knowledge, or some other central cognitive process with broad application.

I have suggested, however, that Hilbert might identify mathematical truth with something like satisfaction in all models, as opposed to derivation. Would that kind of view be friendly to knowledge of mathematical truths that follow from the axioms? Quite plausibly: yes, it would. Since on Hilbert's view the axioms provide implicit definitions of the theoretical concepts, the axioms are conceptual truths. Logical deduction preserves conceptual truth. Accordingly, as Hilbert points out, whenever we have a system of things that satisfies the axioms of Euclidean geometry, that system will also satisfy the Pythagorean Theorem.[39]

Admittedly, it would not be the case on such a view that all truths can be known through logical deduction—at least not if the account were extended from geometry to arithmetic. After all, on Hilbert's view adding axioms changes the implicitly-defined concepts, and if some of those axioms are second-order, some truths will not be provable. But the fact that not all truths can be known through logical deduction is no longer a concern about truth; it is strictly an epistemological problem. And I do not see it as a major epistemological problem. We know many mathematical truths, but we have cognitive limitations that prevent us from knowing them all. And I suspect that unprovable truths are fairly distant cognitively. It may well be that we cannot know unprovable truths through derivation, but that this fact does not interfere with our knowing a great many mathematical truths through derivation.

Combinatorial accounts of mathematical truth—views like Hilbert's—seem to be amenable to plausible theories of mathematical knowledge. At least, they are amenable

[39]Cf. fn.30.

to theories of knowledge that fare well on some of the criteria we have identified. They fare well on (2); combinatorial accounts posit no inscrutable cognitive mechanisms in explaining our knowledge of the foundations of mathematical truth. If they do require inscrutable cognitive mechanisms to explain our knowledge of further mathematical truths, what remains is a problem for inferential knowledge, which is a much broader category than mathematical knowledge. They also fare well on (3); combinatorial accounts must respect mathematical reasoning, especially mathematical proof. And they fare well on (5); combinatorial accounts seem to cohere with plausible general theories of knowledge. Knowledge of the axioms is merely knowledge of stipulative definitions, and knowledge of non-axiomatic truths is generated through cognitive processess that preserve knowledge in non-mathematical cases too.[40]

Combinatorial accounts, which Benacerraf classifies as 'knowledge-motivated', are amenable to theories of mathematical knowledge that look promising in light of (2), (3), and (5). But those are not the only standards that we have seen for an adequate and satisfying account of mathematical knowledge. Such an account also must be compatible with a plausible theory of mathematical truth (4), and it must not require any unexplained antecedent knowledge or cognition of mathematics upon which all of our other mathematical knowledge rests (1). Benacerraf makes it clear that he thinks combinatorial views fare poorly on (4), but his reasons for thinking so seem to emerge from rather restrictive expectations for theories of mathematical truth. Given our considerations of Hilbert, I am inclined to think otherwise. At least one combinatorial account of mathematical truth—Hilbert's—seem to be compatible with a plausible theory of mathematical truth: mathematical truths are conceptual truths, and true in virtue of meaning.

Ultimately, though, I am not concerned here with whether or not combinatorial accounts yield plausible theories of mathematical truth. Regardless of how such accounts—

[40]Benacerraf wants accounts of mathematical knowledge to be causal. But he thinks knowledge-motivated views are combinatorial and not derivation-based. It is not clear to me how logical derivation is causal. It surely is not causal in the sense of physical causation. It likely is causal in a broader sense of causation, e.g. a grounding sense of causation. But while it is fairly clear that Gödelian intuition is not physically causal, it is not at all clear that Gödelian intution is not causal in the broader sense.

accounts that Benacerraf deems ‘knowledge-motivated’—fare as theories of mathematical truth, are they amenable to plausible theories of mathematical knowledge, even disregarding standard (4)? If so, they must also fare well on the first standard we raised, when considering the *Meno*. They must not presuppose any unexplained antecedent knowledge or cognition of mathematics. And, I propose, they also must fare well on a sixth standard:

(6) An adequate account of mathematical knowledge must be compatible with all of our mathematical knowledge, and with the understanding we have of mathematics.

This is a rather weak expectation. Ideally, an adequate account of mathematical knowledge would go further, and would serve to explain or ground all of our mathematical knowledge, and all of our understanding of mathematics.[41]

This expectation is rather different from the others, which are motivated by the discussions of mathematical knowledge in Plato and in Benacerraf. Both Benacerraf and Plato attempt to explain relatively simple examples of mathematical knowledge. Benacerraf uses the example of knowing that there are at least three perfect numbers greater than 17, and Plato uses the example of knowing that a square based on the diagonal of another square has twice the area of the square with that diagonal. But it is important that our account of mathematical knowledge be equipped to explain *all* of our mathematical knowledge, and not only the knowledge we have in these fairly simple cases. And it is important that our account of mathematical knowledge do this without relying on any unexplained antecedent knowledge or cognition of mathematics.

[41] Benacerraf makes an off-hand comment about accounts of mathematical truth that seems related to this standard that I am proposing. He writes, “the concept of mathematical truth . . . must fit into an overall account of knowledge in a way that makes it intelligible how we have the mathematical knowledge we have” (p.667). But he does not explore this consideration at all beyond this comment.

1.4 The Problem of Combinatorial Knowledge

We have seen that there is much to recommend combinatorial accounts of mathematical knowledge. They do well on several of our standards for accounts of mathematical knowledge. All that remains is to address how well they fare on standards (1) and (6)—relying on no antecedent mathematical knowledge or cognition, and being consistent with our having all of the mathematical knowledge and understanding that we actually have. Hilbert's example will help illuminate how combinatorial accounts fare with respect to these expectations.

1.4.1 Hilbert and Antecedent Knowledge

To some extent, Hilbert's account fares well on (1), the standard regarding antecedent knowledge. If a theory grounds our geometric knowledge in the stipulation of Hilbert's axioms as implicit definitions of geometrical concepts, then that theory does not rely on antecedent and unexplained knowledge of geometry. Hilbert's account does, however, rely on some antecedent mathematical knowledge: knowledge of the natural numbers.

As noted earlier, Hilbert only demonstrated the relative consistency of Euclidean geometry; he did not demonstrate absolute consistency. As long as arithmetic is consistent, his models of Euclidean and non-Euclidean geometries appear to be consistent, which suffices to show that the relevant systems of axioms are consistent. But Hilbert's reliance on the structure of the natural numbers is not limited to his *demonstration* of the consistency of arithmetic; it is not only an issue about Hilbert's proof. Rather, Hilbert's very axiomatization also depends on the consistency of arithmetic. One of his axioms is the Archimedean Axiom:

> Let A_1 be any point upon a straight line between the arbitrarily chosen points A and B. Take the points $A_2, A_3, A_4, \ldots$ so that A_1 lies between A and A_2, A_2 lies between A_1 and A_3, A_3 lies between A_2 and A_4, etc. Moreover, let the segments $AA_1, A_1A_2, A_2A_3, A_3A_4, \ldots$ be equal to one another. Then, among this series of points, there always exists a certain point A_n such that B lies between A and

A_n. (*The Foundations of Geometry*, §8)

The basic idea is that a finite number of shorter segments of a given length can always cover a line segment of a longer length. Hilbert relies on this axiom in order to demonstrate that his axioms fully describe a system of Euclidean geomentry. And whether or not a system of axioms satisfies this axiom depends on whether or not there is a suitable number n—an n such that the n^{th} length of a certain size (a size equal to that of the length between A and A_1) covers the point B. Whether or not such an n exists depends on the structure of the natural numbers.[42] But natural number is not otherwise mentioned in the axioms of Euclidean geometry; number is not implicitly defined by the axioms. Accordingly, the natural numbers—arithmetic—must be presupposed by the axioms of Euclidean geometry.

If Hilbert's account of semantics and truth in geometry is to be amenable to an account of mathematical knowledge that satisfies the demands of (1), it must avoid grounding geometrical concepts in any prior mathematical concepts. But his axiomatization of geometry presupposes the prior concept of natural number. Accordingly, Hilbert's account of geometrical truth only succeeds if the concept of natural number is antecedently fixed. And if Hilbert's account is to be extended to mathematical truth generally, as he seems to hope, it must be possible to characterize the system of natural numbers by implicit definition. In fact, this is possible; the details of axiomatizations of arithmetic will be presented in the next chapter. As we will see, some systems of arithmetical axioms are sufficient to characterize the structure of the natural numbers fully and uniquely.

In order for Hilbert's account of mathematical truth to ground a plausible account of mathematical knowledge, then, a theory of mathematical knowledge that grounds all of our knowledge of arithmetic and the natural numbers in axiomatic implicit definitions must satisfy the demands of both (1) and (6). That is, we must be able to acquire all of our knowledge of the natural numbers from arithmetical axioms. Furthermore, we must

[42]If we cannot distinguish the structure of the natural numbers from nonstandard arithmetical structures, then this is not a meaningful axiom. We will revisit this concern briefly at the end of Chapter 2.

be able to do this without relying on any antecedent arithmetical knowledge. If these are not possible, then a Hilbert-style view cannot provide an adequate account of mathematical knowledge; it cannot be that all mathematical truth, and hence all mathematical knowledge, is grounded in the setting-out of axioms.

1.4.2 What Follows: An Overview

In the following chapters of this dissertation, I will argue that Hilbert-style views do not provide adequate accounts of mathematical knowledge. Ultimately, not all of our mathematical knowledge can be grounded in definitions, stipulations, or postulations.

In Chapter 2, I will argue that it cannot be the case that all of our arithmetical knowledge is grounded in axiomatic definitions. In addition to knowing basic arithmetical truths like '$2 + 2 = 4$', we have a fairly robust understanding of the structure of the system of natural numbers. We routinely exhibit this understanding, for example, in our ability to understand and use the concepts 'finite' and 'infinite', which are interdefinable with the natural numbers. In their weaker, first-order forms, standard axiomatizations of arithmetic (most notably Peano Arithmetic) cannot characterize the structure of the natural numbers as well as we understand it. Although stronger, second-order axioms (interpreted robustly) can uniquely characterize this structure, I argue that we can only identify the structure those axioms characterize if we have a robust antecedent understanding of that structure.

In short, an account of mathematical knowledge based on Hilbert's account of mathematical truth can only satisfy (1) by violating (6). Such an account can only avoid relying on antecedent arithmetical understanding if it sacrifices some of our existing arithmetical knowledge. And such an account can only explain all of our mathematical knowledge if it presupposes that we have a substantial antecedent understanding of arithmetic—so substantial that the role of axiomatic definition in such an account is rendered entirely superfluous. An account like Hilbert's that grounds all of our mathematical knowledge in the setting-out of axioms appears to qualify as combinatorial, and hence knowledge-motivated, according to Benacerraf's taxonomy. But regardless of how well this type of view can account for truth, such an view cannot adequately account for mathematical knowledge, at least not arithmetical knowledge.

In Chapter 3, I will address a different type of view that attempts to ground all of our arithmetical knowledge in a stipulated implicit definition. This view, presented and defended by Bob Hale and Crispin Wright, is motivated by Benacerraf's concerns about providing a philosophy of mathematics that sacrifices neither a plausible theory of truth nor a plausible theory of knowledge. Hale and Wright take on the platonist challenge. They assume that numbers are abstract objects, and they attempt to explain one way in which we could come to have knowledge of these objects despite lacking any kind of causal interaction with them.

Hale and Wright respond to the platonist's challenge by attributing all of our arithmetical knowledge—our knowledge of abstract number-objects—to our knowledge of Hume's Principle: "The number of F's equals the number of G's if and only if the F's and G's are equinumerous." They take Hume's Principle to be a stipulated implicit definition that gives identity conditions for natural numbers. Using second-order logic, it is possible to derive a version of the Dedekind-Peano axioms for arithmetic from Hume's Principle, so we would then be able to derive all of our arithmetical knowledge from our knowledge of this stipulated implicit definition.

Even setting aside the concerns that I raise in Chapter 2 about deriving *all* of our arithmetical knowledge from axioms, I will argue that our arithmetical knowledge is not grounded in the way that Hale and Wright suggest, for three reasons. First, ancient mathematicians knew some of the same arithmetical truths that we know, and they did not have the systematized logic required to derive arithmetical axioms from Hume's Principle. Since we learned arithmetical truths from them, the grounds of their knowledge must also be the grounds of our knowledge.

Second, Hume's Principle gives identity conditions for *cardinal* numbers (singleton, pair, triple, etc.), and not for *ordinal* number (first, second, third, etc.). While it is possible to derive ordinal number from cardinal number, and vice versa, the claim that we derive all of our knowledge of arithmetic from Hume's Principle entails that cardinal number is cognitively prior to ordinal number. This claim about cognitive priority is substantive and empirical, and I believe it to be false.

And third, I will argue that Hume's Principle does not give us adequate means with which to pick out the *right* numbers. While it does ensure that the definite descriptions 'the number of fingers on my left hand' and 'the number of fingers on my right hand' pick out the *same* object, it does not uniquely determine *which* object that description picks out. Accordingly, Hume's Principle does not provide us with the uniquely identifying information required to develop knowledge of causally inert number-objects.

In Chapter 4, the last chapter of this dissertation, I will examine a standard approach to our knowledge of basic arithmetical truths like $2 + 2 = 4$. In addition to some general laws, Leibniz and Frege use the definitions '$2 = 1 + 1$', '$3 = 2 + 1$', and '$4 = 3 + 1$' in order to prove that $2 + 2 = 4$. If our basic arithmetical knowledge is grounded in proof, then our basic arithmetical knowledge is grounded in these definitions, and in the proofs that rely on them. Such proofs are those along the lines of the ones that we find in Leibniz and Frege.

My concern in examining this approach will be to determine whether or not it suffices to account for some of our basic arithmetical knowledge without relying on the kind of robust understanding of the structure of the natural numbers that I argue a Hilbert-style axiomatic approach must presuppose. In other words, how much antecedent knowledge or cognition of arithmetic is required for us to use our standard methods to come to know these basic and seemingly easy arithmetical truths?

In order to address this question, I will examine the inner workings of the proof that $2 + 2 = 4$. I will argue that the explicit definitions upon which it relies must be definitions of numerals, i.e. number expressions ('2', '4', etc.), rather than definitions of numbers. Accordingly, we can only understand the relevant proofs—and hence we can only use them to gain arithmetical knowledge—if we are positioned to understand expressions of the form '$n+1$', where 'n' is a numeral. I will then argue that, unless we already have some grasp of the structure of natural numbers, we cannot understand expressions of this form, and hence are not positioned to prove even simple arithmetical truths.

The central lesson to take from the remaining chapters of this dissertation is fairly simple. Accounting for our understanding of the structure of the natural numbers should be one of the key desiderata of a theory of mathematical knowledge. Without understanding this structure, we cannot come to know arithmetical truths like $2+2=4$—truths that we typically identify as some of our most secure knowledge. Definitions and postulations alone cannot ground our understanding of the structure of the natural numbers. Mathematics and mathematical knowledge, I posit, are more substantive than that. In order to account for mathematical knowledge, we must account for arithmetical knowledge. And in order to account for arithmetical knowledge, we must provide an account on which our understanding of the structure of the natural numbers is some of our most fundamental and primitive mathematical cognition.

Chapter 2

Axiomatic Definitions of $\mathbb{N}$

In this chapter, I argue that we could not come to understand the structure of the natural numbers ($\mathbb{N}$) if all of our mathematical knowledge were grounded in arithmetical axioms. Though the subject matter is somewhat technical, the main lessons to take from my argument are not. First, any adequate theory of arithmetical knowledge must be compatible with our fairly robust understanding of the structure of $\mathbb{N}$. Accordingly, it is not (and could not be) the case that our epistemic access to the natural numbers is entirely grounded in axiomatic definitions. Second, and relatedly, there are limits to the power of axiomatic definitions. And third, the more we think about how we might come to understand the structure of $\mathbb{N}$, the more evident it becomes that our understanding of this structure is in some way basic or foundational.

The body of this chapter begins with a presentation of the relevant mathematical framework. Given the subject matter, parts of this mathematical discussion may be unavoidably technical for readers without the relevant background. I have taken some effort to make these sections accessible to such readers. On the other hand, readers with the relevant mathematical background might find these sections familiar; such readers might prefer to skim the chapter until §2 (or §2.3), where the original philosophical work begins.

The first section begins with an introduction to the basics of Peano Arithmetic (PA), the canonical axiomatization of arithmetic. After some discussion of models of PA, I introduce two different languages in which PA can be expressed: the language of *first-order* arithmetic and the language of *second-order* arithmetic. Because the first-order language lacks a certain kind of expressive power, the axioms of first-order PA cannot distinguish $\mathbb{N}$ from recognizably different kinds of mathematical structures. So, I conclude, first-order PA cannot lead us to understand the structure of $\mathbb{N}$.

The second section addresses strategies we might use to come to understand the structure of $\mathbb{N}$ using second-order axioms, which are demonstrably robust enough to uniquely characterize that structure. The bulk of the philosophical work in this chapter occurs in addressing these second-order strategies. I argue that in order to use the axioms of second-order PA to come to understand the $\mathbb{N}$ structure, we must appeal, explicity or implicitly, to an antecedent understanding of this structure.

The chapter ends with a brief discussion of a potential skeptical objection. In response to my arguments, some might be tempted to deny that we have the kind of robust understanding of the natural numbers that my arguments demand. But such an objection is radical, and has significant and deeply implausible implications for the interpretation of mathematics. Rather than denying that we understand the structure of $\mathbb{N}$, we must admit that some of our mathematical knowledge is grounded elsewhere, and not in the axioms.

2.1 First-Order Peano Arithmetic

The Dedekind-Peano axioms, also known as the axioms of Peano Arithmetic (PA), provide the canonical axiomatization of $\mathbb{N}$. If an axiomatic definition is to yield the structure of $\mathbb{N}$, these axioms are the obvious candidates.

In the axioms of PA, we shall treat parentheses, (first-order) quantifiers and variables,[1]

[1]First-order quantification is quantification over individuals, e.g. individual numbers. It is the topic of further discussion in subsequent sections.

and the '=' symbol as logical; they do not require axiomatic definition.[2] Other symbols used in these axioms are not logical. Their meanings must be fully determined by the system of axioms. The first four axioms include two non-logical symbols. One is the '0' symbol, which is a name for an element; the other is the '$S(n)$' symbol representing a function of one argument. We call this 'the successor function'.

Without further ado, here are the first four axioms of PA:

(1) $0 \in \mathbb{N}$ (Zero is a natural number.)

(2) $(\forall n)S(n) \neq 0$ (Zero is not the successor of any natural number.)

(3) $(\forall n)(\exists m)S(n) = m$ (Every number has a successor.)

(4) $(\forall n)(\forall m)S(n) = S(m) \rightarrow n = m$ (No two distinct numbers have the same successor.)

These four axioms guarantee an infinite, non-repeating progression of successor elements that starts with a *ground element*—a number that is not the successor of any other number.[3]

On their own, axioms (1)-(4) are insufficient to describe an order on an arithmetical structure. No sentence limited to the language of axioms (1)-(4) can state, for arbitrary m and n, that m comes before n in the sequence. Sentences are finite, and the only way to capture this coming-before relation using the successor function is to combine an infinite number of specific claims, e.g.: '$n = S(m)$ or $n = S(S(m))$ or $n = S(S(S(m)))$, or $\ldots$'.

Further axioms of PA ensure that there is such an ordering. For the sake of simplicity, we will restrict ourselves to a modified version of PA that yields such an ordering, but only uses one more non-logical symbol: '$m < n$'.[4] We call this the 'less-than' or 'predecessor' relation ('m is less than n' or 'm is a predecessor of n'):

[2]Some variations of PA do not treat the '=' symbol this way, and include four axioms to govern it. The relation the '=' symbol picks out is (i) reflexive ($(\forall n)n = n$), (ii) symmetric ($(\forall m)(\forall n)m = n \rightarrow n = m$), (iii) transitive ($(\forall l)(\forall m)(\forall n)((l = m \wedge m = n) \rightarrow l = n)$), and (iv) $\mathbb{N}$ is closed under this relation ($(\forall n)(\forall m)n \in \mathbb{N} \wedge n = m \rightarrow m \in \mathbb{N}$). To use Dedekind's words, "That a and b are symbols for one and the same thing is indicated by the notation $a = b$." ("The Nature and Meaning of Numbers" pp. 44.)

[3]Nothing of import rides on the decision to take 0 instead of 1 to be the first natural number.

[4]The arguments I give will apply equally well to other versions PA, including the standard version that axiomatically defines addition and multiplication, and uses addition to explicitly define the less-than relation.

(5) $(\forall n)(0 < n \vee 0 = n)$ (Zero is less than every other number.)

(6) $(\forall n)n < S(n)$ (Each number is less than its successor.)

(7) $(\forall n_1)(\forall n_2)(\forall n_3)\big((n_1 < n_2 \wedge n_2 < n_3) \rightarrow n_1 < n_3\big)$ (Less-than is a transitive relation.)

(8) $(\forall n_1)(\forall n_2)(n_1 < n_2 \rightarrow n_2 \not< n_1)$ (Less-than is an anti-symmetric relation.)

(9) $(\forall n)(\forall m)(n = m \vee n < m \vee m < n)$ (If two numbers are not equal, then one of them is less than the other.)

The less-than relation, then, totally orders the natural numbers. Intuitively, we might say, for arbitrary m and n, m is less than n when m comes before n in the progression of successors.

Axioms (1)-(9), then, ensure that PA characterizes an infinite, non-repeating progression of successors that has a ground element and is ordered by the predecessor relation.

2.1.1 Uniqueness and Isomorphism

If there is a particular model $\mathbb{N}$ that satisfies axioms (1)-(9), the axioms do not uniquely describe it. Ordinarily, we think of $\mathbb{N}$ this way:

$$0,\ 1,\ 2,\ 3,\ \ldots,\ n,\ \ldots$$

But (1)-(9) equally well describe the progression of natural numbers starting with 1, or starting with 2, or starting with 3, etc. Of course, these progressions have different ground elements; the name '0' in the axioms applies in these progressions to 1 or 2 or 3 (etc.) instead of the standard 0.

Axioms (1)-(9) also describe progressions in which the implicitly-defined '$S(n)$' symbol picks out successor functions other than one it picks out in $\mathbb{N}$. Such progressions include that of the even numbers, that of the multiples of three, etc. The interpretations of '$S(n)$' in these cases (e.g. 'next even number') are consistent with the restrictions imposed by (1)-(4); the relevant progressions have ground elements, each of their elements has a successor, and no two distinct elements have the same successor. They also are ordered by predecessor relations, defined by (5)-(9), that coincide with the ordering on $\mathbb{N}$. All of these progressions will satisfy (1)-(9).

While the progressions just described do not have all the same elements as $\mathbb{N}$, they are just like $\mathbb{N}$ structurally. More precisely, these progressions are all *isomorphic* to $\mathbb{N}$. Two models are isomorphic if there is a mapping (an *isomorphism*) from one model to the other that preserves structural features. An isomorphism has two key features. (i) It is a bijection. That is, distinct elements of the first model get mapped to distinct elements of the second model, and each element of the second model is the image of some element of the first model.[5] (ii) It is a homomorphism. That is, structurally relevant functions and relations of the first model are preserved by the mapping. Since an isomorphism is a bijection, the image of a structurally relevant function or relation in the first model is a function or relation with the same structural role in the second model. In models of PA, the structurally relevant relations and functions are the ones defined using these axioms: successor and predecessor.[6] An isomorphism between two models of these axioms, then, maps pairs of successors in the first model to pairs of successors in the second model. It also maps elements related by predecessor in the first model to elements related by predecessor in the second model.

Consider two structurally alike models: $\mathbb{N}$ and the progression of even numbers. Intuitively, the successor function on $\mathbb{N}$ is 'next natural number', and the successor function on the evens is 'next even number'. Let f be a mapping that takes a number n to the number $2n$. f is a bijection from $\mathbb{N}$ to the evens; each even number is in the image of the mapping, and no two distinct natural numbers map to the same even number. Also, f preserves successor; if m is the successor of n in $\mathbb{N}$, then $f(m)$ (i.e. $2m$, or $2(n+1)$, or $2n+2$) is the successor of $f(n)$ (i.e. $2n$) in the sequence of evens. f preserves the progression ordering too. If $m < n$ on the first model, then $f(m) < f(n)$ (i.e. $2m < 2n$) on the second.

[5]The image of an element (or set of elements) under a mapping is the element (or set of elements) to which the original element (or set of elements) gets mapped. E.g. if $f(a) = b$, then b is the image of a under f.

[6]In other versions of PA, addition and multiplication are also defined by axioms and must also be preserved by isomorphisms.

So, f is both a bijection and a homomorphism; successor and predecessor on $\mathbb{N}$ correspond under the mapping with successor and predecessor on the evens. So, f is an isomorphism. Hence, as claimed before, $\mathbb{N}$ and the progression of even numbers are isomorphic—these two models are structurally the same.

Unless axioms describing $\mathbb{N}$ identify a particular ground element and a particular successor function, they cannot distinguish $\mathbb{N}$ from models isomorphic to it—models like the progression of even numbers. But the axioms of PA can only provide us with our grasp of $\mathbb{N}$ if they are purely structural axioms, and presuppose no prior familiarity with $\mathbb{N}$, its ground element, or its successor function. The axioms of PA cannot distinguish $\mathbb{N}$ from models isomorphic to it.

Although the axioms of PA cannot uniquely characterize any one particular model, perhaps they nonetheless can allow us to understand the structure-type of $\mathbb{N}$. Call a system of axioms *categorical* and say that the axioms yield *categoricity* if all of the models of those axioms are isomorphic to one another. If the axioms of PA are to ground our structural grasp of $\mathbb{N}$, they must at least be categorical. The question remaining, then, is this: Are the axioms of PA categorical? Do they uniquely characterize the structure-type of $\mathbb{N}$?

2.1.2 Standard and Nonstandard Models

Axioms (1)-(9) are not categorical. Imagine, for example, a model composed of a full copy of the natural numbers followed by a full (and fully distinct) copy of the integers. Call this model $\mathbb{N} + \mathbb{Z}$. Intuitively, $\mathbb{N} + \mathbb{Z}$ looks like this:

$$0', 1', 2', 3', \ldots, n', \ldots$$

$$\ldots, -n, \ldots \text{-2}, \text{-1}, 0, 1, 2, \ldots, n, \ldots$$

Although this model includes an additional integer-chain ($\mathbb{Z}$-chain) of elements, it satisfies axioms (1)-(4). It too has a ground element, every element has a successor, and no two distinct elements have the same successor. As long as we supplement the expected predecessor orderings on the $\mathbb{N}$-chain and $\mathbb{Z}$-chains by setting $m < n$ for any elements m of the

$\mathbb{N}$-chain and n of the $\mathbb{Z}$-chain, this model also satisfies axioms (5)-(9). Similarly, models with additional $\mathbb{Z}$-chains, such as $\mathbb{N} + \mathbb{Z} + \mathbb{Z}$ and $\mathbb{N} + \mathbb{Z} + \mathbb{Z} + \mathbb{Z}$, will also satisfy (1)-(9), as long as the different chains of elements are ordered with respect to one another by the predecessor relation. Models with many more $\mathbb{Z}$ chains satisfy the axioms too. So does a model with the form of $\mathbb{N} + \mathbb{Q}_{\mathbb{Z}}$, composed of an $\mathbb{N}$-chain followed by infinitely many densely-ordered $\mathbb{Z}$-chains, one for each rational number in the open interval between 0 and 1. Similarly, models of $\mathbb{N} + \mathbb{R}_{\mathbb{Z}}$, which have a $\mathbb{Z}$-chain for each real number, also satisfy (1)-(9).

But these models with $\mathbb{Z}$-chains are structurally different from $\mathbb{N}$; the models are not isomorphic. Take $\mathbb{N} + \mathbb{Z}$ as an example. Suppose a mapping f from $\mathbb{N}$ to $\mathbb{N} + \mathbb{Z}$ is a homomorphism; it preserves structural relations. Then f maps ground elements of $\mathbb{N}$ to ground elements of $\mathbb{N} + \mathbb{Z}$ (in each case, there is only one). Also, f preserves the successor relation amongst other elements. But the successor of any element of an $\mathbb{N}$-chain must also be in that $\mathbb{N}$-chain. So, the image of any element of $\mathbb{N}$ under f must be in the initial $\mathbb{N}$-chain of $\mathbb{N} + \mathbb{Z}$; the mapping does not reach any elements of the integer chain structure within the $\mathbb{N} + \mathbb{Z}$ model. Since some elements of $\mathbb{N} + \mathbb{Z}$ are not in the image of f, f is not a bijection. No function from $\mathbb{N}$ to $\mathbb{N} + \mathbb{Z}$ can both be a bijection and a homomorphism (preserving of structural relations). Hence $\mathbb{N}$ and $\mathbb{N} + \mathbb{Z}$ are not isomorphic models.

We now have multiple non-isomorphic models of (1)-(9). For now, let us assume that $\mathbb{N}$ has some particular type of structure, and that there is a fact of the matter about whether or not an arithmetical structure $\mathbb{M}$ is isomorphic to $\mathbb{N}$. We call a structure that satisfies PA a *standard* model of arithmetic if it is isomorphic to $\mathbb{N}$; we call it a *nonstandard* model of arithmetic if it is not. For example, the progression of even numbers is a standard model, while $\mathbb{N} + \mathbb{Q}_{\mathbb{Z}}$ is a nonstandard model. Our assumption, then, is that there is a legitimate distinction between standard and nonstandard models of arithmetic. This assumption, though rarely made explicit, is common in both philosophy and mathematics.

Let us assume further, for now, that we understand the progression of natural numbers well enough to distinguish standard models of arithmetic from nonstandard ones. That is, we understand the overall structure of the natural numbers. We can recognize, for example, that $\mathbb{N} + \mathbb{Q}_{\mathbb{Z}}$ is not a standard model of arithmetic. Let us make no assumptions about *how* we are able to distinguish standard models from nonstandard ones; let us only suppose that we are able to do so.

It is clear, then, that (1)-(9) do not describe the structure of $\mathbb{N}$ as well as we understand it. These axioms are not categorical. If PA is categorical, and if PA thereby gives us a means through which to grasp the structure of $\mathbb{N}$, then PA must exclude structures like $\mathbb{N} + \mathbb{Q}_{\mathbb{Z}}$ and $\mathbb{N} + \mathbb{Z}$. PA can only generate our understanding of $\mathbb{N}$ if some further feature(s) of PA can successfully exclude the extraneous chains of elements found in nonstandard models. Mathematical induction is all that remains of PA to play such a role.

2.1.3 Mathematical Induction

Mathematical induction uses the properties of elements to introduce further restrictions on what qualify as arithmetical structures. It guarantees that all of the natural numbers have some specified property whenever two conditions are met: (i) 0 has the specified property, and (ii) for each number n, if n has the specified property, then so does $n + 1$. To put this symbolically, mathematical induction is as follows:

(**MI**) For a property P:

$$\Big(P(0) \wedge (\forall n)\big(P(n) \rightarrow P(S(n))\big)\Big) \rightarrow (\forall n)P(n)$$

If there is a property that is unique to the standard natural numbers, then mathematical induction ensures that there are no nonstandard natural numbers, thereby ensuring the categoricity of PA.[7]

[7]Suppose there is a property P^* that every element of $\mathbb{N}$ has, but that none of the nonstandard elements of a model of arithmetic has—i.e. $P^*(n) \leftrightarrow n \in \mathbb{N}$ for n in a model of PA. If such a property P^* exists, the ground element has that property ((i) $P^*(0)$), since all of the elements of the initial $\mathbb{N}$-chain have it. Furthermore, every number that has the property is in the initial $\mathbb{N}$-chain. Since no particular nonstandard element of the model is the immediate successor of any element of the initial $\mathbb{N}$-chain, and since the immediate successor of any number in the initial $\mathbb{N}$-chain is also in the $\mathbb{N}$-chain ($n \in \mathbb{N} \leftrightarrow S(n) \in \mathbb{N}$), such a property also must be preserved by the successor operation ((ii) $P^*(n) \rightarrow P^*(S(n))$). Applying mathematical induction on the property P^* then ensures that all natural numbers have the property ($(\forall n)P^*(n)$).

The key question, then, is whether or not there is a property that successfully distinguishes the elements of the initial $\mathbb{N}$-chain from any putative nonstandard numbers. If so, mathematical induction excludes those putative nonstandard elements and yields categoricity. This would leave open the possibility that our understanding of the structure of the natural numbers is grounded in the axioms of PA.

The statement of mathematical induction in (**MI**), however, is imprecise. It does not specify standards governing what it is to be an admissible property. In contemporary mathematics, there are two main ways to delineate admissible properties. These two sets of standards for admissible properties correspond with the two versions of Peano Arithmetic: first-order PA (PA^1), and second-order PA (PA^2).

2.1.4 First-Order Induction

In PA^1, eligible properties are restricted to those that can be expressed as open sentences written in the first-order language of arithmetic. First-order languages are limited; in addition to logical connectives, our arithmetical version includes identity, 0 (zero), successor ($S(n)$), predecessor ($m < n$), and quantification over numbers ($\forall n$, $\exists n$). This first-order language is important because we ought to be able to fully understand it; axioms (1)-(9) have defined every non-logical symbol in the language.

Consider the property of either being 0 or being the successor of some number. This property is expressible in the first-order language of arithmetic:

$$\phi(n) = (def): \ n = 0 \vee (\exists m) n = S(m).$$

Since it can be expressed in the limited first-order language, this property ϕ is admissible as a property in mathematical induction. As it happens, ϕ also satisfies the antecedent of induction; it is easy to see that $\phi(0)$ holds, and that if $\phi(n)$ holds, then so does $\phi(S(n))$. Applying mathematical induction, then, yields that $(\forall n)\phi(n)$ (i.e. every number is either zero or the successor of some number). So, first-order induction guarantees that the ground element is unique, i.e. there is only one zero in an arithmetical structure.

Clearly, first-order mathematical induction adds nontrivial restrictions to the ones imposed by axioms (1)-(9).[8] But the highly accessible language of first-order arithmetic has significant limitations. In particular, it does not allow quantification over subsets of $\mathbb{N}$, relations, or properties. This limitation has at least two noteworthy consequences. First, first-order mathematical induction cannot be expressed as a single statement. Instead, it becomes an axiom schema into which first-order sentences can be substituted:

(**MI–Schema**) Let $\phi(n)$ be an arithmetical sentence in a first-order language. Then $\Big(\phi(0) \wedge (\forall n)\big(\phi(n) \rightarrow \phi(S(n))\big)\Big) \rightarrow (\forall n)\phi(n)$

Second, some properties that seem potentially useful for our present purposes cannot be expresssed in the first-order language. The property 'finite' cannot be taken as primitive in the first-order language, so we cannot use a property like 'n has finitely many predecessors' in mathematical induction to yield categoricity. We also cannot express anything that involves quantification over sets of numbers, properties of those sets, or correspondences or relations between those sets. The only quantification in first-order arithmetic is quantification over individual elements of arithmetical structures.

These limitations of first-order languages, including that of arithmetic, entail that PA^1 ((1)-(9) plus the first-order schema of induction) cannot uniquely describe the structure of the natural numbers. PA^1 cannot distinguish $\mathbb{N}$ from $\mathbb{N} + \mathbb{Q}_{\mathbb{Z}}$. PA^1 is not categorical. This is a well-known result.[9] Since PA^1 is unable to characterize the structure of $\mathbb{N}$ up to isomorphism, it cannot ground our understanding of the structure of the natural numbers.

[8] In versions of PA that include addition and multiplication, first-order induction ensures that $\mathbb{N} + \mathbb{Z}$ is not a model of PA^1. However, there will be many different isomorphism types structured like $\mathbb{N} + \mathbb{Q}_{\mathbb{Z}}$ and $\mathbb{N} + \mathbb{R}_{\mathbb{Z}}$ (differing in their addition/multiplication operations) that satisfy PA^1.

[9] Appendix I explains this result.

2.2 Second-Order Peano Arithmetic

First-order languages only license quantification over individuals in the relevant domain. Second-order languages are more expressive in one of two ways. One way a language can be second-order is by licensing quantification over all properties and relations (with determinate extensions) on the domain. The second way a language can be second-order is by licensing quantification over all sets of individuals in the domain. Ultimately, though, quantification over properties and quantification over sets of numbers are equivalent in strength. Each determinate property has a determinate extension that can be treated as a set, and each set corresponds with the extension of the property of being a member of that set.

The expanded second-order language of arithmetic licenses quantification over all sets, or all properties, of numbers. This allows us to express, for example, the *least number principle*: every set of numbers has a least number (a number that is the predecessor of every other number in the set). Although claims like the least number principle are common in ordinary mathematics, they cannot be stated in first-order arithmetic.

PA^1 and PA^2 ((1)-(9) plus second-order mathematical induction) differ only in mathematical induction. In keeping with the two ways in which a language can be second-order, there are two ways to state second-order mathematical induction. The first way takes the second-order quantifier to range over properties of the elements of a model of arithmetic:

$$\textbf{(MI–Properties):}\ (\forall P)\Big(\big(P(0) \wedge (\forall n)\big(P(n) \rightarrow P(S(n))\big)\big) \rightarrow (\forall n)P(n)\Big)$$

That is, for any property P, if (i) P is a property of 0, and if (ii) P is preserved by successor (i.e. for all numbers n, $P(n) \rightarrow P(S(n))$), then P is a property of every number ($(\forall n)P(n)$). In other words, if (i) and (ii) hold for a property, then elements that lack the property are not elements of the model of second-order arithmetic.

The second way to express second-order induction is to take the second-order quantifier to range over sets of elements:

$$\textbf{(MI–Sets):}\ (\forall P)\Big(\big(0 \in P \wedge (\forall n)(n \in P \rightarrow S(n) \in P)\big) \rightarrow (\forall n) n \in P\Big)$$

That is, for any set of elements of a model, if (i) 0 is in the set, and (ii) the successor of any element of the set is also in the set, then every element of the model is in the set.

2.2.1 Second-Order Categoricity

If the axioms of PA^2 are to characterize the $\mathbb{N}$-structure well enough to ground our understanding of it, they must characterize a structure-type up to isomorphism. Provably, they do. Using second-order resources, we can demonstrate that any two models of PA^2 are isomorphic to each other. PA^2 is categorical.

Suppose $\mathbb{M}$ and $\mathbb{M}'$ are both models of PA^2. Then construct a mapping $f : \mathbb{M} \rightarrow \mathbb{M}'$ as follows:

$$f(0_{\mathbb{M}}) = 0_{\mathbb{M}'}$$

$$f(S_{\mathbb{M}}(n)) = S_{\mathbb{M}'}(f(n))$$

The mapping f clearly preserves the ground element and the successor relation. To show that f is an isomorphism, we must only show that it is a bijection and that it preserves the predecessor (less-than) relation. This is done in Appendix II. So, any two models of PA^2 are isomorphic to each other.[10]

Since PA^1 is provably not categorical, this categoricity proof requires second-order resources. The proof that f is a bijection proceeds by applying induction to two uncontroversially determinate sets defined using the mapping f.[11] Since f is a mapping between

[10] This proof of categoricity is similar to Zermelo's 1930 proof of the categoricity of second-order ZF (Zermelo-Fraenkel set theory).

[11] One is the set of elements of $\mathbb{M}'$ that are in the f-image of $\mathbb{M}$. The other is the set of all m in $\mathbb{M}$ such that, for all $m' <_{\mathbb{M}} m, f(m) \neq f(m')$. See Appendix II for verification.

models of PA^2, these sets are identified by taking another model as a reference. They are not defined by first-order sentences—sentences of the language of arithmetic with quantification only over elements of the relevant model. But nonetheless they are in the scope of the second-order quantifier of **MI–Sets**.

While this proof entails that an isomorphism can be built between any two models of PA^2, it does not identify the structure-type that PA^2 characterizes. It does not ensure that these axioms can ground our understanding of the $\mathbb{N}$-structure. To do that, the axioms must characterize the *right* structure-type—that of standard models, and not nonstandard ones.

And it is obvious to us that they do. It is obvious to us that any given model of PA^1 begins with an initial $\mathbb{N}$-chain containing the model's ground element and all of its finite successors. The elements of this $\mathbb{N}$-chain can be collected into a set, and there is a property of being in the initial $\mathbb{N}$-chain of the model. Call this property P'. Although P' cannot be expressed in the language of first-order arithmetic, it does seem to be a perfectly good, determinate property.

So, we can apply **MI–Properties**. The ground element of a given model of PA^2 is in its initial $\mathbb{N}$-chain ($P'(0)$). And if an element is in the initial $\mathbb{N}$-chain of a model, then so is its successor ($P'(n) \rightarrow P'(S(n))$). So, the the model satisfies both antecedents of mathematical induction for the property P'. Hence, if a model satisfies PA^2, the property must hold for all elements of the model ($(\forall n)P'(n)$). That is, if the model is a model of PA^2, then all of its elements are in its initial $\mathbb{N}$-chain. The model has the structure of $\mathbb{N}$. Given the categoricity result, then, every model of PA^2 has the structure of $\mathbb{N}$. In the analogous case of the set whose members include all and only the elements of a model's initial $\mathbb{N}$-chain, **MI–Sets** will yield the same result: PA^2 uniquely characterizes the structure of $\mathbb{N}$.

2.2.2 Potential Limitations of Second-Order Quantification

The result that PA^2 uniquely characterizes the structure of $\mathbb{N}$, however, depends on the range of the second-order quantifier. If being in the initial $\mathbb{N}$-chain of a given model is a genuine property, or if there is a set of the elements of the model's initial $\mathbb{N}$-chain, then the desired result follows. But if the elements of the initial $\mathbb{N}$-chain do not comprise a set, or if being in the initial $\mathbb{N}$-chain is not a genuine, determinate property, then the result may well fail. We normally assume that this is a genuine property, and that the standard elements do comprise a set. But there are nonstandard interpretations of the second-order quantifier on which the elements of the initial $\mathbb{N}$-chain of a model of arithmetic do not comprise a set—or, at least, they do not comprise a set within the range of the second-order quantifier.[12]

If PA^2 is to uniquely characterize the structure of $\mathbb{N}$, the second-order quantifier must have its standard meaning. Call this its *full semantics*. The quantifier must range over *all* of the sets of elements of the model in question (or *all* of the properties). Let us make a potentially controversial assumption: that there *is* some standard semantics for the second-order quantifier.[13] That is, let us assume that there is a matter of fact about what *all* the subsets of some given set are. And, however it might come to pass, let us assume that the second-order quantifier has its standard, full semantics. Somehow, the second-order quantifier comes to range over the right collection of sets. Let us also assume that under the standard, full second-order semantics, one of the subsets of a model of arithmetic is the set containing all and only the elements of the model's initial $\mathbb{N}$-chain. Given these fairly robust

[12] Such nonstandard interpretations of the second-order quantifier are due to Henkin. First, terminology: the *full powerset*, intuitively speaking, is the set of *all* the subsets of the elements of some given domain. In standard, full semantics, the second-order quantifier ranges over the full powerset of the given domain. In Henkin semantics, the second-order quantifier can range over any of a class of subsets of the full powerset of the given domain. To be in this class, a subset of the powerset must satisfy some minimal conditions, such as including all definable subsets. The full powerset itself would be one class of subsets that satisfies the relevant minimal conditions. But it is not the only such class.

[13] Part of the motivation behind Henkin semantics is the suspicion that there is no such standard semantics, or at least that we could not possibly understand standard semantics in the case of infinite domains—domains such as the elements of an arithmetical structure.

assumptions, PA^2 does uniquely characterize the structure of $\mathbb{N}$.[14]

But the fact that PA^2 uniquely characterizes a structure does not entail that we can use PA^2 to come to understand that structure. In particular, although the reasoning above demonstrates for us that every model of PA^2 has the structure of $\mathbb{N}$, this reasoning cannot help us come to understand the structure of $\mathbb{N}$. We cannot use a property (being in the initial $\mathbb{N}$-chain) to come to understand the structure of $\mathbb{N}$ if we must understand the structure of $\mathbb{N}$ in order to understand that property.

2.2.3 Understanding through Axioms

We return, then, to the question of understanding. Assuming that PA^2 uniquely describes the structure of $\mathbb{N}$, and that nonstandard models fail to satisfy second-order mathematical induction, how could PA^2 ground our knowledge of the structure of $\mathbb{N}$? How could we use PA^2 to come to understand that structure-type?

Before we identify the distinctively second-order resources required to come to understand the structure of $\mathbb{N}$, let us take stock of the other resources at our disposal. Assuming we have the basic ability to iterate, axioms (1)-(4) position us to make sense of a non-repeating progression of successors with a ground element but no final element. They provide us with the resources required to think through some initial segments of such a progression, and to understand that any given element, no matter how far down the progression of successors, is followed by another. Axioms (5)-(9) then position us to understand this progression as totally ordered by the predecessor relation.

First-order mathematical induction contributes more to our understanding. Together with a very accessible first-order sentence, it restricts us to models with only one ground element. But PA^1 does not uniquely characterize the structure of $\mathbb{N}$; it also characterizes nonstandard models of arithmetic. Accordingly, without further resources, the axioms of

[14] I am not entirely convinced that these assumptions are correct. If not, so much the worse for PA grounding our understanding of the structure of $\mathbb{N}$.

PA^1 cannot lead us to understand the structure of $\mathbb{N}$ as well as we do; they do not give us a thorough enough understanding to distinguish $\mathbb{N}$ from nonstandard models of arithmetic. PA^1 does not, so to speak, sufficiently limit the length of progressions of successors.

With limited second-order resources, as seen in §2.1, we can prove that PA^2 only characterizes one structure-type. The proof only requires induction on two sets that are not defined in the first-order language, both of which are expicitly defined using the mapping f, which turns out to be an isomorphism. And the mapping f is cognitively accessible. We are positioned to understand it as used in the proof as long as we understand that there can be mappings between mathematical structures, and also understand some basic structural features addressed previously, viz. the zero and successor functions defined in axioms (1)-(4). Since the proof is straightforward and uses fairly accessible second-order resources, we are able to appreciate the categoricity of PA^2 without much prior knowledge, and without antecedently understanding the structure of $\mathbb{N}$.

But without revealing the structure-type of any model that satisfies the axioms of PA^2, the categoricity proof cannot yield understanding of the structure of $\mathbb{N}$. Indeed, if we (wrongly) took $\mathbb{N} + \mathbb{Q}_{\mathbb{Z}}$ to be a model of second-order PA, the categoricity proof on its own would lead us to think (again wrongly) that all models of PA^2 have that structure. So, if PA^2 leads us to understand the structure of $\mathbb{N}$, then a more robust use of second-order resources must unambiguously illuminate that structure.

We have already identified one way second-order induction uniquely characterizes the structure of $\mathbb{N}$. We identified a property that satisfies the two conditions of mathematical induction—$P(0)$, and $P(n) \to P(S(n))$. We then observed that, while all of the elements of standard models of arithmetic have this property, not all of the elements of nonstandard models do. In other words, we identified a property that distinguishes standard models from nonstandard models, and that makes nonstandard models fail second-order mathematical induction. But the property was that of being in the initial $\mathbb{N}$-chain of the model. Obviously, we cannot use this property to come to understand the structure of $\mathbb{N}$; understanding the property depends upon already understanding this $\mathbb{N}$-structure.

Another property that allows us to uniquely characterize the structure of $\mathbb{N}$, fairly trivially, is the property of being a *standard* element, as opposed to a *nonstandard* one. It satisfies the two conditions in the antecedent of mathematical induction—0 is a standard element, and if n is a standard element, then so is its successor. By second-order induction, if the model satisfies PA^2, then all of its elements are standard ones. But, obviously, this property also will not serve our purposes. By definition, to be a standard element of a model of arithmetic is simply to be an element in the model's initial $\mathbb{N}$-chain. So, understanding this property also depends upon understanding the structure of $\mathbb{N}$.

There might, however, be other suitable properties. Second-order mathematical induction can ground our understanding of the $\mathbb{N}$-structure provided that there is some property—any property at all—that the ground element has, that successor preserves, and that we can understand without antecedently understanding the structure of $\mathbb{N}$. Ultimately, I shall argue that we must already understand the structure of $\mathbb{N}$ in order to use any such property to recognize that PA^2 uniquely characterizes the structure of $\mathbb{N}$.

2.3 To Distinguish Standard Models

As we have seen, one way to distinguish standard models of arithmetic from nonstandard ones is to explicitly envoke the structure of $\mathbb{N}$. That is how I introduced the distinction in this paper. But that way of distinguishing the different types of models relies on our already understanding the $\mathbb{N}$-structure. If the axioms of PA^2 are to ground our understanding of that structure, there must be some other way to identify standard models.

There are three standard approaches to distinguishing standard models of arithmetic from nonstandard ones without explicitly envoking $\mathbb{N}$. One approach relies on the notions of *finitude* and *infinity*; someone taking such an approach might, for example, claim that every standard element of an arithmetical structure is finite, and every nonstandard element is infinite. The second approach is to use the *least number principle* (briefly mentioned in §2), which only holds for standard models. And the third approach relies on the set-theoretic idea of taking the *intersection* of all models with the same zero element and successor function; $\mathbb{N}$ is the only arithmetical structure-type that can be found within every model of arithmetic.

We shall consider each of these approaches in turn, starting with the first. Can we ground our understanding of the structure of $\mathbb{N}$ in an application of mathematical induction to the property of finitude? It seems as though being finite is a distinguishing feature of standard elements of arithmetical structures. But, as it turns out, the term 'finite' can be applied to any of a family of properties. Ultimately, none of these properties can lead us to understand the structure of $\mathbb{N}$.

2.3.1 Traditional and Intuitive Kinds of Finitude and Infinity

We seem to have an intuitive understanding of the difference between finitude and infinity, and it is tempting to appeal to this intuitive difference to distinguish standard elements of arithmetical structures from nonstandard ones. But unless we apply this difference suitably, it will not distinguish standard models from nonstandard ones. Accordingly, it is important to identify two features of how we are using this intuitive difference. First, we must identify whether we are discussing finitude or infinity as intrinsic features of elements, or whether we are discussing finitude and infinity in virtue of the place elements have in the arithmetical structures in which they occur. And second, since there are distinct conceptions of infinity, we must ensure that the intuitive difference we are using is in keeping with a conception of finitude and infinity that successfully distinguishes standard elements from nonstandard ones.

It might be the case that something is (or some things are) intrinsically infinite—that something is infinite by its very nature. Perhaps this is the case with Anaximander's *apeiron*, or with the Abrahamic God, or with Descartes's plenum (comprising space and its contents).[15] Or perhaps certain large magnitudes—or large sets, pluralities, or collections—are intrinsically infinite. Whether or not this is the case, the intrinsic finitude or infinity of things will not be a property capable of distinguishing standard elements of arithmetical structures from nonstandard elements. Suppose there is a nonstandard model of arithmetic $\mathbb{M}$ in which all standard elements are intrinsically finite and all nonstandard elements are instrinsically infinite. Then there are models isomorphic to $\mathbb{M}$ in which some standard elements are infinite and some nonstandard ones are finite.[16]

Since it cannot generally be the case that standard elements of models of arithmetic are finite and nonstandard elements are infinite, other features must get us to focus on models that have intrinsically finite standard elements and intrinsically infinite nonstandard ones. There must be some other relevant feature about the way the model is structured that makes it the case that only standard elements are intrinsically finite. Two set-theoretic ways of structuring models to ensure the finitude of (only) the standard elements are fairly prominent in philosophy, thanks to Paul Benacerraf.[17] One is the model of von Neumann numbers, structured thus:

$$\emptyset;\ \{\emptyset\};\ \{\emptyset, \{\emptyset\}\};\ \ldots n;\ n \cup \{n\}\ \ldots$$

Another is the model of Zermelo numbers, structured thus:

$$\emptyset;\ \{\emptyset\};\ \{\{\emptyset\}\};\ \ldots n;\ \{n\}\ \ldots$$

The kind of intrinsic finitude found in these models is finite cardinality (finitely *many*), which we shall address in §3.2. There are finitely many members of any standard element on the von Neumann model, and there are finitely many pairs of brackets in any standard

---

[15]Descartes actually took the plenum to be *indefinite*, as opposed to *infinite*, because he was concerned about the contrast with God's infinity. But, although Descartes did not do this, one might take the plenum to be intrinsically infinite.

[16]Consider the model $\mathbb{M}'$ with all and only the same elements that $\mathbb{M}$ has, but in which the places of two elements are swapped. One of those elements, n, is in the initial $\mathbb{N}$-chain of $\mathbb{M}$, and the other element, m, is not in the initial $\mathbb{N}$-chain of $\mathbb{M}$. Hence, n is standard in $\mathbb{M}$ and nonstandard in $\mathbb{M}'$; likewise, m is nonstandard in $\mathbb{M}$ and standard in $\mathbb{M}'$. Since all of the standard elements of $\mathbb{M}$ are intrinsically finite, n is intrinsically finite, and so a nonstandard element of $\mathbb{M}'$ is intrinsically finite. Similarly, m is intrinsically infinite because it is nonstandard in $\mathbb{M}$, so $\mathbb{M}'$ has an intrinsically infinite standard element.

[17]See "What Numbers Could Not Be."

element on the Zermelo model. Other approaches of this type are similar.[18] Setting aside the details of these models, it seems that intrinsic finitude and infinity—in these models and others—can only help us to understand the structure of $\mathbb{N}$ if they are tied to more general structural features of the model. But then we are not using the elements' intrinsic finitude or infinity to understand the structure of $\mathbb{N}$. Rather, we are using those other structural features, to which intrinsic finitude and infinity are tied, to come to understand.

If finitude is to distinguish standard models of arithmetic from nonstandard ones, then the relevant kind of finitude cannot be an intrinsic feature of elements. Rather, it must be a feature that an element has in virtue of its place in an arithmetical structure. The intuitive thing to say is that an element is finite or infinite in virtue of being finitely or infinitely distant from zero. Presumably distance here is measured either using the elements between the zero element and the element in question, or the iterations of the successor function required to get from zero to the element in question.[19]

Given that the relevant kind of finitude or infinity has something to do with distance from zero, several traditional conceptions of finitude and infinity will be unable to distinguish standard elements from nonstandard ones. One idea commonly found in the history of philosophy is that of the infinite as all-encompassing, complete, and whole. According to this conception of infinity, the half-plane is finite because the other half of the plane is not included in it, and likewise the entire plane minus one point is finite because of the single point it lacks.[20] Under this conception, even nonstandard elements are finite. The distance

[18]For example, Frege's account is analogous to Zermelo's and von Neumann's, but makes it so that standard elements are given in terms of there being finitely many objects in the extension of a concept. See *The Foundations of Arithmetic*, §70-83.

[19]One might similarly say that a model is standard if any two elements of the model are finitely distant from each other. This claim entails that any element of the model is finitely distant from zero. Since we only need this latter property to ensure categoricity, and this latter property uses the same conceptual machinery as the former does to get a categoricity result, we shall focus on the distance from zero.

[20]Spinoza suggests this in Definition 2 of Part I of his *Ethics*: "A thing is said to be finite in its ofn kind when it can be limited by another thing of the same nature. For example, a body is said to be finite because we can always conceive of another body larger than it." Descartes also seems to present such an idea in the *Principles of Philosophy*(Part I, §26), but he is careful to call this 'indefinite' as opposed to 'infinite'. He thinks we ought not to reason about the properly infinite.

between zero and a given element could only include all of the elements of the model if there were some last element of an arithmetical structure. But every element n of an arithmetical structure, standard or nonstandard, has a successor that is not in the distance between 0 and n (inclusive). Hence, no element of an arithmetical structure is infinite in this sense. Accordingly, the relevant conceptions of finitude and infinity cannot be the ones on which the infinite is all-encompassing and the finite somehow incomplete or lacking.

Similarly, nonstandard elements will be finite under the common conception of the infinite as boundless, unlimited, or unending, and the finite as bounded, limited, or coming-to-an-end. There is a limit and a bound to the distance from zero to any given nonstandard natural number. This is evident from the fact that any given nonstandard element has a successor that is beyond the bound or limit of this distance. Since the conception of finitude as bounded or limited admits nonstandard elements as finite, and not as infinite, this kind of finitude and infinity cannot distinguish standard models from nonstandard ones, and so cannot help us come to understand the structure of $\mathbb{N}$.

Aristotle's definition of infinity is of no help either. According to him, "A quantity is infinite if it is such that we can always take a part outside what has already been taken."[21] If the distance between zero and nonstandard elements of arithmetical structures is such that we can always take a part outside what has already been taken, then nonstandard elements of arithmetical structures are infinite by Aristotle's definition. But since we are operating with second-order resources, we can take sets of elements of arithmetical structures as individuals.[22] So, for any n, we can take the set of n and its predecessors all at once; once we do that, there is no part of the distance between 0 and n that has not already been taken. Because we are using second-order resources that allow us to take large sets (and all

[21] *Physics* III.6, 207a8.

[22] Technically, we are only using properties of elements in this version of second-order induction, and not sets. We shall address sets in the next section. But Aristotle's definition is about parts of wholes, and we are not talking about intrinsic features of elements of arithmetical structures. Accordingly, discussion of parts of wholes here must be modeled as analogous to susets of sets. To more faithfully restrict ourselves to the property version of second-order PA, we should think of the relevant sets as extensions of properties.

of their elements) all at once, nonstandard elements are again finite in the sense of Aristotle's definition. We cannot use this conception of infinity to distinguish the standard elements of arithmetical structures from the nonstandard ones.

A fourth conception of the finite and infinite is that of the finite as traversable and the infinite as not.[23] I take this to be the most intuitive idea of finitude and infinity we have in mind when we take such properties to distinguish standard elements of arithmetical structures from nonstandard ones. The idea is approximately this: if you start at zero and go through the numbers towards some nonstandard number n, you will never get there—you will never reach n. Since it is not possible to traverse the distance from 0 to n, n is infinite.

But traversability alone will not suffice; we need some restrictions on what it is for a distance to be traversable. Given infinite time, we could reach a nonstandard number; we must restrict ourselves to finite time. Counting infinitely fast or with infinitely large intervals, we could reach a nonstandard element; we must restrict ourselves to finite counting speeds and finitely large counted intervals. The restrictions we place on what it is for a quantity to be traversable, then, end up presupposing a robust notion of finitude. In fact, they presuppose what seems to be the very notion of finitude that we are attempting to articulate.

Our idea of a distance that is traversable in the relevant sense is our idea of something that is accessible in finitely many steps—that is, accessible in a *finite number* of ordered steps. If we have the idea of something being traversable in the relevant sense, then we already have the idea of a standard element of an arithmetical sequence. Our intuitive understanding of what it means for a distance to be traversable depends on our understanding of the structure of $\mathbb{N}$. If we can traverse something, then there is some (standard) natural number of steps we can take to reach the end; if we cannot traverse something, then we can follow through an entire $\mathbb{N}$-chain worth of steps without reaching the end. Our idea of the finite as traversable presupposes our having the idea of finitely many ordered steps or iterations, which is just the same as having the idea of the structure of $\mathbb{N}$. Since understanding the relevant property

[23]This is found in Aristotle's *Metaphysics K* and *Physics* III, among other places.

of finitude requires understanding the structure of $\mathbb{N}$, we cannot use this traversability kind of finitude (together with mathematical induction) to come to understand the structure of $\mathbb{N}$. If we use this property, it is because we understand the structure of $\mathbb{N}$ already.

2.3.2 Technical Definitions of Finitude

Intuitive ideas of finitude and infinity cannot help us come to understand the structure of $\mathbb{N}$. But perhaps technical definitions of the finite and infinite can help; several have been given in mathematics. All of these are (or are derived from) definitions of finitude and infinity in *sets*. They are definitions of what it is for there to be finitely or infinitely many things—that is, what it is for a set to have finite or infinite *cardinality*. I will address the two most prominent definitions; others are similar.[24]

Most commonly, a set is called finite if it can be put into one-to-one correspondence with the natural numbers between zero and some natural number n. The presupposition here, of course, is that n must be a *standard* natural number. The circularity problem arises again. If we rely on an antecedent understanding of the structure of $\mathbb{N}$ to develop our understanding of the property of finitude, then it cannot be that we rely on an antecedent understanding of the property of finitude to come to understand the structure of $\mathbb{N}$.

A second technical definition of finitude and infinity in mathematics is attributed to Dedekind.[25] A set is *Dedekind infinite* if its elements can be put into one-to-one correspondence with the elements of one of its proper subsets; if they cannot, the set is *Dedekind finite*.[26] The predecessors of standard natural numbers comprise Dedekind finite sets, and

[24]At least, they are similar enough that my arguments can easily be adjusted to account for them. On one definition, for example, a set is infinite if there exists a relation on it such that (a) for every n, there is some m such that nRm; (b) the relation is antisymmetric ($mRn \to \neg mRn$); and (c) the relation is transitive ($nRm \wedge mRl \to nRl$). Hopefully, after reading my arguments against using Dedekind infinity to account for our understanding, it will be fairly clear why this definition would not work either.

[25]Dedekind gives the definition in §64 of *The Nature and Meaning of Numbers*. In the preface to the second edition of that work, he acknowledges that Cantor and Bolzano also might legitimately be credited with this definition, though they did not use it as a foundation for the theory of natural numbers, as Dedekind did.

[26]Technically, a set is Dedekind infinite if there is a bijection (one-to-one and onto mapping) between the set and one of its proper subsets. According to the Schroeder-Bernstein Theorem, if there is a one-to-one

the predecessors of nonstandard natural numbers comprise Dedekind infinite sets.[27] So, the property of Dedekind finitude distinguishes standard natural numbers from nonstandard ones. This fact opens up two strategies we might use to come to understand the structure of $\mathbb{N}$. We might use the property of Dedekind infinity to understand how to rule out non-standard models, which have elements whose predecessors comprise Dedekind infinite sets. Or, we might use the property of Dedekind finitude to build up a positive understanding of features peculiar to standard models.

Intuitively, the strategy of ruling out nonstandard models seems to get something wrong in the order of explanation. If we are to categorize something, we must know what we are categorizing. So, in order to categorize a model as nonstandard on the grounds that it has an element with Dedekind infinitely many predecessors, we must already understand the structure of the model we are so categorizing. But it seems like we come to understand nonstandard structures like $\mathbb{N} + \mathbb{Q}_{\mathbb{Z}}$ and $\mathbb{N} + \mathbb{R}_{\mathbb{Z}}$, for example, through our understanding of $\mathbb{N}$, and not vice-versa; after all, every nonstandard model of PA^1 begins with an $\mathbb{N}$ structure.

Furthermore, ruling out nonstandard models could not lead us to understand the structure of standard models. Categorizing structures as nonstandard will only help us to know what the structure of $\mathbb{N}$ is not, rather than what it is. While the process of elimination might help us identify an already-understood structure that is not eliminated in this way (i.e. that has no elements whose predecessors comprise Dedekind infinite sets), eliminating alternatives will not help us understand the structure of $\mathbb{N}$. We will only understand the structure of standard models after completing the process of elimination if we already understand that structure before beginning the process.

---

mapping from set A into set B, and a one-to-one mapping from set B into set A, then there is a bijection between sets A and B. Since there is clearly a one-to-one mapping from any subset into the set of which it is a subset (the identity mapping), so long as there is a one-to-one mapping from the set into its subset, there will be a bijection between the two.

[27]Let n be a nonstandard natural number. Let P_n be the property of being a predecessor of n (in the relevant model). Let P_n' be the property of both being a predecessor of n and not being zero. Then for all of the elements m in the initial $\mathbb{N}$-chain of the model (which all satisfy P_n), we can map m to $S(m)$. For all of the elements m' that satisfy P_n but are not in the initial $\mathbb{N}$-chain, we can map m' to itself. Since this is a one-to-one mapping, the set determined by P_n is Dedekind infinite.

Perhaps, though, we use the property of Dedekind infinity to come to understand the structure of $\mathbb{N}$ without relying on the process of elimination. Rather, using the fact that every element of a standard model has a particular feature (viz. that its predecessors comprise a Dedekind finite set), we might build up a positive understanding of the structure of standard models. To do this, we must make sense of the feature in a way that gives us insight into how elements with that feature are situated in models of arithmetic.

In order to come to understand the structure of $\mathbb{N}$ through the property of Dedekind finitude, we must consider the mappings (and/or types of mappings) that Dedekind finite sets of predecessors can and cannot support. In particular, the set of predecessors of an element of a standard model of arithmetic cannot support one-to-one mappings into any of its subsets. So, every one-to-one mapping from such a set into itself is also an onto mapping; every element of the set is in the image of every one-to-one mapping from the set into itself. But we cannot survey all of the one-to-one mappings from the set of predecessors of an arbitrary element into itself, since we are cognitively incapable of surveying all of the elements of a sufficiently large set, finite or infinite. There are many more one-to-one mappings from a sufficiently large set into itself than there are elements of that set, and it is only possible to think through all such mappings individually if we are already able to think through the entire set of predecessors.

In order to use the property of Dedekind finitude to come to understand the structure of $\mathbb{N}$, then, we must consider fairly specific types of mappings that the set of predecessors of a standard element cannot support, but that the set of predecessors of a nonstandard element can support. Suitable types of mappings both must be cognitively accessible to us and must in some way guide us to understand the key structural difference(s) between Dedekind finite and Dedekind infinite sets of predecessors. Presumably suitable types of mappings will distinguish Dedekind finite sets from Dedekind infinite sets in virtue of the fact that whenever the set is Dedekind infinite, at least one mapping of the type will be a one-to-one mapping into a proper subset.

The cognitive accessibility constraint on mappings is significant. If a type of mapping is to distinguish Dedekind finite sets of predecessors from Dedekind infinite sets of predecessors, it must be a type of mapping that can be applied to sets with many more elements than we are cognitively capable of surveying individually. The mappings, therefore, must be in accord with relatively simple patterns or rules—patterns or rules that we can understand. And while there may well be different mapping rules (or disjunctive mapping rules) for different subsets of the set of predecessors (e.g. m maps to $S(m)$ for all odd numbers of the set, and m maps to itself for all even numbers of the set), cognitive accessibility demands that the mappings not involve different rules for too many disjoint sets of predecessors.

In fact, any suitable mapping type must involve disjunctive mapping rules, in which the rule (or part of a rule) that determines where an element gets mapped depends on which (proper) subset(s) of the domain contain the element. Without disjunctive mapping rules, it is not possible to map even a Dedekind infinite set of predecessors into one of its proper subsets in a one-to-one and cognitively accessible way. This is the case because every set of predecessors of an element of an arithmetical structure has both a first element and a last element, according to the predecessor ordering. Intuitively, if the mapping does not shift the placement of any elements (i.e. every element maps to itself), then its image is not a proper subset. If, on the other hand, the mapping shifts all elements by the same rule (e.g. each element m maps to $S(m)$), it shifts the first element and the last element in the same way as it does all the others; its image then includes things outside the subset (e.g. the successor of the last element). In any case, non-disjunctive mappings cannot witness the Dedekind infinity of any set of predecessors with a first and last element.[28]

A suitable class of mappings must witness the difference between Dedekind finite and Dedekind infinite sets of predecessors. Accordingly, such a class must include disjunctive mappings (at least on all Dedekind infinite sets of predecessors). Furthermore, if all of the

[28]Admittedly, I have not established this point beyond doubt. But I challenge the reader to come up with a counterexample—a cognitively accessible, non-disjunctive mapping from a set of predecessors into itself that can witnesses the set's being Dedekind infinite, if indeed it is.

subsets on which the disjuncts of a disjunctive mapping rule are defined have both a first and a last element, the mapping cannot witness that a set of predecessors is Dedkind infinite. In order for a mapping to witness the Dedekind infinity of a set of predecessors, at least one disjunct of the mapping rule must only apply on a subset with either no first element or no last element. So, in order for a class of mappings to help us distinguish Dedekind finite sets of predecessors from Dedekind infinite ones, the class must include mappings of this sort—mappings with at least one disjunct of the mapping rule defined on a subset with either no first or no last element—precisely when the class contains mappings over Dedekind infinite sets of predecessors. If we are to use Dedekind finitude and Dedekind infinity to come to understand the structure of $\mathbb{N}$, we must be able to recognize when a set of predecessors contains a subset with no first element or with no last element. We must be able to recognize when a set of predecessors contains a progression of elements, ascending or descending, with no final element.

An example will help illustrate this point. Consider the class of mappings that take m to $S(m)$ for m in some initial segment of the progression of successors up to some element n, and take m to itself for all of the other predecessors of n. Intuitively, these mappings shift elements at the beginning of the set of predecessors up one spot, and hold stationary elements at the end of P_n. Notice that the image of the set of predecessors of n under any such mapping is a proper subset of itself because it does not include the element 0. This class of mappings will include one-to-one mappings (and hence will be Dedekind infinite) exactly when there are progressions of successors of n with no final element. So, if we understand what it is for the predecessors of some arbitrary element n to include a progression of successors with no final element, we can understand the difference between Dedekind finite and Dedekind infinite sets, and hence can understand the structure of $\mathbb{N}$.

In any model of arithmetic, the full set of predecessors of any nonzero element is a set with a least element, viz. zero, and a greatest element, viz. the immediate predecessor of the element whose predecessors comprise the set. If we are to distinguish Dedekind finite sets of predecessors from Dedekind infinite sets of predecessors, we need to understand some difference between the structures of these two kinds of sets. It seems that our route to understanding this difference requires us to understand what it is for an ordered set of predecessors to contain within it a progression with no final element. This could be an ascending progression that goes from each element to a greater element, or it could be a descending progression that goes from each element to a lesser element.

Either way, if the properties of Dedekind finitude and Dedekind infinity help us to understand the structure of $\mathbb{N}$, we must be able to understand the difference between Dedekind finite and Dedekind infinite sets of predecessors, which requires us to understand what it is for a set of predecessors to contain such progressions. We shall soon see that this same type of antecedent understanding is required for us to come to understand the structure of $\mathbb{N}$ through the two other standard second-order means: the taking of intersections and the least number principle.

2.3.3 Taking Intersections

Another second-order method is rather different, at least superficially. Instead of using a property and second-order mathematical induction to establish the structure of $\mathbb{N}$, this approach relies on the set-theoretic operation of intersection to define the $\mathbb{N}$-structure.[29] In brief, a standard model of arithmetic is one whose domain is the intersection of—the set of elements that are common to—all the domains of all the models of arithmetic that share some particular ground element and some particular successor function.[30] The domain of this standard model, then, is that of the initial $\mathbb{N}$-chain part of the progression of successors

[29]Dedekind, for example, presents a definition much like the one presented here. He then uses that definition to prove the validity of second-order induction. See *The Nature and Meaning of the Natural Numbers*.

[30]Two issues deserve note here. One is that two models of arithmetic cannot share the exact same successor function. But one can be a restriction or an extension of the other. The other is that we must quantify over all sets of elements if we are to take the interection of *all* of the models of PA^1 that share an element and a successor function. This is a case of unrestricted comprehension, which leads to paradox. But the approach is easily salvaged if we take some particular model of PA^1 to quantify over all of its submodels.

beginning with the identified ground element.

If we are to use this type of definition to come to understand the structure of $\mathbb{N}$, we must recognize that the relevant intersection is nonempty, and in fact is itself a model of PA^1. Presumably we can recognize this. Trivially, 0 must be in the intersection of all the domains of models of PA^1 that begin with 0. Since every element in these domains has a successor, $S(0)$ must also be in the intersection, and $S(S(0))$, and so on. No element contained in the intersection can be the last such in the progression of successors.

Let us try to build up an understanding of $\mathbb{N}$ from the resources currently available. Let m be an element that is nonstandard according to the intersection definition. We are aware that such elements exist in some models because we appreciate the fact that taking intersections sets a restriction on models of PA^1; otherwise, taking intersections would not help us hone in on the $\mathbb{N}$-structure. Then, since m is nonstandard, it is not in the intersection of all models beginning with 0. So, there is some progression of successors beginning at 0 that has no last element, and all of the elements of which are predecessors of m. Clearly m is not the ground element of a model in which it is a nonstandard element, so m has an immediate predecessor. Its immediate predecessor also is not in the intersection, since by hypothesis m is not in the intersection and the immediate successor of any element in the intersection is also in the intersection. Neither is the predecessor of m's predecessor, nor its predecessor, and so on. Hence there is a progression of immediate predecessors, starting with m, that are not in the intersection of the domains of all the models of PA^1 that begin with 0. Furthermore, this progression cannot have a final element.[31]

Thus, it seems that we can use the intersection method to provide a sufficient, thorough account of the structure of $\mathbb{N}$. The intersection definition gives us the resources to de-

[31]This result is easy to demonstrate. If there is a final element, m', then either m' lacks an immediate predecessor in the model, or it has one. If m' has no immediate predecessor in the model, it is the ground element; m' would then be in the intersection. If m' has an immediate predecessor, that predecessor must be in the intersection; otherwise it would be in the sequence. But then m' would be the immediate successor of an element in the intersection, and so would also be in the intersection. But m' was defined as an element of a progression that is not in the intersection. So m' can neither have nor lack an immediate predecessor; we reach a contradiction if we assume that there is a last element to this progression of predecessors.

scribe the structure of the standard model as that of the smallest successor- and 0-preserving model contained in any model of PA^1. We can use this description to come to understand that structure, provided that we understand what it is for every element of the model to have a certain feature: every progression of immediate predecessors beginning with that element must have a final element. This, of course, is the very same feature we need to understand if we are to come to understand the structure of $\mathbb{N}$ through the properties of Dedekind finitude and Dedekind infinity.

2.3.4 The Least Number Principle

The third second-order method of definitionally distinguishing the structure of $\mathbb{N}$ from the structures of nonstandard models of PA^1 uses the least number principle. Recall this principle from §2: every non-empty set of numbers has a least number. That is, if the set is non-empty, then there is an element of the set that is less than every other member of the set. In symbols:

$$(\forall P)\Big((\exists x)Px \rightarrow (\exists x)\Big(Px \wedge (\forall y)\big(Py \rightarrow (y = x \vee x < y)\big)\Big)\Big).$$

The least number principle holds for standard models of arithmetic, but not nonstandard ones. Intuitively, this is clear from the fact that nonstandard models contain $\mathbb{Z}$-chains, whose elements comprise sets with no least element. But this insight cannot explain how we come to understand the structure of $\mathbb{N}$ through axioms. If we come to understand the structure of $\mathbb{N}$ because we antecedently understand the structure of $\mathbb{Z}$, our understanding of the $\mathbb{N}$-structure instead is grounded in whatever grounds our understanding of the $\mathbb{Z}$-structure, which will encounter analogous challenges.

Presumably there are far too many sets of elements in a model of PA^1 for us to survey to ensure that every set has a least element. Perhaps, however, we can again focus on fairly specific and accessible classes of sets—ones that distinguish standard models of PA^1 from nonstandard ones in virtue of whether or not the classes include any sets that violate the least number principle. The paradigmatic example of such an accessible class is, yet again, that of the descending progressions of immediate predecessors of elements. An element of a standard model cannot have a descending progression of predecessors with no final element because the elements of such a sequence would comprise a set with no least element.

2.3.5 Descending Progressions

We have already seen that first-order axioms are not robust enough to ground our understanding of the structure of ℕ. While second-order axioms (interpreted with their full semantics) are robust enough to uniquely characterize that structure, the more intuitive approaches we might use to identify the structure they characterize cannot help us to understand the structure of ℕ; these approaches require us to antecedently understand that structure. We are left, then, with more technical second-order approaches. It has become apparent that the three standard technical definitions of the ℕ-structure that we have considered—definitions of infinity, the intersection of sets, and the least number principle—are only able to guide us to the structure of ℕ, in all three cases, if we understand what it is for there to be a descending progression of immediate predecessors that begins with one element but has no final element.[32] We must then come to understand what it is for each element of a model of arithmetic to have no such descending progression of predecessors. To understand this is to understand the structure of ℕ.

That is not to say, however, that we have discovered how to use axioms to come to understand the structure of ℕ. Making sense of verbal expressions ('the element begins no decending progressions of predecessors with no final element' or 'every descending progression of immediate predecessors beginning at the element has a final element') is not enough. To

[32] Alternatively, these approaches might require us to understand what it is for the predecessors of an element to include an ascending progression of successors starting at zero that has no final element. But these two ideas amount to the same thing. One is a descending progression starting at some element n and ending before 0; the other is an ascending progression starting with 0 and ending before some element n.

come to genuinely *understand* the structure of $\mathbb{N}$, we must fully understand the structural limitations that the property so described imposes on an element of an arithmetical structure. And if we genuinely understand what it is for an element to have this feature—for it to begin no descending progressions of predecessors that lack a final element—we must understand the most simple and basic structure of a progression with no final element. And that is the structure of $\mathbb{N}$. It seems, then, that we must already understand the structure of $\mathbb{N}$ if we are to successfully identify that structure by second-order means. We cannot use PA^2 to come to understand the structure-type of $\mathbb{N}$.

2.4 Resisting Relativism

One stance that could be taken in response to this result is the stance that there is no *standard* or *intended* model of arithmetic. There are only models of arithmetic. Some interpret Skolem to have taken such a view in his later years,[33] and Putnam seems also to have entertained a view of this sort.[34] In both cases the view (or putative view) appears in response to the fact that first-order axioms cannot describe a structure up to isomorphism.

We made two fairly substantive assumptions early in this paper. (1) There is a standard model-type of arithmetic, with the structure of $\mathbb{N}$. (2) We understand that model well enough to distinguish it from nonstandard models. Taking a skeptical or relativistic position on either of these assumptions has serious consequences—consequences that ultimately serve to undermine the motivational grounds of that denial. Denying (1) renders the idea of a 'finite' set ultimately meaningless. Denying (2) interferes with our ability to use the term 'finite' (or the concept) to pick out an unambiguous property in the world. Denying either of these assumptions completely undermines our ability to use reasoning about arbitrary finite things (e.g. finite sets) or anything infinite.

Notice that the very idea of a sentence of a first-order language becomes ambiguous

[33] See Benacerraf and Wright, "Skolem and the Skeptic."

[34] See Putnam "Models and Reality."

when we deny (1) or (2); a sentence has finite length. Accordingly, the very set-theoretic results that we use to conclude that first-order sentences cannot characterize structures up to isomorphism all end up relying on the assumed categorical (and understood) structure of $\mathbb{N}$. These results also all presuppose the idea of finitude in more explicit ways.[35] If the set-theoretic results depend on the idea of a standard model, we ought not to use those results to motivate the view that there is no standard model.

But lastly, it is just deeply implausible that we are unable to reason about finitude and infinity. Surely we are making an unambiguous claim when we claim that there are infinitely many prime numbers, or that Euclidean space is Archimedean (i.e. that we can cover any distance of Euclidean space with finitely many copies of any other distance of Euclidean space). Surely we are disinclined to deny the meaningfulness of these sorts of claims.

If we are to do any interesting math at all, it seems that we must accept the uniqueness (up to isomorphism) of the structure of $\mathbb{N}$. We also must understand that structure-type. If, as I have argued, we cannot get this understanding from axioms, then we must get it elsewhere. The structure of $\mathbb{N}$, it seems, is basic, primitive, and foundational.

[35] The statement of Compactness, for example, makes use of idea of finite subsets. The Lowenheim-Skolem Theorem makes use of countable sets—i.e. sets that can be put into one-to-one correspondence with the (standard) natural numbers.

2.5 Appendix I

We must show that no first-order set of sentences can characterize $\mathbb{N}$ up to isomorphism. In order to do so, we shall use the Compactness Theorem, which is the theorem that a set of first-order sentences has a model as long as all of its finite subsets have models. And we shall argue by contradiction.

Suppose that there is a set, Γ, of sentences that characterizes the structure of $\mathbb{N}$ up to isomorphism. Then every finite set of sentences of Γ is satisfied by $\mathbb{N}$, so every finite subset of Γ has a model. Now, we shall expand Γ to Γ', whose members include all of the members of Γ, and also all of the following sentences:

$(\exists n)n \neq 0,$

$(\exists n)(n \neq S(0) \wedge n \neq 0)$, [i.e. $n \neq 0 \wedge n \neq 1$]

$(\exists n)(n \neq S(S(0)) \wedge n \neq S(0) \wedge n \neq 0)$, [i.e. $n \neq 0 \wedge n \neq 1 \wedge n \neq 2$]

and so on, for all of the sentences that can be so formed. (Recall that all sentences are finite.) Now, notice that any *finite* subset of these new sentences will be satisfied in $\mathbb{N}$. So, since evry finite subset of Γ is satisfied in $\mathbb{N}$, and every finite subset of these additional sentences is satisfied in $\mathbb{N}$, any finite set of sentences of Γ' is satisfied in $\mathbb{N}$. By Compactness, then, Γ' has a model. But the model of Γ' must contain an element not in $\mathbb{N}$, since it satisfies these additional sentences. So, Γ' has a model that is not iomorphic to $\mathbb{N}$. Since this model, $\mathbb{M}$, satisfies Γ', it must also satisfy its subset Γ. So, Γ did not uniuely characterize the structure of $\mathbb{N}$. We reach a contradiction with our supposition. So, no first-order theory can uniquely characterize the structure of $\mathbb{N}$.

2.6 Appendix II

Suppose $\mathbb{M}$ and $\mathbb{N}$ are both models of second-order PA. Then construct a mapping $f : \mathbb{M} \to \mathbb{N}$ as follows:

$$f(0_{\mathbb{M}}) = 0_{\mathbb{N}}$$

$$f(S_{\mathbb{M}}(n)) = S_{\mathbb{N}}(f(n))$$

Clearly f preserves the ground element and the successor relation. To show that f is an isomorphism, we must only show that it is a one-to-one and onto mapping, and that it preserves predecessors.

Onto: Suppose f is not onto. That is, suppose there are $n \in \mathbb{N}$ that f does not map to. Then we can define a proper subset $\mathbb{N}^*$ of $\mathbb{N}$: $\mathbb{N}^*$ is the image of $\mathbb{M}$ under f. Now consider induction using the subset $\mathbb{N}^*$. Since $f(0_{\mathbb{M}}) = 0_{\mathbb{N}}$, $0_{\mathbb{N}} \in \mathbb{N}^*$. Also, for any $n \in \mathbb{N}$ to also be in $\mathbb{N}^*$, there must be some $m \in \mathbb{M}$ such that $f(m) = n$. Since $\mathbb{M}$ satisfies second-order PA, it is closed under successor, so $S_{\mathbb{M}}(m) \in \mathbb{M}$. Using the definition of f, then, $f(S_{\mathbb{M}}(m)) = S_{\mathbb{N}}(f(m)) = S_{\mathbb{N}}(n)$. So, if n is in the image of $\mathbb{M}$ under f, then so is $S_{\mathbb{N}}(n)$. So, if $n \in \mathbb{N}^*$, then $S_{\mathbb{N}}(n) \in \mathbb{N}^*$. But since $\mathbb{N}^*$ is a proper subset of $\mathbb{N}$, $\mathbb{N}$ could not have satisfied second-order PA. Since $\mathbb{N}$ does satisfy second-order PA, the supposition must have been false, and f must be onto.

One-to-one: Suppose f is not one-to-one. That is, suppose f maps some distinct m and m' in $\mathbb{M}$ to the same n in $\mathbb{N}$. Then let $\mathbb{M}^*$ be the set of those m in $\mathbb{M}$ such that for all $m' <_{\mathbb{M}} m$, $f(m) \neq f(m')$. Then, given our supposition, $\mathbb{M}^*$ is a proper subset of $\mathbb{M}$. Since there is no $m <_{\mathbb{M}} 0_{\mathbb{M}}$, it is clear that $0_{\mathbb{M}} \in \mathbb{M}^*$. Now suppose $m \in \mathbb{M}^*$, and $f(m) = n$. Then no $m' <_{\mathbb{M}} m$ is such that $f(m') = f(m)$. Now suppose $S_{\mathbb{M}}(m)$ is not in $\mathbb{M}^*$. Then for some $m' <_{\mathbb{M}} S_{\mathbb{M}}(m)$, we get $f(m') = f(S_{\mathbb{M}}(m)) = S_{\mathbb{N}}(f(m)) = S_{\mathbb{N}}(n)$. If $m' = 0_{\mathbb{M}}$, then $0_{\mathbb{N}} = f(0_{\mathbb{M}}) = f(m') = S_{\mathbb{N}}(n)$, which violates the axioms of PA. If $m' \neq 0_{\mathbb{M}}$, then $m' = S_{\mathbb{M}}(m'')$ for some $m'' \in \mathbb{M}$. But then $S_{\mathbb{N}}(f(m'')) = f(S_{\mathbb{M}}(m'')) = f(m') = S_{\mathbb{N}}(n)$. Since $\mathbb{N}$ satisfies second-order PA, $S_{\mathbb{N}}$ is one-to-one, so $f(m'') = n = f(m)$. But if $S_{\mathbb{M}}(m'') = m' <_{\mathbb{M}} S_{\mathbb{M}}(m)$, then $m'' <_{\mathbb{M}} m$, and we already have from the supposition that $m \in \mathbb{M}^*$ that no $m' <_{\mathbb{M}} m$ is such that $f(m') = f(m)$. So, if $m \in \mathbb{M}^*$, then $S_{\mathbb{M}}(m) \in \mathbb{M}^*$. But since $\mathbb{M}^*$ is a proper subset of $\mathbb{M}$, $\mathbb{M}$ then fails to satisfy second-order induction. So f must be one-to-one.

Predecessor: We proceed by induction. We will show that for arbitrary n, the following holds for all m: $m <_{\mathbb{M}} n \rightarrow f(m) <_{\mathbb{N}} f(n)$.

Case 1: $m = 0_{\mathbb{M}}$. Suppose $0_{\mathbb{M}} <_{\mathbb{M}} n$. Then $0_{\mathbb{M}} \neq n$, by antisymmetry. And by the definition of f, $f(0_{\mathbb{M}}) = 0_{\mathbb{N}}$. Since f is one-to-one, $f(n) \neq 0_{\mathbb{N}} = f(0_{\mathbb{M}})$. So, then, $f(0_{\mathbb{M}}) = 0_{\mathbb{N}} <_{\mathbb{N}} f(n) \neq 0_{\mathbb{N}}$. The result holds.

Case 2: Suppose $m <_{\mathbb{M}} n \rightarrow f(m) <_{\mathbb{N}} f(n)$. We must show $S_{\mathbb{M}}(m) <_{\mathbb{M}} n \rightarrow f(S_{\mathbb{M}}(m)) <_{\mathbb{N}} f(n)$. Suppose not. Then $S_{\mathbb{M}}(m) <_{\mathbb{M}} n$ and $f(S_{\mathbb{M}}(m)) \not<_{\mathbb{N}} f(n)$. By the latter, and the definition of predecessor, (i) $f(S_{\mathbb{M}}(m)) = f(n)$ or (ii) $f(n) <_{\mathbb{N}} f(S_{\mathbb{M}}(m))$. If (i), then since f is one-to-one, $S_{\mathbb{M}}(m) = n$, contradicting $S_{\mathbb{M}}(m) <_{\mathbb{M}} n$. If the latter, $f(n) <_{\mathbb{N}} f(S_{\mathbb{M}}(m)) = S_{\mathbb{N}}(f(m))$ by the definition of f. Now, notice that if $f(n) <_{\mathbb{N}} S_{\mathbb{N}}(f(m))$, then $f(m) \not<_{\mathbb{N}} f(n)$. This is a specific case of if $A < S(B)$ then $B \not< A$. Since, then, $f(m) \not<_{\mathbb{N}} f(n)$, by the original supposition, $m \not<_{\mathbb{M}} n$. But then $S_{\mathbb{M}}(m) <_{\mathbb{M}} n$. cannot be true, as we had concluded it was. So we reach a contradiction. Thus, the second supposition must be false, and the case must hold. So, by mathematical induction, the result holds for all m.

Thus, we have that f is an isomorphism, so $\mathbb{M}$ and $\mathbb{N}$ are isomorphic. Since these were two arbitrary models satisfying second-order PA, the axioms of second-order PA are categorical, and pick out the standard models of arithmetic.

Chapter 3

On Interaction Problems and Hume's Principle

Platonists about mathematics face two intuitively gripping problems that I call 'the interaction problems'. This chapter addresses a fairly prominent neo-Fregean attempt to resolve these problems by appealing to an implicit definition. I argue that this approach, proposed by Bob Hale and Crispin Wright, cannot succeed. The view seems to be committed to an implausible claim about the origin of mathematical thought, viz. that all cognition of the natural numbers is grounded in a particular truth-making stipulation. The stipulation required is only possible with fairly robust logical resources that were not available until long after people had established arithmetical thought.

3.1 Interaction Problems for Platonism

Mathematical platonism is the view that there are abstract mathematical objects, and that mathematical truths describe the properties of and relations between these objects. For example, it is true that 2+3=5 in virtue of relations that hold between 2, 3, and 5, all of which are abstract objects. Of course, it is difficult to explain what it is to be an abstract object. The standard characterization is largely negative: abstract objects are not spatial, they are not temporal, and they are not in the causal order of things. This characterization of abstract objects as outside the causal order leaves platonism about mathematics vulnerable to two challenges: the interaction problems.[1]

The first interaction problem, posed by Paul Benacerraf,[2] highlights an apparent tension between our inability to causally interact with abstract mathematical objects and our ability to know things about such objects. If abstract mathematical objects are not causally efficacious, how do we come to have knowledge of their properties and relations? That is, how do we come to know mathematical truths? We cannot, for example, come to know the properties of causally inert objects through ordinary perception, as we come to know so many other truths. The platonist is challenged to provide an epistemology that accounts for mathematical knowledge in a way that is contiguous with a reasonable account of knowledge in non-mathematical cases.

The second interaction problem, posed by Bob Hale and Crispin Wright,[3] is similar to the knowledge problem.[4] There is an apparent tension between our inability to causally interact with abstract mathematical objects and our ability to *refer* to such objects. How do we account for the fact that abstract objects are the referents of our mathematical terms? A causal-historical picture of reference, for example, seems not to cover reference to mathematical abstracta, again because they are causally inert. Thus, the platonist faces a second challenge: identifying a reference-fixing mechanism for mathematical terms that is contiguous with a reasonable account of reference-fixing for non-mathematical terms.

The interaction problems typically strike philosophers as intuitively compelling.[5] Pre-

[1]Perhaps other veins of platonism are also vulnerable to the interaction problems. In this paper, I only consider the case of platonism about mathematics.

[2]See "Mathematical Truth" in *The Journal of Philosophy* 1973.

[3]See "Benacerraf's Dilemma Revisited," in *European Journal of Philosophy*, 2002. Until otherwise noted, all references to Hale and Wright are references to this paper.

[4]Benacerraf does not directly present this problem, but comes very close.

[5]Benacerraf's knowledge problem, in particular, has been one of the central problems in the philosophy of mathematics since "Mathematical Truth" was published in 1973.

sumably, the fact that we cannot causally interact with mathematical abstracta makes it obscure to us how we might come to think about them at all. A worry about our ability to come to think about mathematical abstracta in the first place is, I think, at the heart of the interaction problem for reference. Very likely, the interaction problem for knowledge initially was presented to address that same worry. But the knowledge case is potentially even more difficult. Not only must we be able to think about these objects, we must be able to think about them in a truth-tracking way. After all, mathematical knowledge is paradigmatically secure.

3.1.1 Descriptions, Interaction, and Object-Directed Thought

The source of our worry about coming to think about causally inert mathematical abstracta, however, is not entirely clear. We routinely come to think about, know about, and refer to things with which we do not *directly* interact causally. I can refer to and know truths about Julius Caesar, though we have had no direct causal interaction. That said, I have interacted with Caesar causally, albeit indirectly, through a chain of causal interactions of communication somehow grounded in long-ago perceptions of the man himself.

It also is fairly clear that we refer to things with which we have not causally interacted, directly or indirectly, in any way that grounds object-directed thought. We do so using definite descriptions. For example, I can think about and refer to the oldest living man who has summitted Mt. Everest. In fact, in saying that, I just *did* refer to him. My ability to refer to this man through a definite description in no way presupposes any kind of interaction with him through any of the relevant causal channels, such as perception or acts of communication.[6]

I even have identifying knowledge of the individual that I am able to think about through this description. I know that he is the oldest living man who has summitted Mt.

[6]Presumably I do have some very indirect physical interaction with the man in question, for butterfly-effect reasons. But that is not the kind of physical interaction that is usually taken to ground the ability to think about a thing.

Everest. All I had to do to get this knowledge (that the oldest living man who has summitted Mt. Everest is the oldest living man who has summitted Mt. Everest) was state the law of self-identity in the case of a definite description that I know has *some* unique satisfier, I know not what. And not all of my knowledge of this man is trivial. I know that he has summitted a taller mountain than I have, and that he is older than I am.

What, then, is the problem with coming to think about causally inert mathematical abstracta? A definite description provides the resources required to refer to and know things about an object, including nontrivial things, without having the right kind of causal connection to it. But in the case just considered, we do have the right kind of causal connections to the elements of that description. Perhaps we haven't seen Mt. Everest (I haven't), but all of us have heard about it from others.[7]

Mathematical abstracta also satisfy definite descriptions. Five, for example, is the number of fingers on my left hand. But we do not have causal interactions with all of the parts of this description. In order to refer to the number five using this description, I presumably must already be able to think about number. No causal story can explain how that is possible. This is what is at issue in the interaction problems.

It is somewhat tempting to think that we cannot think about an object unless we have the right sorts of causal connections, either to it or to the elements of an identifying description of it. Hale and Wright credit a related and objectionable assumption with making many philosophers take the interaction problems to be rather more hopeless than they actually are. The assumption is that some kind of interaction with an object is required to think about that object, and that this interaction depends on a causal connection—in particular, a *physical* connection—with the object.[8] This assumption is objectionable because, as

[7]I am not entirely sure what to say about the other terms used in the description. I would like to say that I am familiar enough with the concepts of 'older' and 'man' and 'living' to make sense of the description. I have causally interacted with instances of each of these concepts, and I have done so in ways that seem to be relevant for thinking about things.

[8]Strictly speaking, they only identify this assumption in the case of the knowledge problem. But that is only because they identify the assumption before presenting the reference problem. They also do not consider the possibility of definite descriptions in this discussion.

Hale and Wright point out, it "is obviously inimical to the abstract." That seems right. To *assume* that only physical interaction will do is to make an extreme physicalist assumption. Even if it is not immediately obvious to us how we might come to think about objects with which we do not causally interact, we ought not to assume it to be impossible.

3.2 The Fregean Solution

Hale and Wright claim that a neo-Fregean solution can resolve the interaction problems for some mathematical abstracta: the natural numbers. The general strategy is to devise an account of how we could come to understand descriptive phrases like 'the number of fingers on my left hand' that does not require causal interaction with numbers. We could then provide those numbers with new names, e.g. 'five'. On Frege's account, we could ground our grasp of the meanings of number-words in our grasp of the meanings of words of other sorts.[9] If this account succeeds, it both gives us the ability to think about some mathematical abstracta and provides a sufficient basis for us to come to know quite a bit about arithmetic.[10]

In our discussion of this Fregean solution, let us grant two controversial claims. First, let us grant that number-words pick out abstract number-objects;[11] our goal is to explain

[9]See *The Foundations of Arithmetic* §62-70, pp. 73-81 in the English translation.

[10]From Hume's Principle, which is central to Frege's account, it is possible to derive a version of the Dedekind-Peano axioms.

[11]Frege does not assume this, and takes some effort to establish it. The first step in his argument for this in the *Foundations* is to establish that, grammatically, number words are singular terms.

On the surface, there are two grammatical uses of number words. Number-words can be used as singular terms, as in, 'The number of fingers on my left hand is *five*'. But they can also be used as modifiers, as in, 'I have *five* fingers on my left hand'. Wright (in *Frege's Conception of Numbers as Objects*, pp.11-12) identifies three reasons that Frege has for believing that number-words fundamentally figure as singular terms and not as modifiers. First, the same content that is expressed by sentences using number-words as modifiers can also be expressed by sentences that use number-words as singular terms. In the examples above, the content of the sentence featuring a modificational use of the word *five* can be expressed equally well by the sentence using the word *five* as a singular term. Second, "we readily apply the definite article to number words and expressions," as in 'the number of fingers . . . ' or 'the number five'. And third, identity statements involving numbers are central to number theory.

After establishing that number words are singular terms, Frege argues further in §55-62 that "number words are to be taken as standing for self-subsistent objects." In "Frege's Platonism," (*The Philosophical Quarterly* 1984) Hale gives a succinct account of a Fregean argument for this claim:

> [I]f there is a range of expressions members of which function as singular terms in true

how number words come to do this. Second, for the sake of argument, let us grant the *context principle*, which is an important starting point for Frege in the *Grundlagen*. This is the principle that words only have meaning in the context of sentences.[12] It is a general principle for Frege. So, the name 'Mt. Everest' only has meaning in the context of sentences involving the name 'Mt. Everest'. Number words (e.g. 'five') only have meaning in virtue of the meanings of sentences involving them (e.g. 'The number of fingers on my left hand is five' or 'Three and two add up to five'). If Hale and Wright can use the context principle to resolve the interaction problems in a plausible way, they have made considerable progress.

Assuming that numbers are self-subsistent abstract objects and that we can only grasp the meanings of number words in the context of sentences containing them, the status of the remaining problem stands thus, in Hale and Wright's words:

> The outstanding question, after Frege's re-orientation of the issues, is how an understanding may be acquired of those forms of statement which, if true and taken to involve the reference to and quantification over abstracta which they seem to involve, serve to give us the means to think of and refer to such objects. (22)

One relevant form of statement is that of identity statements. Identity statements are central in arithmetic; this observation plays a significant role in Frege's argument that number-words pick out self-subsistent objects. Given their centrality to the science of arithmetic, identity claims involving number-words must be meaningful. Providing the meanings of identity statements involving number-words appears to be a promising approach to grasping the meanings of number-words.

> statements, then there exists a range of objects corresponding to them. But numerals and other numerical expressions do so function—notably, though not exclusively, in ... arithmetical equations. Hence there exists a range of numerical objects to which reference is made in arithmetical statements. (pp. 40-41)

[12]Frege himself puts it slightly differently, as a principle: "never to ask the meaning of a word in isolation, but only in the context of a proposition" (Introduction p.x). For our purposes, my paraphrase will do.

Unfortunately, textual difficulties leave unclear the sort of meaning that must be provided for number-identity statements in order for us to grasp the meanings of number-words. Frege wrote *The Foundations of Arithmetic* before making his significant distinction between Sinn (sense) and Bedeutung (denotation).[13] When Frege discusses the *meanings* of number-words, or of numerical identity statements, it is not clear whether he intends to talk about their senses, their denotations, or something else entirely. He appears to use 'Sinn' and 'Bedeutung' (and their variants) interchangeably.[14] Hale and Wright suggest, quite reasonably, that providing truth-conditions for numerical identity statements will suffice for providing their meanings in the relevant sense.[15]

We follow Hale and Wright (and perhaps also Frege), then, in supposing that we can grasp the meaning of numerical identity statements if we give the truth-conditions of identity statements involving number-terms. Once we grasp the meanings of numerical identity statements, we have identifying knowledge of numbers; we know the conditions of identity for numbers. Since identifying knowledge of numbers suffices for us to have the ability to engage with (or think about) the identified numbers, we would then be able to have knowledge about numbers and refer to them. In other words, the interaction problems will be resolved—at least, for numbers—if we can give truth-conditions for number-identity statements without making reference to any numbers or other mathematical objects (or other potential unknowns) in the process of giving those truth conditions.

[13]See "Sinn und Bedeutung."

[14]When Frege gives the context principle, he uses a variant of 'Bedeutung' ("bedeuten die Wörter"—literally 'the words mean/denote'); words have denotations only in the context of sentences. He then switches to 'Sinn', and says that once we give the sense (Sinn) of numerical identity sentences, we can refer to number objects (§62). A few sections later (§65), he takes himself to have successfully completed this task by giving another sentence with the same meaning as a relevant identity sentence, using a variant of 'Bedeutung' again ("sei gleichbedeutend mit"—literally 'are same-meaning with' or 'are same-denoting with').

[15]There might seem to be a *prima facie* worry about using truth conditions to provide the meanings of statements about mathematical objects. This does tend to present a problem for certain kinds of truth conditions, as we see in possible world semantics—or at least in its unsophisticated versions, which are unable to distinguish between distinct necessary truths. But Frege's approach to truth conditions are of a rather different sort. It might be worth attending to the question of whether or not Frege's truth-conditions adequately distinguish between the meanings of distinct mathematical truths. My instinct is to believe that they can so distinguish.

3.2.1 Abstraction Principles: Hume's Principle and Direction Equivalence

Consider a number-identity claim:

$$\text{The number of F's } \textit{is identical to} \text{ the number of G's.}^{16} \tag{3.1}$$

According to Hale and Wright, the trick to resolving the interaction problems is to explain how we could come to know the truth-conditions for (1) without relying on any antecedent ability to think about numbers. They suggest that we could come to know these truth-conditions through Hume's Principle:[17]

$$\begin{aligned} &\text{The number of F's} = \text{the number of G's} \\ &\iff \text{there is a one-one correspondence between the F's and the G's.} \end{aligned} \tag{3.2}$$

Since we can account for one-one correspondence without appealing to numbers, it is possible to know when the right-hand side of the biconditional is true without relying on knowledge of numbers.[18] Hume's Principle, then, seems to offer a means by which we could come to

[16] Hale and Wright state (1) as, "The number of F's = The number of G's." Frege writes, "The number which belongs to the concept F is the same as that which belongs to the concept G." These slight variations are of no substantive import for our purposes.

[17] Frege presents this, as Hume did, in terms of one-to-one correspondence. I follow suit. Hale and Wright sometimes present it this way, and sometimes present it in terms of *equinumerocity*: The number of F's = the number of G's $\iff$ the F's and the G's are equinumerous. In "Cardinal, Counting, and Equinumerosity" (*Notre Dame Journal of Formal Logic*, 2002) Richard Heck focuses on the equinumerocity idea, as contrasted with the one-to-one correspondence idea.

[18] Frege gives a purely logical account of one-one correspondence in *Grundlagen* Sections 70-73. To give a more modern version, we can say that for the F's and G's to be in one-one correspondence is for the following condition to be met:

$$\exists R\Big(\forall x\big((Fx \to \exists y(Gy \land Rxy)) \land (Gx \to \exists y(Fy \land Ryx))\big)\land$$
$$\forall x \forall y \forall z\big(((Fx \land Gy \land Rxy \land Rzy) \to x = z) \land ((Fx \land Gy \land Rxy \land Rxz) \to y = z)\big)\Big).$$

This expression of one-one correspondence, like Frege's version, is a second-order statement, but the first order schema is also available. For the purposes of demonstrating that we don't need to use numbers to account for one-one correspondence, I don't think it matters which version we use. If we are further concerned about the existence of relations (as entities), that might be another matter.

know the truth-conditions of number-identity statements in the form of (1) without having antecedent knowledge or cognition of numbers.

According to Hale and Wright, the truth conditions of a statement fix its meaning. In light of the context principle, if the meaning of (1) is in keeping with the meanings of other kinds of identity claims, then knowing the truth-conditions of (1) positions us to individuate numbers. So, knowing Hume's Principle would position us to have identifying knowledge of numbers, even without antecedent cognition of any mathematical objects. As we saw with definite descriptions, identifying knowledge gives us a way to think about objects. Thus, if Hale and Wright are correct, knowledge of Hume's Principle would position us to have knowledge of numbers, and to begin to think about them. If they are right, we can resolve the interaction problems by accounting for our knowledge of Hume's Principle without relying on antecedent knowledge of mathematical objects.

To explain more fully how this approach works, Hale and Wright turn to the more general class of what they call 'abstraction principles'. Abstraction principles have the form:

$$(\forall\alpha)(\forall\beta)(R(\alpha) = R(\beta) \leftrightarrow \alpha \approx \beta) \tag{3.3}$$

The 'α' and 'β' denote objects or concepts that bear some equivalence relation ($\approx$), and 'R' denote a function that takes those objects (or concepts) to some abstracta, viz. the abstracta in virtue of which α and β are equivalent in the relevant sense. In the case of Hume's Principle, 'α' and 'β' are 'the concept F' and 'the concept G', '$R(x)$' is 'the number belonging to x', and the equivalence relation is that of being in one-one correspondence. But Hume's Principle isn't the only abstraction principle. Frege also makes use of another, which Hale and Wright call 'Direction Equivalence':

$$\begin{aligned} &\text{The direction of line } a = \text{the direction of line } b \\ &\iff \text{lines } a \text{ and } b \text{ are parallel.} \end{aligned} \tag{3.4}$$

Frege claims that a statement about lines being parallel "can be taken as an identity" (§64). To support this claim, he points out that being parallel is a transitive relation.[19] But when we take a statement like '$a//b$' as an identity, we do not simply interpret it as '$a = b$'; a and b are two different lines. Rather, we interpret '$a//b$' as in (4)—'the direction of line a = the direction of line b'. As Frege says, "We carve up the content [of the parallel relation between the two lines] in a way different from the original way, and this yields us a new concept" (§64). We start with what we grasp in intuition (in something like the Kantian sense), viz. lines and being parallel. From these, we extract things we do not grasp in intuition, viz. directions. The same sort of phenomenon occurs in the case of the equinumerocity of concepts and the identity of the numbers belonging to those concepts. We can grasp numbers by conceptually extracting them from things we already are familiar with—the concepts F and G, and the relation of equinumerocity.

Notice that in his explanation of how we conceptually access directions through parallelism, Frege provides evidence for thinking that the notion of identity we use when we carve up conceptual content through Direction Equivalence is our usual notion of identity. It is the same notion of identity that we use in the case of all abstraction principles, including

[19]Something bothers me about the claim that '$a//b$' is an identity. Frege is onto something in claiming that the relation of being parallel is transitive (if line a is parallel to line b, and line b is parallel to line c, then line a is parallel to line c) . The relation is clearly symmetric; if $a//b$, then $b//a$. But it seems unnatural to me to think of it as reflexive. Hence, it seems unnatural to me to think of it as entirely transitive; I'm not convinced that transitivity holds if $c = a$ in the above. I can get myself to think that $a//a$ is true, but only if I define 'parallel' for myself using the notion of direction. Of course a has the same direction as itself. But if I take the sense associated with 'parallel' as prior to the sense associate with 'direction', as Frege requires in this case, then it doesn't seem quite right that $a//a$. Another candidate notion of parallelism is Euclid's:

> Parallel straight lines are straight lines which, being in the same plane and being produced indefinitely in both directions [alt. translation: towards both the parts/regions], do not meet one another in either direction [alt. translation: region]. - Part I, Definition 23 in Euclid's *Elements*

But this notion of parallelism does not accomodate reflexivity either, since every line meets itself at every point.

The same kind of concern about reflexivity might—or might not—apply to concepts being equinumerate. To be honest, I have so much more practice using my concept of number than my concept of equinumerocity that I have trouble prodding my intuitions about the proper application of the concept of equinumerocity in abstraction from a definition involving the concept of number. The account given in fn 18 does imply that equinumerocity is reflexive.

Hume's Principle. But even more important, we haven't made any effort at all to define identity in the case of directions. Likewise, we make no effort to define identity in the special case of numbers. In Frege's words,

> We are therefore proposing not to define identity specially for this case, but to use the concept of identity, taken as already known, as a means of arriving at that which is to be regarded as being identical. (§63)

The abstraction principles have simply given the truth-conditions of identity for directional (and numerical, etc.) objects. So the notion of identity in these cases is no different from the ordinary notion we use in the case of, say, physical objects. We are only able to carve up the content of parallelism (or equinumerocity) to render the new concept of direction (or number) because we already understand the concept of identity; the concept of identity is antecedently fixed.

3.2.2 Implicit Definitions

As Hale and Wright point out, this talk of carving up content is metaphorical. Somehow we are able to move from the concept of parallelism to the new sortal concept of direction. They claim that the work is being done by an implicit definition. In this case, the implicit definition is Direction Equivalence. When we take $a//b$ to be an identity, in Hale and Wright's words, we "view the Direction Equivalence as an implicit definition of the direction operator—'the direction of . . . '—and thereby of the sortal concept of *direction*" (HW 23). Analogously, Hume's Principle is an implicit definition, one that implicitly defines the number operator—'the number of . . . '—and thereby the sortal concept of *number*.

Let's think about how these ideas work out in a non-mathematical case: color. Suppose your shirt and mine hue-match. Then they have something in common–they have the same blueish hue. The hue of your shirt is identical with the hue of mine. We have a way of thinking about whatever it is that our shirts have in common, because we know its identity conditions, viz. it is the hue of something that hue-matches my shirt, or hue-matches your shirt. That is, something is identical with this particular hue, the hue that I can pick out with the phrase 'the hue of my shirt', if and only if it is the hue of something that is hue-matching with my shirt.

Furthermore, there will be a sortal of which this-bluish-hue and that-reddish-hue (and so on) are exemplars. By getting onto one of the hues, e.g. this-blueish-hue, we can recognize the identity conditions for members of the hue class in general. We have what we might call the hue operator; this operator ('the hue of ...') picks out the hue of any given physical object, be it this-blueish-hue or that-reddish-hue.

Each thing picked out through the hue operator falls under a sortal concept, *hue*. We even have proper identity conditions for the things that fall under the sortal concept of hue, that is, the things picked out by the hue operator: the hue of physical object x = the hue of physical object y if and only if x and y hue-match. Since we have and understand the identity conditions for the things—the *objects*—of the sort *hue*, we have a way of thinking about objects of that sort.

Now, let's return to the implicit definition issue. We have here another application of an abstraction principle. That is *Hue Equivalence*:

$$\begin{aligned} &\text{The hue of physical object } c = \text{the hue of physical object } d \\ &\iff \text{physical objects } c \text{ and } d \text{ hue-match.} \end{aligned} \tag{3.5}$$

We can access the relevant equivalence between physical objects, like your shirt and mine, using our ordinary and immediate faculties—in this case, visual perception. When we stipulate Hue Equivalence, we use our stipulation to implicitly define something, viz. the hue operator—the operator that takes us from an object c to 'the hue of c'.[20] Hue Equivalence thereby implicitly defines the sortal *hue*, just as Hale and Wright tell us Direction

[20]There is something of an issue about whether 'the hue of c' is an object or a concept. This is related to Frege's comment that, "The concept *horse* is not a concept," found in his paper "On Concept and Object" in Geach and Black.

Equivalence implicitly defines the sortal *direction*.[21]

They explain:

> The *import* of the stipulation of the equivalence is simply that corresponding instances of the left and right sides—matching sentences of the shapes 'the direction of line a = the direction of line b' and 'lines a and b are parallel'—are to be alike in truth-value, i.e. materially equivalent. But because the stipulation is put forward as an explanation, its *effect* is to confer upon the statements of direction-identity the *same truth-conditions* as those of corresponding statements of line-parallelism.[22]

In other words, the stipulation of an abstraction principle—such as Hume's Principle, Direction Equivalence, or Hue Equivalence—only states that the equivalence relations and associated identity statements are materially equivalent. But the stipulation is also the manner of introduction of these phrases—'the number of ...', 'the direction of ...', or 'the hue of ...'. Accordingly, the stipulation sets the truth-conditions of the relevant identity statements. It doesn't just happen to be the case that Hume's Principle holds. Hume's Principle *determines* what it is for two numbers to be identical—the concepts that those numbers belong to, or the sets that have that number, are equinumerous. The stipulation effectively defines the meaning of the relevant number-identity claims, and thereby makes its content—the content of Hume's Principle—true in virtue of meaning.

According to Hale and Wright, the stipulation of Hume's Principle resolves the interaction problems. Since it is a stipulated truth that introduces a new linguistic expression ('the number of ...'), Hume's Principle is true in virtue of meaning, and we can know it

[21] Notice, though, that we don't give identity conditions for the sortal *hue* using this. We give a schema for identity conditions of members of the class *hue*. We give identity conditions for this-blueish-hue and that-reddish-hue and whatnot. We give conditions that tell us whether this hue is that hue, and so on. (What conditions? They are hue-matching.) But we don't do the second-order thing and give identity conditions for *hue*. We don't give identity conditions for being this *sortal* or that. We simply give conditions for being an *instance* of the hue sort—conditions for being some particular hue.

[22] All italics are in the original.

without having antecedent knowledge of numbers. Since it is a biconditional, and we assume that the right-hand side of the biconditional has an antecedent meaning that we already understand, Hume's Principle provides us with truth-conditions for statements of the form of (1). By providing us with truth-conditions for such statements, it gives us the meaning of such statements. And since we use the usual notion of identity in these statements, we can extract from our understanding of these statements the meaning of the previously unknown and undefined expression 'the number of ...'. In doing so, we gain the ability to individuate numbers, which is a kind of identifying knowledge of numbers. In gaining identifying knowledge of numbers, we gain the ability to think about numbers, and Hume's Principle ensures that we are able to do so in a truth-tracking way.

3.3 Identifying Knowledge

According to Hale and Wright, our stipulation of Hume's Principle determines which objects get picked out when we apply the operator 'the number of ...' to various Fs and Gs. In particular, the operator maps concepts (or collections) like 'the Fs' or 'the Gs' onto number-objects, mapping two concepts onto the same number-object if and only if the concepts are equinumerous. This information about the operator is enough to provide us with information about the individuation of numbers. Using Hume's Principle, then, we are able to come to understand what it is to be an instance of the sort 'number'. Once we understand the number operator, we are able to think about numbers in general. Once we apply the number operator to a particular F (e.g. fingers on my left hand), we are able to think about a particular number (e.g. 5).

3.3.1 Fixing Reference

On the neo-Fregean account, stipulating Hume's Principle fixes reference for number-phrases like 'the number of fingers on my left hand'. Given that we are attempting to resolve the interaction problems, we can assume that there are entities (viz. the natural numbers) satisfying the conditions laid out by Hume's Principle. But the existence of entities satisfying given conditions is not sufficient to establish reference; a denoting phrase fails to refer if it fails to distinguish between candidate denotations. Hume's Principle can only establish the reference of the phrase 'the number of fingers on my left hand' if there is an entity that uniquely satisfy the conditions that Hume's Principle lays out for this particular F, or if one of the entities satisfying the conditions laid out by Hume's Principle is privileged in some way (e.g. by intention or salience) that ensures that it becomes the referent of this number-phrase.[23] If there is no unique or privileged schema according to which the number operator assigns particular number-objects to concepts, then Hume's Principle does not succeed in fixing reference, and hence does not resolve the interaction problems.

Hume's Principle simply requires that, given an equivalence class of equinumerous collections, the number operator ('the number of ...') picks out the same object for each member of that equivalence class. But these conditions are not terribly strong, and many distinct operators could satisfy them. It would be consistent with these conditions, for example, for an operator to map all and only the members of some particular equivalence class of equinumerous collections to Julius Caesar. Intuitively, of course, such an operator would not be the number operator; Julius Caesar is not a number, and the number operator maps collections to numbers. This is a worry that bothered Frege. He thought that, from the proper definition of number, we should be able to determine that Julius Caesar is not a number. Since he thought that Hume's Principle was insufficient for this purpose, he ultimately rejected a definition of number from Hume's Principle.

Hale and Wright, however, are not swayed by the Julius Caesar problem; though they take it seriously, they think it is surmountable. Their proposed resolution centers

[23]For example, I can use the definite description 'the pen on the table' to refer to a particular pen sitting before me. Though many entities satisfy the description, somehow the pen before me is privileged in a way that makes it the referent of my use of the definite description, perhaps by my intention to refer to it, or perhaps because this pen is the most salient one in my current physical environment.

on an appeal to sortals and to the fact that Hume's Principle specifically provides identity conditions for numbers; it does not provide identity conditions that apply to things like lines, hues, or people.[24] Accordingly, Hume's Principle—like Direction Equivalence—gives sortal identity conditions. It expresses the conditions of individuation for objects of a certain sort, viz. number-objects.

Furthermore, Hale and Wright think, Julius Caesar is not a number because there is no overlap between the number sort and the human sort. They justify this claim through a more general principle that they call 'Sortal Inclusion': an object x can only be an instance of two different sortals (e.g. *human* and *animal* or *human* and *number*) if the identity conditions for that object can be the same regardless of the sortal under which it is considered. Thus, x can be an instance of both *human* and *animal* because the conditions that make human x identical with human y are the same conditions that make x identical with animal z. Similarly, if the human x is to be identical with the number w, then the conditions under which x is the same human as y must be the same conditions under which x is identical with the number w. But the identity condition for the latter is equinumerocity, and equinumerocity is not the appropriate kind of identity condition for human $x =$ human y. Thus, humans and numbers are not overlapping sortals, and Julius Caesar is not a number.

Even if we grant their account of why Julius Caesar is not a number, however, the Julius Caesar problem can be generalized in a way that Hale and Wright seem to lack the means to resolve. In light of their Sortal Inclusion principle, let us restrict the candidate numbers to those entities whose conditions of individuation could plausibly thought to be cashed out in terms of equinumerocity. One such view is Frege's. According to his view, n is a number if and only if, for some concept F, n is the extension of the concept *equal to the concept F*. What it is for a concept G to be equal to the concept F is for the Fs and Gs to be equinumerous. So, on Frege's view, identity conditions for numbers (considered as

[24] This proposal is found in the paper "To Bury Caesar . . . ," which appears in their book *The Reason's Proper Study*.

the extensions of certain kinds of concepts) can be explained in terms of conditions about equinumerocity. His analysis of number, according to which numbers are entities of the concept-extension sort, satisfies the restrictions imposed by Sortal Inclusion.

But Frege's view is not the only reductive analysis of number that satisfies Sortal Inclusion. Van Neumann's set-theoretic reduction of number does too. On Van Neumann's reduction, the numbers are a class of certain kinds of sets: any given natural number is the set of all of the smaller natural numbers (and zero is the empty set). The sets of this class are such that sets x and y are equinumerous if and only if they have the same members. And according to the individuation conditions for sets given by the Axiom of Extension, sets x and y are the same set if and only if they have the same members. Accordingly, for the class of sets to which Van Neumann reduces the numbers, the individuation conditions for sets are tantamount to the equinumerocity conditions that individuate numbers. Thus, it appears to be consistent with Sortal Inclusion that these set-objects be identical with the number-objects.

Van Neumann's reduction is not even the only set-theoretic reduction of number that appears to be consistent with Sortal Inclusion. Zermelo's set-theoretic reduction of number also seems to work. According to Zermelo's reduction, each number is the set whose sole member is its immediate predecessor (again, zero is the empty set). Again, two sets are identical if they have the same members. In the case of the class of sets to which Zermelo reduces the numbers, two sets have the same members exactly when they have the same *transitive closure*. The transitive closure of a set x is the smallest set that contains all the members of x and also satisfies an additional condition: if y is a member of the set, and z is a member of y, then z is a member of the set. In other words, the transitive closure of a set contains all of the set's members, and all of the members of those sets, and so on. In the case of the class of sets to which Zermelo reduces the numbers, because each number-set's sole member is its immediate predecessor, the transitive closure of a number-set turns out to be the set of its predecessors. So, using reasoning similar to that used in the Van Neumann case, two sets of the class to which Zermelo reduces the numbers are the same set if and only if their transitive closures are equinumerous. The individuation conditions for sets of the relevant type can be expressed in terms of equinumerocity, and so Zermelo's set-theoretic reduction of number seems to be consistent with Sortal Inclusion.

Given that Sortal Inclusion is consistent with Frege's, Van Neumann's, and Zermelo's respective reductions of number, there are multiple kinds of candidate entities to which the number operator might map.[25] It is not clear how we would decide which potential number operator is *the* number operator. Hale and Wright do not offer a method, and different philosophers presumably would want to give different methods, given their different accounts of number. This is a serious problem, and closely related to the topic of another paper of Benacerraf's, "What Numbers Could Not Be." But it seems perfectly plausible that there is one number operator that maps each F onto one of Frege's preferred entities, another that maps each F onto one of Van Neumann's preferred entities, and yet a third that maps each F onto one of Zermelo's preferred entities. All of these candidates satisfy both Hume's Principle and Sortal Inclusion. If we want to claim that there is only one number operator, we either have to privilege one of these number operators over the others (and there seems no reason to do so), or we have to claim that not all of these entities exist. But if we claim that some of these entities do not exist, we must give up the existence of at least one of the set-theoretic options. This is a bullet that most platonists about mathematics would be loath to bite; it does not provide a satisfactory resolution to the interaction problems.

If Hume's Principle is to resolve the interaction problems, it must give us identifying knowledge of numbers. That is, it must give us *uniquely* identifying knowledge of numbers. If there are several candidate classes of number-objects that satisfy Hume's Principle, and if additional principles like Sortal Inclusion do not distinguish one preferred class, then

[25]The Axiom of Extension ensures that the objects of Van Neumann's reduction are not the same as the objects of Zermelo's reduction (except for zero). I am making the substantive, but I believe unobjectionable, assumption that the objects of Frege's reduction are not identical with the objects of either type of set-theoretic reduction.

Hume's Principle cannot provide uniquely identifying knowledge of the numbers. Hence, Hume's Principle cannot resolve the interaction problems.

3.3.2 Distinguishing Numbers

Suppose, however, that additional metaphysical principles could be introduced to ensure that there is a unique sortal whose instances have the individuation conditions set out by Hume's Principle. Then stipulating Hume's Principle as an implicit definition does seem to uniquely pick out the entities of that sort; those entities are the numbers. Even under such auspicious circumstances, I contest, Hume's Principle would not suffice to fix reference for phrases of the form 'the number of Fs'. Hale and Wright's explanation of how to use Hume's Principle to resolve the interaction problems relies on the assumption that knowledge of individuation conditions constitutes identifying knowledge of the things so individuated. I claim that knowing what individuates one thing of a sort from another thing of that same sort does not, on its own, position us to have identifying knowledge of any particular thing of that sort. Even if Hume's Principle provides conditions for individuating numbers from each other, it does not provide uniquely identifying knowledge of any particular numbers. The latter is required to fix the reference of a phrase of the form 'the number of Fs'.

Suppose we accept whatever metaphysical suppositions are required for the individuation conditions found in Hume's Principle to uniquely apply to a certain class of entities. Then the objects of this class are objects of the number-sort. Consider, then, the phrase 'the number of fingers on my left hand'. Since this phrase takes the proper form, it picks out one of the number-objects. It picks out the same number-object that the phrase 'the number of fingers of my right hand' picks out. This fact is clear from Hume's Principle and the fact that the fingers on my left hand are equinumerous with the fingers on my right hand.

But *which* of the entities in the class of number-objects is picked out by these phrases? It is not sufficient to use the stipulation to get onto the right *sortal*; we must be able to identify some *particular* object of the relevant sort. If Hume's Principle is to fix reference for ‘the number of fingers on my left hand’, it must either identify one particular number-object as the unique number satisfying the relevant conditions, or it must somehow privilege one particular number as the one that the phrase is to pick out. It does not suffice to know what sort of objects numbers are and what sorts of things the fingers on my left hand are. In order to fix the referent of the phrase ‘the number of fingers on my left hand’, we must know how numbers are related to concepts like ‘the fingers on my left hand’.

An analogy might help. We might know individuation conditions for people (e.g. spatio-temporal continuity), and we might know what sorts of things constitute wealth (money, things that can be sold for money, etc).[26] Nonetheless, we might not know which person is the wealthiest in the room. I am not merely suggesting that we might not know the wealthiest person by name or by a different description. In order to know who the wealthiest person in the room is—under that very description—we need to know what it is for a person to be wealthy. It does not suffice to know what a person is and what wealth is; we also need to know something about what it is for a person to *have* wealth.

Similarly, it does not suffice for us to know conditions of individuation for number, and to know the concept F. If Hume’s Principle is to fix the referents of statements of the form ‘the number of Fs’ by providing identifying knowledge, it must provide knowledge about the relation between numbers and concepts—what it is for a number to be the number *of* some concept F. Without doing this, Hume’s Principle cannot fix any particular referent for the phrase ‘the number of fingers on my left hand’. As far as Hume’s Principle is concerned, the object we call ‘two’ is just as well qualified to be the referent of this phrase as is the number we call ‘five’, so long as it is also the referent of ‘the number of Fs’ for all and only the Fs that are equinumerous with the fingers on my left hand.

Even if Hume’s Principle fixes which entities are the numbers, it is ill-equipped to fix which of those entities is the number of fingers on my left hand. Hume’s Principle seems to

[26]Individuation conditions for people are obviously controversial, and I do not intend to take a side on that debate. The controversy should have no bearing on the point I am trying to make.

be consistent with the phrase 'the number of fingers on my left hand' picking out what is intuitively the wrong number. If it is to fix reference, then there must be other factors at play that somehow privilege one particular number-entity to be the referent of that phrase. One possibility would be that some other metaphysical principle leaves some particular entity privileged. If so, Hale and Wright must adequately address that metaphysical principle in order to resolve the interaction problems. Knowledge of individuation conditions does not on its own constitute the kind of identifying knowledge that can be used to fix reference.

It might seem like there are some intuitive solutions. For example, if Frege's preferred class of entities (extensions of concepts of the form 'equal to the concept F') are the numbers, then it seems obvious to fix the referent of 'the number of fingers on my left hand' as the extension of the concept 'equal to the concept *fingers on my left hand*'. If Van Neumann's preferred entities (sets containing all of a number's predecessors) are the numbers, then it seems obvious to fix the referent of 'the number of fingers on my left hand' as the set of the relevant type that is equinumerous with the numbers on my left hand. And if Zermelo's preferred entities (sets containing a number's immediate predecessor) are the numbers, then it seems obvious to fix the referent of 'the number of fingers on my left hand' as the set whose predecessors are (or, what amounts to the same, whose transitive closure is) equinumerous with the fingers on my left hand. While Hume's Principle does not itself privilege any of these assignments, they nonetheless seem to be the natural ones.

While these assignments might seem intuitive, there are reasons to hesitate before assuming that we can use them to fix the reference of phrases of the relevant form. First, if further relevant metaphysical principles are going to be strong enough to pick out a unique class of entities, they must be strong enough to ensure that all or almost all reductive analyses of number do not in fact yield number objects; the number operator does not map to the entities described by these reductions. But it seems likely that principles strong enough to eliminate *almost* all reductions ultimately will be strong enough to eliminate *all* reductions. So, it seems likely that, if further metaphysical principles ensure that Hume's Principle gives the individuation conditions for a unique class of entities, then the numbers will be *sui generis*—that they will not be reducible to things like sets or extensions of classes.[27]

Second, the assignments that seem to be intuitive seem to emerge from the specific reductive analyses under consideration, and if number cannot be reduced, then there does not seem to be an intuitive assignment system. The most intuitive method of assignment outside of a reductive analysis would seem to rely on something like assigning the referent of 'the number of Fs' as the number whose predecessors are equinumerous with the Fs. But to do this would be to presuppose knowledge of the structure of the number sequence. That is, it would presuppose knowledge of *ordinal* number—first, second, third, fourth, etc.—in order to explain knowledge of *cardinal* number—singleton, pair, threesome, foursome, etc.[28] Our initial ability to think about number, then, would be an ability to think about ordinal number. Our stipulation of Hume's Principle would not resolve the interaction problems. Even presupposing additional (and presumably very substantive) metaphysical principles that ensure that Hume's Principle uniquely picks out a class of number-objects, our stipulation seems like it would only help us to access the cardinal numbers if we already knew the ordinals. We will return to this point.

3.4 The Ancients

I have argued that Hume's Principle does not suffice to fix reference for phrases of the form 'the number of Fs'. But even if it did suffice, I will argue that Hale and Wright's attempt to resolve the interaction problems could not explain how we actually do come to think about numbers. There can be no doubt that ancient mathematicians knew arithmetical truths. They made the same basic arithmetical claims that we make today, albeit in different

[27]I make this point with some hesitance, primarily because it strikes me as unlikely that metaphysical principles could bring us from the conditions set out in Hume's Principle to a unique class of number-objects without presupposing knowledge of numbers. It is difficult to reason about things that strike one as impossible.

[28]Notice also that Zermelo and Van Neumann provide set-theoretic analyses of ordinal number, while Hume's Principle gives individuation conditions for cardinal number.

languages. Those claims are true today, and they were true thousands of years ago when expressed by the ancients. If knowing those mathematical truths requires the ability to think about abstract mathematical objects, then ancient mathematicians were able to think about mathematical objects, and refer to them, and know truths about them. The solution Hale and Wright provide for the interaction problems—their account of our ability to refer to such objects and know truths about them—must be compatible with this fact. I will argue that it is not.

3.4.1 The Possibility of an Alternative Method of Engagement

Suppose that my arguments in the previous sections are wrong, and that Hume's Principle can fix reference for phrases of the form 'the number of Fs'. Suppose it is possible to use Hume's Principle to come to think about the natural numbers. Then, either (a) the ancients had Hume's Principle and used it to think about the numbers, or (b) they came to think about the same number objects through some other route.

Let us consider (b) first. Suppose the ancients initially accessed arithmetical objects using some means other than the stipulation of Hume's Principle. Like Hume's Principle, this alternative method allows people to think about number objects, and it gives rise to arithmetic.

Supposing, then, that there are two different methods of accessing the subject matter of arithmetic, do these two methods of engagement give us access to the *very same* number-objects? Without a particular alternative method of engagement (and possibly more information about the objects picked out by Hume's Principle), we might not be able to answer this question. Unless the alternative method also gives us access to numbers in the form 'the number of Fs', we will be unable to determine whether or not the two methods yield the same objects. The only identity statements that the neo-Fregean route positions us to evaluate are those that occur in the form found in Hume's Principle.[29] We cannot, for

[29] Perhaps the alternative method would give us insight into another form of identity statement that would

example, take the name of a (putative) number, e.g. '5', that we access using the alternative method and substitute in for it something of the right form, perhaps 'the number of fingers on my left hand'. To do so would be to assume that the object named '5' accessed using the alternative method is the same as the object named '5' accessed using Hume's Principle. Most likely, we would lack the tools to determine whether or not the entities picked out by Hume's Principle were the same picked out using the alternative method.

Why, one might ask, does this matter? Prima facie, unless the two methods give access to the very same number objects, and not just similar ones, the arithmetic based on Hume's Principle will be a different arithmetic from that based on the alternative method. After all, those who use the alternative method for engaging with number objects would have knowledge about and refer to different objects from those who get onto number objects using Hume's Principle. Presumably the two engagement methods would give rise to two similar-looking arithmetics. But they would involve completely different truths; there would be two fundamentally different studies of arithmetic, studies that explored two entirely different subject matters. It would appear to be sheer coincidence if the class of numbers accessed through Hume's Principle and the class of numbers accessed through the alternative method have entities with similar properties and relations.

An initial response to this line of reasoning might be that two people can engage with the same arithmetical project even when they are referring to and discovering truths about different objects. After all, Zermelo and von Neumann did not seem to be engaged in two *fundamentally different* projects of arithmetic, even though their accounts of number were rather different. For Zermelo, $n+1 = \{n\}$; for von Neumann, $n+1 = n \cup \{n\}$. Nonetheless, they were doing the same arithmetic. The common-looking arithmetical equalities and other results that the Zermelo and von Neumann numbers yielded (e.g. 2+3=5) were not accidentally similar. The fact that the two mathematicians were concerned with fundamentally

be of some use in resolving the identity questions here, but we are just operating with the possibility of such an account, not an example of one.

different number objects did not prevent them from being involved in the same arithmetical project.

Surely Zermelo and von Neumann were engaged in the same arithmetical project. But their situation does not appear to bear directly on the proposal under consideration. Zermelo and von Neumann both were already familiar with the numbers before giving their set-theoretic accounts of those numbers. Both were trying to ground an already-existing arithmetic;[30] they were trying to understand a common subject-matter, and the result was two different substantive accounts of the set-theoretic composition of numbers. This is analogous to a case in which two biologists examine the same animal and come up with two different substantive accounts of the internal structure of that animal. Zermelo and von Neumann were engaged in the same arithmetical project because they were trying to describe a common subject-matter. People were familiar with the numbers long before they were familiar with any set-theoretic underpinnings that arithmetic may have.[31]

Zermelo and von Neumann, though, were already engaging with the same numbers *before* they theorized about the internal structure of those numbers. The assumption of prior engagement with the numbers is not available in the kind of scenario we are trying to consider: the case in which there are two different ways of *initially* engaging with number-objects. We are considering the case in which one person's initial (and continued) engagement with numbers comes through Hume's Principle, and another person's initial (and continued) engagement with numbers comes through some other means, and these two methods yield engagement with two different classes of objects. So, the fact that Zermelo and von Neumann were working on the same arithmetic does not bear on the question of whether or not the two people interacting with fundamentally different entities could nonetheless do the same arithmetic. It seems that they could not.

[30] I would claim that Frege was also engaged in the same project of grounding an already-existing arithmetic. Unlike Wright and Hale, he was not using Hume's Principle to account for our initial interaction with numbers.

[31] I do not mean this to be an endorsement of any attempts to reduce arithmetic to set-theory.

Perhaps another kind of mathematical consideration will prove more fruitful in this discussion: the distinction between ordinal and cardinal numbers. Consider the number 4. It could be interpreted as a cardinal number—a size of collections, the size that all foursomes have in common. Or, it could be interpreted as an ordinal—a place in an order, following the third and preceding the fifth. Since they take equinumerocity to give truth conditions for numerical identity, and thereby account for our ability to engage with number, Hale and Wright must be concerned with cardinal number, not ordinal number.[32] But perhaps there could be a method that gave access to ordinals rather than cardinals. If so, then this method would likely get us onto different objects than would Hume's Principle. After all, there is no *ordinal* number of fingers on my left hand, the way there is a cardinal number of fingers on my left hand (five). And when the ordinal number of my finish in the race is third, the cardinal number is, if anything, one—only one event fits the description of being my finish.

But even though ordinal numbers and cardinal numbers are different, they yield the same arithmetic.[33] And it is no coincidence that they do; ordinal and cardinal numbers are closely related. There is a natural one-one correspondence between the cardinals and the ordinals; one gets mapped to first, two to second, three to third, and so on. And there are straightforward ways of defining this natural correspondence from the ordinals to the cardinals, or vice versa. We can create an ordering of cardinal numbers according to whether or not there is an injection from a concept that has one cardinal into a concept that has another cardinal.[34][35] And we can always map the ordinals onto the cardinals using the operator 'the (cardinal) number of (ordinal) numbers (including zero) less than x'.

Given this natural bijection between the ordinal and cardinal numbers, it is no ac-

[32] Recall that Zermelo and von Neumann gave accounts of ordinal number.

[33] At least, they yield the same arithmetic in the finite case. They do not yield the same arithmetic in the infinite case.

[34] That is, if n is the (cardinal) number of the Fs, and m is the (cardinal) number of the Gs, then the number of Fs is greater than the number of Gs if and only if there is an onto mapping of the Fs into the G but no onto mapping of the Gs into the Fs.

[35] Cf. Frege's version in *Foundations* §76.

cident that the arithmetical structure of the ordinals is the same as that of the cardinals. Yes, the person who gets onto this common arithmetical structure through the ordinals will be thinking about different numbers from the person who gets onto it through the cardinals (say, through Hume's Principle). But although a person may be stating a truth about the ordinals, there will be a corresponding truth—a dual truth, if you will—about the cardinals.

Thus, if there are two methods of initially engaging with number-objects, and one method provides access to the ordinals while another (Hume's Principle) provides access to the cardinals, it seems that a person who engages with number-objects through the ordinal route is perfectly capable of engaging in the same arithmetical project as the person who engages with number-objects through the cardinal route. So if there were a method capable of establishing engagement with ordinal numbers the way that Hume's Principle establishes engagement with cardinal number, and the ancient mathematicians used this alternative ordinal method to establish interaction with number-objects, then it seems that it will nonetheless be the case that the ancient mathematicians were making the same basic arithmetical claims that we make today. At worst, they were making corresponding claims, claims whose duals are still made today.

3.4.2 Continuity of Object-Directed Thought

For all that has been said thus far, it seems perfectly possible that the ancients' initial engagement with number-objects was an engagement with one sort of object, and today the typical initial engagement is with a different sort of object. But this might not be the right approach.

There is continuity between our arithmetic and the arithmetic of the ancients. We learned much of our arithmetic from them. Presumably it is not the case that the ancients discovered arithmetic through one means, and we discovered it completely independently, using a completely different route. No, rather, the knowledge of basic arithmetic has traveled down through the linguistic and cultural community, much as knowledge of Julius Caesar has. And much as the ability to speak of Julius Caesar gets communicated to those who have not personally interacted with him, talk of numbers can get communicated through the community. So if someone gained the ability to think about numbers at some point long ago, and passed knowledge about and reference to number objects down to us through communication chains, then we too have the ability to think about those same number-objects, thanks to the help of our teachers.

If the ancients used one means of engaging with the numbers, and we initially learned about arithmetic from communication chains that connect to them, then we initially came to think about the same number-objects through those ancients and so through their route of interaction. If they initially accessed numbers using some method other than Hume's Principle, than so did we. Contrapositively, if our initial access to the numbers comes through Hume's Principle, then so did theirs. If Hale and Wright's solution to the interaction problem is the method that we actually use to initially come to think about the numbers, then it must be the case that the ancients also had Hume's Principle.[36] Furthermore, it must be the case that the ancients relied upon Hume's Principle to think about number objects.

Granted, it seems unlikely that our intellectual predecessors in ancient times had Hume's Principle in explicit form. To state the principle without presupposing the ability to refer to number, in particular in the parts that involve '*one-one* correspondence' (or 'equi*numerocity*'), the principle must be stated using polyadic logic (that is, nested quantifiers). Frege was the first to develop polyadic logic, and his doing so was a significant acheivement.[37] Ancient philosophers could not have formulated Hume's Principle in a logically rigorous way.

[36] At least, the ancients who were our intellectual predecessors must have had Hume's Principle. It seems possible that the Greeks and the Mayans, for example, had different origins for their arithmetical thought. Perhaps the Greeks got onto number using ordinality while the Mayans got onto number using cardinality, given that the Mayans had zero and the Greeks did not.

[37] Ancient philosophers did not have the logical resources to nest quantifiers and so could not have explicitly formulated Hume's Principle. But I might be shortchanging medieval logicians in making this claim. Their logic was considerably more sophisticated than we usually credit it with being. But nonetheless, medieval logicians could not have formulated Hume's Principle because they did not have quantification over relations.

Most likely, if the ancients who thought about arithmetic accessed numbers through Hume's Principle, then they had Hume's Principle implicitly. They used it in some abstract, logically imprecise form.[38] But they were not consciously aware of the fact that they were using it, and probably did not attempt to formulate it. Nonetheless, their subconscious use of Hume's Principle allowed them to latch onto these abstract number-objects, and gave the means necessary to refer to these objects and have knowledge about them.

3.4.3 The Very Idea of Implicit or Innate Stipulation

The fact that Hume's Principle is a stipulation does much work in the account Hale and Wright give to resolve the interaction problems. Since it is a stipulation, we need not justify our claim that number-identity claims have the particular truth-conditions it sets out. Our stipulation makes it so; our stipulation introduces those truth-conditions. Hume's Principle determines *what it is* for two numbers to be identical. If the principle merely *reported* the truth-conditions of identity statements, relying on it would be unjustified without some other knowledge of number. But because Hume's Principle is a stipulation and *establishes* those truth-conditions, relying on it is perfectly acceptable.

Stipulations are declarations that introduce truths. But in the previous section, we suggested that Hume's Principle was known implicitly, or innately. Somehow the ancients subconsciously knew and applied this truth. But if Hume's Principle is *innate*, then it cannot be a *stipulation* that serves to implicitly define number, as Hale and Wright claim. Stipulations cannot introduce innately known truths. Stipulated truths are not even true before they are stipulated; they are not knowable before they are stipulated. The very idea of an innately known stipulation is incoherent.

If it is possible to have Hume's Principle innately, then Hume's Principle is a so to speak non-manufactured truth. If we are to know it innately, we must be able to recognize

[38] I do not mean to say that their version of Hume's Principle was logically imprecise in a pejorative sense. I just mean that they did not use, even subconsciously, any particular logical formation of the principle. The ancients used a pre-theoretic version.

its truth. But if we are to be able to recognize something as true, then we must already understand the content of that truth. If we are to innately know Hume's Principle, we must have some sort of innate access to number.

This leaves us in the following position. If we are to accept Hume's Principle as the way that we are able to engage with non-intuitable, non-perceivable number objects, then we need Hume's Principle to be a stipulation. If we cannot otherwise interact with number objects, then no justification is available for this truth, so we cannot know it or rely on it in building our theory. But if we do in fact use Hume's Principle to initiate our interaction with number-objects, then we also seem to be committed to thinking that the ancients used it in the same way, to give them access to number-objects. And the most reasonable way to suggest that the ancients had Hume's Principle is to suggest that they had it implicitly, or innately. Something about the idea of the ancients going about stipulating Hume's Principle seems strange. So we are torn, then, between Hume's Principle being a stipulation and its being innate. It cannot be both.

3.5 Final Thoughts

Hale and Wright seem to want to have mathematics both ways, so to speak. They want mathematical platonism to be true; they want our mathematics to be a science of a special kind of abstract entities. But they also want us to be able to interact with those entities without relying on any rather mysterious faculty of intuition. Ultimately they rely on a stipulation of Hume's Principle in order to resolve the interaction problems and make sense of holding these two positions simultaneously. I have argued that this solution is not tenable. But even setting aside my arguments, it should seem implausible to us that we come to think of numbers through Hume's Principle for the same reasons that it should seem intuitively implausible to us that we come to think about the colors of objects through Hue Equivalence. Instead, we *recognize* Hume's Principle—and likewise Hue Equivalence—to be true. Before we encounter these principles, we already know quite a lot about the putative abstract objects they involve.

Chapter 4

2+2=4: A Basic Arithmetical Proof

In the previous two chapters, we considered two candidate methods of using definitions to ground our mathematical knowledge. Both methods involved implicit definitions. In Chapter 2, we saw that axiomatic definitions could not ground all of our knowledge and understanding of arithmetic. In Chapter 3, we saw that an implicit definition could not ground our initial cognition of the objects of arithmetic if those objects (the natural numbers) are mathematical abstracta.

In this chapter, we will step away from implicit definitions and consider the role of explicit definitions in arithmetic. I will argue that we could not use explicit definitions to gain secure knowledge of basic arithmetical truths unless we already had a rather significant antecedent understanding of the structure of the natural numbers. Accordingly, we cannot ground our understanding of the structure of arithmetic in explicit arithmetical definitions any better than we can ground our understanding of the structure of arithmetic in implicit definitions.

4.1 Leibniz's Proof

In *New Essays Concerning Human Understanding*, Leibniz gives a seemingly solid proof of the arithmetical truth that $2 + 2 = 4$. Explicit definitions of the form '$m = n + 1$' play a central role in this proof; one of the proof's key definitions, for example, is '$4 = 3 + 1$'. In addition to these definitions, the proof relies on two fundamental assumptions. One, Leibniz notes, is that the substitution of equals preserves equality in arithmetical equations. Another, Frege notes in *The Foundations of Arithmetic*, is that the proof requires the associativity of addition. Any viable account of mathematical truth and knowledge ought to preserve both of these assumptions.

Leibniz's proof that $2 + 2 = 4$ goes as follows:

> It is not an immediate truth that 2 and 2 are 4; provided it be granted that 4 signifies 3 and 1. It is can be proved, as follows:
>
> Definitions:
>
> 1. 2 is 1 and 1
> 2. 3 is 2 and 1
> 3. 4 is 3 and 1
>
> Axiom: If equals be added to equals, the equality remains.
>
> Proof: $2 + 2 = 2 + 1 + 1$ (by Def. 1) $= 3 + 1$ (by Def. 2) $= 4$ (by Def. 3)
>
> $\therefore 2 + 2 = 4$ (by the Axiom).[1]

Frege quotes this passage in *The Foundations of Arithmetic* §6, and then goes on to observe that Leibniz's proof depends on an additional assumption that Leibniz fails to make explicit: the associativity of addition ($(a + b) + c = a + (b + c)$). But aside from suggesting that this assumption be made explicit, Frege praises Leibniz's approach to arithmetical proof:

> If we assume this law [associativity], it is easy to see that a similar proof can be given for every formula of addition. Every number, that means, is to be

[1]There are several minor variations in the translation here. Frege's quotation, and Austin's translation of it, uses numerals here: "4 is 3 and 1" ("4 ist 3 und 1"). Leibniz's original French uses number-words: "four is three and one" ("quatre est trois et un"). Langley's translation of Leibniz's *New Essays* italicizes one of those number-words: "*four* is three and one." In addition, the formatting of the proof is different in all three versions. I have used Frege's formatting here.

defined in terms of its predecessor. I do not see how a number like 437986 could be given to us more aptly than in the way that Leibniz does it. (§6)

The key points are that (a) Leibniz's approach offers an account of each of the natural numbers in terms of its predecessors; (b) Frege thinks that the explicit definitions in such proofs provide an apt explanation of how the numbers are given to us—that is, how we cognitively access the numbers; and (c) this approach gives us a straightforward method with which to prove arithmetical truths. Taking these points together, Frege seems to think that the kind of approach we see in Leibniz's proof exhibits a plausible account of how we come to think about numbers (at least, numbers larger than 1) and also of how we can securely establish (prove) arithmetical truths.

A variation of this approach has become standard in axiomatic accounts of arithmetic. Rather than starting from the number 1 and using Leibniz's '$n + 1$' operation to get the other natural numbers, modern mathematicians tend to start from 0 (zero) and get the other natural numbers using the successor operation ($S(n)$). While Leibniz's approach to generating the numbers starts at one and modern axiomatic arithmetic tends to start at zero, they are fundamentally minor variations of the same approach. For the purposes of this paper, we shall go with Leibniz, and assume that the process begins at 1. Nothing should be lost by using this convention.

4.2 Leibniz's Definitions

For at least two reasons, it is worth examining what cognitive resources are required to understand Leibniz's proof. First, if we are to take seriously Frege's apparent suggestion that Leibniz's explicit definitions give us cognitive access to numbers (at least the fairly large ones), then we should ensure that these definitions do not presuppose any antecedent cognitive access to those numbers. Second, if the type of reasoning demonstrated in Leibniz's proof that $2 + 2 = 4$ is the type of reasoning that ultimately grounds our secure arithmetical knowledge, it is worth determining what cognitive resources that reasoning relies upon. That is, it is worth determining what resources we rely upon in acquiring our most secure arithmetical knowledge.

In order to determine what cognitive resources are required to understand Leibniz's proof, we must first investigate the nature of the definitions that the proof employs. What exactly are they definitions *of*? One possibility is that they are definitions of numerals (e.g. the symbol '6') or number-words (e.g. the word 'six'). A second possibility is that they are definitions of number-objects or number-concepts—objects or concepts that are neither symbols nor pieces of language.

It is fairly clear that Leibniz means to treat definitions (1)–(3) as definitions of terms, that is, definitions of numerals or of number-words. After all, he writes that "4 signifies 3 and 1." Usually symbols or words signify things or objects; non-linguistic objects like numbers—objects that have not been stipulated in the relevant context to have some particular symbolic meaning—do not signify anything at all. So, this statement suggests that Leibniz is interested in the signification of the symbol or word '4'.[2] In fact, he makes explicit the fact that he thinks of these as definitions of terms when he writes, two paragraphs before the selection excerpted by Frege:

> . . . the statement one and two is three is only the definition of the term three, so that to say that one and two is equal to three is to say that a thing is equal to itself. (Bk IV, Ch. VII, §10)[3]

Given our modern sensitivity to the use/mention distinction, a philosopher writing today would be better off putting it this way: " '4' signifies 3 and 1," and "the statement 'one and two is three' is only the definition of the term 'three'."[4]

[2]In Leibniz's original French, it is actually the written-out word 'four', rather than the numeral.

[3]Here I use Langley's translation, but I remove Langley's italics, which are not in the original French.

[4]The segments placed in single quotation marks here are the same segments italicized by Langley. There should be little doubt about the interpretation.

Frege appears to treat Leibniz's definitions a little differently. In describing Leibniz's approach, he writes that, "Every number is to be defined in terms of its predecessor." He uses the term 'number' here, and not the term 'numeral'.[5] It appears that Frege understands Leibniz's definitions (1)–(3) to be definitions of objects.[6] Frege might have just been speaking loosely here, and might have meant to suggest that the definitions were of terms (numerals) or of concepts. But let us consider the possibility suggested by Frege's wording—the possibility that these are definitions of number-objects.

4.2.1 Defining Number-Objects

I think the interpretation according to which number-objects are defined in terms of their predecessors makes it implausible that Leibiz's proof is a source of secure arithmetical knowledge. On such an interpretation, a definitional statement, e.g. '4 is 3 and 1', is a definition of the object (a particular number-object) picked out by the definiendum (the numeral '4'). Let us follow both Frege and Leibniz in treating '4 is 3 and 1' as equivalent to '$4 = 3+1$'; let us assume with them that the "=" symbol represents identity. Then, the definition states that this object, viz. the one picked out by '4', is identical with whatever gets picked out by the expression '$3 + 1$'. Since the number 4 is an object, that with which it is identical must also be an object. So, the expression '$3+1$' must also pick out an object; it must pick out the same object that '4' picks out.

If the definition '$4 = 3+1$' is a definition of the number-object picked out by '4' (and not a term, concept, or mode of presentation), then there are two possibilities regarding how this number-object is being defined. One possibility is that the definition merely defines this

[5] Frege uses 'Zahl' in the German.

[6] There is something a little intuitively strange, to me at least, about the idea of defining an object. But there is plenty of historical precedence for engaging in such a project. There is a significant literature on Aristotle's real and nominal definitions, and at least some of that literature interprets these definitions (or just the real definitions) to be definitions of objects. For example, on such an interpretation the real definition 'Man is rational animal' identifies the essence of the thing in question, viz. humans, or perhaps humanity. The definition provides no information about e.g. the term 'man' or its proper application; the term already has a meaning fixed. Rather, the definition aims to provide information about or an analysis of what the term picks out.

number-object as identical to another object, viz. the one to which the expression '$3+1$' refers. If so, the role of the definiens is to provide an *object*, rather than the term that picks it out or the mode of presentation through which it is picked out. Such a definition is independent of the name or mode of presentation through the definiens picks out the object. In such a case, the way in which *this thing*, viz. the relevant number-object, gets picked out should not affect its definition.

Accordingly, if the definition '$4 = 3+1$' defines the bare object picked out by '4' as identical to that picked out by '$3+1$', then the definition does one of two things: either it defines the number-object as identical to itself, or it is a substantive claim that is false because '4' and '$3+1$' refer to different objects. Even disregarding the clearly undesirable possibility that the definition is a false substantive claim, this ought not to be a satisfying definition; every object is self-identical, so a definition of something as identical with itself provides neither analysis of the object nor insight into it. Since this definition is an identity statement, it does not, on the surface, provide any information about the object.

The other possibility is that the expression used as the definiens does not merely pick out an object (as the expression used in the definiendum does), but provides further information as well, such as perhaps a mode of presentation. But in this case, the definition '$4 = 3+1$' provides no analysis or insight that '$3+1 = 3+1$' does not; in both cases, the definiendum merely picks out the relevant object. And if the definition provides no analysis or insight beyond what could be provided by an instance of self-identity, it could not play the substantive role it is required to play in proofs. The definition is not used merely as an instance of self-identity in the proof. If the definition is to be at all useful in proving arithmetical truths, it must be a definition that relies on the term (or perhaps the mode of presentation or the concept) used in the definiendum.[7] But then the definition no longer

[7] There are at least two notions of proof. One notion is that of *formal* proof, which involves the manipulation of statements using syntactical rules (e.g. modus ponens, or the substitution of equals). This is the standard notion of proof, and proofs of this sort rely on the forms of statements. So, for example, since '4' and '$3+1$' have distinct syntactical forms, the identity '$4 = 3+1$' potentially could play a different role in proofs of this kind than could the identity '$3+1 = 3+1$'.

defines the *object*, but rather it defines something that represents or picks out the object.

More generally, it seems to be fruitless to define particulars—that is, to define individual things. Definitions seem to define representations: words, or perhaps modes of presentation. Perhaps it is more sensible to define universal things—that is, kinds, like *human* or *triangle* (assuming that there are many triangles). To be honest, I remain dubious about this latter possibility, but that is a topic for another time. Of relevance to the issue under discussion is the fact that identity claims will not suffice to define number-objects. If definitions of the sort provided by Leibniz are to be useful in arithmetical proofs, those definitions cannot be definitions of objects.

4.2.2 Defining Numerals

Now that we have rejected the interpretation of definitions as defining number-objects, we turn to Leibniz's idea that the definition '4 is 3 and 1' defines the term (the numeral or number-word) '4'. Under this interpretation, the definition is non-trivial; it fixes the referent of the numeral '4' by giving it the same referent as the expression '$3+1$'. Once the referents of the numerals get fixed in this way, arithmetical truths can be proved using the method Leibniz models in his proof of '2 and 2 is 4'.[8]

This interpretation seems promising, provided that the definition succeeds in providing the numeral with a referent. The definition of the numeral '4', for example, succeeds if

The other notion of proof is not formal. Descartes suggests one such method of proof (deduction), according to which deduction involves a movement of the mind. (See his *Rules for the Direction of the Mind*, in *The Philosophical Writings of Descartes, Vol I*, edited by Cottingham, Stoothoff, and Murdoch.) But truths of the form '$3+1=3+1$'—statements of *self*-identity—play no substantive role in such deductions because no movement of the mind is required to notice that an object is self-identical, even if a movement of the mind is required to recognize that two distinct terms co-refer. While Descartes's notion of proof is rather tricky, it still seems to require movement of the mind between something like distinct modes of presentation of an object or perspectives on it, or having distinct means of representing a single object.

[8] In the French, Leibniz says, "two and two are four" ('sont') in one place; he says, "two and two is four" ('est') in another. The difference between singular and plural here does seem relevant, but it is not clear whether or not Libniz was attending to that difference. If he was, he meant to treat the number four as singular; the character representing his views, Theophilus, uses the singular form, while the character recommending Locke's views, Philalethes, uses the plural. I will stick to the singular in virtue of the fact that we are assuming that e.g. the number four is one object, rather than multiple objects.

and only if the expression '$3+1$' has a particular number-object as its referent. We must suppose, of course, that '1' has its referent fixed, as do all the other predecessors of the particular numeral getting defined. So, in this case, we assume '1', '2', and '3' all have referents fixed. But what about '$3+1$'?

4.2.3 Compositional Number-Expressions

There are a few ways that the referent of '$3+1$' might be fixed. One possibility is that the expression is primitive in a certain way; in particular, the referent of '$3+1$' is not dependent on the composition of '1' and '3' and '$a+b$' (the part the '+' contributes to the expression), but rather the expression acts as an indivisible unit and picks out the relevant number object immediately, so to speak. Of course, this would render the definition of '4' as '$3+1$' uninformative, at least in the sense that it would provide no information about the relation between 3 and 4. Further, it would introduce complications in arithmetical proofs of the style modeled by Leibniz's proof that $2+2=4$. The non-compositional expression would accomodate neither the valid application of Leibniz's axiom (substitution of equals) nor Frege's (associativity of addition). Just as we cannot validly perform substitution of equals on terms that occur as part of a name, we cannot substitute into '$(2+1)+1$' to get '$3+1$' if '$3+1$' is to be taken as a single, non-compositional term. A similar problem arises with the associativity of addition, assuming that '$(2+1)+1$' is taken to be a non-compositional term as well.

Accordingly, if we are to validate the uses of substitution and associativity in Leibniz's proof that $2+2=4$, we must take expressions of the form '$a+b$' to be compositional in nature. The referents of such expressions depend on their component terms. In particular, the success that such expressions have in fixing reference depends on the role of the symbol that allows the numerals to be joined together into a single term: the '+'.

4.3 The Meaning of '$a + b$'

We turn our attention, then, to the role of the '+' symbol in fixing the reference of expressions of the form '$a + b$'. One fairly intuitive idea is that the '+' sign, in conjunction with the open variables for numerals flanking it (i.e. '$a+b$'), represents an operation of some sort—an operation of *adding*.[9] But there is a prima facie implausibility about treating '$a + b$' (where a and b are unbound variables) as representing an operation. While perhaps 5 is the *result* of adding 2 and 3, 5 is not the actual *adding* of 2 and 3. The implausibility of the idea that '$a + b$' represents a binary operation performed on number-objects, accordingly, stems from our assumptions that numbers are objects and the '=' symbol represents identity. It is not enough for the composition '$2 + 3$' to *yield* an object or give instructions on how to get an object; the referent of the composition '$2 + 3$' must itself *be* an object.

It might be tempting to think that this is no more than a mathematical example of Frege's puzzle about identity. One might think that the challenge I raise to the idea of '$a+b$' representing an operation is the problem of one object (the number 5) getting picked out by two distinct senses (the sense associated with the expression '2+3' and that associated with the numeral '5'). Surely such a puzzle will arise for number-objects if it arises for other sorts of objects. The substantive side of such a challenge obviously would be far more general than the mathematical case, and would give us no reason to worry about arithmetical identities beyond the reasons we already have to worry about identity in general.

But Frege's puzzle does not apply here, because the number 5 is *not* identical with the operation of adding the numbers 2 and 3. The number 5 is not the performance of an action, while the adding of 2 and 3 surely is. Intuitively, the number 5 is an object, while the summing of 2 and 3 is an action.[10] But let us even suppose that the number 5

[9] Honestly, I am not certain where this view can be found in print as a view held by any living philosopher. But I routinely hear philosophers *say* things along these lines. It is also the way basic arithmetic is commonly explained to children. It is easy to imagine a parent or teacher saying, "Take two. Now add 3. What do we have now? 5. Good. See, 2+3=5!"

[10] Admittedly, depending on our ontology, there may well be objects that are actions; perhaps the bombing of Dresden, for example, is an event-object. But to my knowledge, nobody believes that numbers are actions,

is an action-object, and that it is the very action-object of summing 2 and 3. Then, since '$4 + 1 = 5$' is also true, the number 5 must also be the action-object of summing 4 and 1. By the transitivity of identity, then, the action of summing 2 and 3 must be identical to the action of summing 4 and 1. That is, these actions are not just equivalent actions thanks to associativity; they are the very same actions. Certainly such a result is counter-intuitive. We must conclude that it is deeply implausible that the '+' represents an operation, at least if an operation is a performed action.

Since the '+' symbol, together with its two unbound variables, presumably does not represent the performance of an operation of adding or summing, it must represent something else. Perhaps, rather than representing the performance of an operation, '$a + b$' represents the result of an operation. That is, perhaps the expression '$2+3$' is polysemous between the performance of the operation (the adding itself) and the result of performing the operation. If so, the latter meaning (the result of adding) is the meaning that makes the equation sentence ('$2 + 3 = 5$') true.

From here on out, let us assume that '$2 + 3$' represents the result of the addition operation on the numbers 2 and 3. It will be useful to consider how the results of the operation represented by '$a + b$' are determined. What is the '$a + b$' operation, fundamentally?[11]

4.3.1 Collections, Combinations, and Sums

One common way of thinking about addition is as an operation of combining, collecting, or putting together. After all, if we put two apples together with three apples, we get five apples. It is natural to think that we have performed an operation of adding apples. We might think about this operation we have performed—the adding, or putting together, or combining, or collecting—either in terms of mereology or in terms of the union of sets of objects; the mereological sum of two apples and three apples is five apples, or the union of

or in any way active. Such a view would be intuitively implausible.

[11]For the sake of brevity, I will use the term 'addition' and the phrase 'the + operation' in place of 'the operation represented by the '+' symbol, flanked by two unbound variables'.

the two-apple set with the three-apple set is a five-apple set.

There are a number of problems with this kind of account of the addition operation. First, apples are not themselves number-objects; I can eat apples, but I cannot eat numbers. We have described an operation on apples, rather than an operation on number-objects. There may well be some non-arbitrary relation between this operation on apples and the addition of numbers, but in describing the operation on apples we have not, strictly speaking, described the addition of numbers. It might be that the operation on apples is an *application* of the addition operation on numbers. Perhaps we even could explain such an application.[12] But we are interested in explaining the addition operation itself, rather than methods of or criteria for its application to situations involving things other than number-objects.

It also might be that the relevant number-objects (viz. the numbers 2, 3, and 5) somehow involve the relevant apples, either directly or indirectly; it might be that apples somehow *fall under* numbers. Frege's view, according to which numbers are the extensions of second-level concepts, is an example of something like this.[13] Such views require positing internal structure to individual number-objects. More particularly, it seems that the relevant kinds of internal structure could not be given in strictly mathematical terms. At the very least, if the number 3 somehow covers our trio of apples, the number must have more structure than simply being identical with 1+1+1, or the structural relations of addition, multiplication, etc. that it has with the other natural numbers. In addition to the problem raised previously that we do not yet know what operation the '+' sign represents, such a simple account of internal structure provides no explanation of how our trio of apples might be connected to the nature of the number 3. So, the operation on the apples can only be a case of addition if number-objects have some kind of extra-mathematical structure or anchor. In that case, further knowledge would be required to understand Leibiz's proof: knowledge

[12] We could, for example, explain something about the one-to-one correspondence between the collections of apples and the sets of predecessors of the numbers in question.

[13] Charles Parsons is a more modern example of someone with such a view. He expounds his view in *Mathematical Thought and Its Objects*.

of the metaphysical relation between collections of objects (such as apples) and the numbers. I assume that most of us lack such knowledge. Since we do understand Leibniz's proof, it seems that such metaphysical knowledge is not required to understand the proof, and so it seems unlikely that the numbers have such non-mathematical structure.

Second, the operation on the apples will not always work as advertised. Two apples put together with three apples do not always yield five apples; sometimes the two apples and the three apples have an apple or two in common. The two apples and the three apples must be disjoint, or non-overlapping, for the operation to work. This point extends to addition on the number-objects too. The collecting and mereological summing of number-objects is largely the same as the collecting and mereological summing of objcts of other sorts. For example, the mereological sum of an object x and the very same object x is just x, whether x is an apple or the number 2. Since $2+2=4$, and $2+2 \neq 2$, the operation of addition cannot be a mereological summing of number objects. For similar reasons, addition also cannot be the union of sets of ones; it cannot be a putting-together operation in any ordinary sense.[14]

We now have two reasons for rejecting the picture of addition as an operation of collecting, combining, or mereologically summing objects—assuming, of course, that number-objects are not to be identified with entities having non-mathemetical internal structures. One reason to reject the picture of addition as combining is that this picture does not seem to arise from the addition operation itself. Addition is an operation on number-objects, but this picture seems to arise from our common use of arithmetic to model non-mathematical cases (like combining collections of apples). In doing so it gets the order of explanation wrong. We use mathematical structures to explain and understand the non-mathematical world, not the other way around. The second reason we have for rejecting this picture of addition is that operations of collecting, combining, or mereologically summing number-objects seem to

[14]Frege argues similarly in the *Foundations of Arithmetic*, §29-45. As he writes in §45:

> The word 'one', as the proper name of an object of mathematical study, does not admit of a plural. Consequently, it is nonsense to make numbers result from the putting together of ones. The plus symbol in $1+1=2$ cannot mean such a putting together.

yield results that we do not expect from addition (e.g. the undesirable result that $2+2=2$).

Assuming that numbers do not have non-arithmetical internal structures, it seems deeply implausible that addition is an operation of combining, collecting, or mereological summing. We must consider other possibilities. Presumably the genuine addition operation, the one on the number-objects, cannot be any sort of ordinary physical action or activity; after all, numbers are not ordinary physical objects like apples. But if addition is not such an operation, what else might it be? One possibility is that it is an extensional function. Another is that it is a Fregean function, or something in that vicinity.

4.3.2 An Extensional Function

Perhaps addition—that is, the operation that the '+' sign represents—is not an activity, nor any other kind of operation that takes time. We might think that addition is a function in the extensional sense. There are a number of ways to cash out this extensional sense of function. An extensional function (in the case of addition) might be a set of ordered triples, of form $< a, b, c >$, where intuitively the triple means that $a + b = c$. If we want to avoid a set-theoretic reduction, we might think of the extensional function as a graph; one such graph might be the 3-dimensional one in Cartesian coordinates in which, intuitively, the ordered triples mentioned before give the graph's coordinates. We also might think of an extensional function in terms of inputs and outputs; inputs a and b yield output c, and this is just a brute fact with no further explanation.

There seem to be many different frameworks for thinking about what it is to be an extensional function, but let us not dwell long on any puzzles related to that concern. Regardless of the account of extensional function that we give, I propose, we will not be able to account for Leibniz's proof that $2+2=4$ (and similar proofs) if addition is an extensional function. We will run into one of two problems with the proof. Which problem we encounter will depend on what the proof is directed towards demonstrating. The proof might be directed towards demonstrating something about number-objects; e.g. the number-object that results from the addition of 2 and 2 is identical to the number object 4. Or, it might be directed towards proving something about language or notation; e.g. if the expressions '$2+2$' and the numeral '4' refer at all, they refer to the same object, whatever that object might be.[15]

In the former case, in which the proof is directed towards demonstrating a fact about a relation holding between objects, no non-trivial proof of '$2+2=4$' is required. Because addition is an extensional function, the addition of ones in no way grounds the addition of larger numbers. Every result of adding number-objects is an equally fundamental part of the definition of the extensional addition function. Accordingly, there is no need for the middle steps in Leibniz's proof. The fact that the object referred to by the numeral '4' (*that* object) is the very same object as the object that results from the adding of 2 and 2 (the adding of *this* object and *this* object) is simply included in the extension that defines the addition function. Any true sentence of the form '$a+b=c$' is trivial in virtue of the extensional definition of addition.

In the latter case, the proof is directed towards demonstrating a fact about language—in particular, a fact about the co-reference of numerical expressions, regardless of what the common referent happens to be. Rather than demonstrating a truth about number-objects, the proof demonstrates that there are certain kinds of linguistic connections between numerals.

One problem with taking the proof to demonstrate this linguistic connection is that it is prima facie implausible. It seems to get the subject matter wrong; rather than proving something about numbers, we prove something about language. As a matter of course, we

[15]The two possibilities here of the content of what we are trying to prove map onto the two different interpretations Frege offers in the first paragraph of 'Sinn und Bedeutung' about what the relation of identity is. Identity, according to Frege, is either a relation between objects, or a relation between the names of objects. Although Frege earlier thought (in the *Begriffschrift*) identity to be a relation between the names of objects, here he seems to think identity to be a relation between objects. There is, however, some debate on how Frege is properly interpreted on this point. See, e.g. Thau and Caplan 'What is Puzzling Gottlob Frege?' and Heck 'Frege on Identity and Identity-Statements: A Reply to Thau and Caplan', both in the *Canadian Journal of Philosophy*.

mention numerals in our basic mathematical practices, rather than *using* them. Frege puts it this way:

> In that case the sentence $a = b$ would no longer refer to the subject matter, but only to its mode of designation; we would express no proper knowledge by its means.[16]

But prima facie implausibility is not the only impediment to taking this approach; it has more serious difficulties.

If this sort of account is to work, we must think of the definitions of numerals in a slightly different way. We had been thinking of the definition '$4 = 3 + 1$' as directly fixing reference; the numeral '4' is defined as a name for *that number-object*, viz. the one to which the expression '$3 + 1$' refers. But a purely reference-fixing definition does little to identify the connections between the relevant linguistic items; such a definition only serves to identify the connection between one linguistic item (the numeral '4') and one non-linguistic item (the relevant number-object). If the proof is to demonstrate a fact about language, the definitions of numerals cannot be purely reference-fixing; they must provide a linguistic connection between the definiendum numeral and the numerals used in the definiens.

The most natural way for there to be a linguistic connection between the definiendum and the definiens is for a definition to stipulate co-reference. For example, the definition '$4 = 3+1$' is the stipulation that the numeral '4' co-refers with the expression '$3+1$'. Leibniz's proof then gets its legs from the combination of these facts about co-reference, given by the definitions of the numerals, with facts about the compositional nature of numeral-expressions involving the '+' sign.

So long as this linguistic approach allows us to account for the relevant aspects of the compositional nature of expressions using the '+' sign, we have a potentially viable story about how the definitions and proof work. Unfortunately, no such account is available.

[16]'Sinn und Bedeutung', first paragraph.

Leibniz's proof requires the associativity of addition, but this feature will not be available as required to prove that '4' and '2 + 2' co-refer.

The unavailability of associativity for Leibniz's proof is a consequence the fact that '$a + b$' represents addition, which we are supposing to be simply an extensional function on number-objects. Addition is not an operation defined on numerals. Were this operation defined extensionally on numerals, a definitional circle would arise. The numerals would be defined in terms of the + operation, and the + operation would be defined in terms of the numerals. So expressions of the form '$a+b$' must represent an operation on number-objects.

Since expressions of the form '$a + b$' must represent operations on number-objects, features of the way the '+' sign (together with its unbound variables) works compositionally must parallel features of the addition operation on number-objects. It ought to be fairly clear that the only plausible explanation of this parallelism of features is that the compositional features of the sign are grounded in the compositional features of the operation on numbers.[17] One of the relevant compositional features of the symbol in arithmetical language is the feature that parallels the associativity of addition. For any three numbers α, β, and γ, $(\alpha + \beta) + \gamma = \alpha + (\beta + \gamma)$; similarly, for any three numerals a, b, and c, expressions of the form '$(a + b) + c$' and '$a + (b + c)$' co-refer.[18] Call this linguistic feature 'associative co-reference'.

The version of Leibniz's proof under consideration relies upon associative co-reference, since the claim we are interested in proving is a linguistic claim about co-reference. But even

[17] In case it is not clear, here is a brief argument. There are three possibilities. Either the features of the sign and the features of the operation on numbers have analogous features as a matter of happy coincidence, or the features of one are grounded in its relation to the other, or the analogous features have a common grounding in something else entirely. Having a common grounding in something else is not a viable option because the addition operation on number-objects is an extensional function; it has no grounding aside from its extension. This fact also undermines the possibility of the features of the operation on numbers being grounded in the features of the '+' sign. And if the analogous features are the result of a happy coincidence, then it is mere coincidence that the proper use of a symbol representing a function reflects the nature of that very function. But that is absurd; surely the proper use of a symbol is non-coincidentally related to that which it represents.

[18] Notice that this is in keeping with the style of numerical definition: '$4 = 3 + 1$' means '4' and '$3 + 1$' co-refer.

though associative co-reference is linguistic, it is grounded in a feature of the numbers and an operation on those numbers (viz., the associativity of the operation). Furthermore, since the linguistic features of the '+' sign are grounded in the operation of addition, every proper use of the '+' sign must be grounded in this extensional function on number-objects.[19]

Our problem arises here, in the grounding of the proper use of the '+' sign, and in particular the grounding of associative co-reference, which is part of the proper use of this sign. Given the rigorous nature of arithmetical proofs, any assumption used in a proof must be solidly available to mathematicians.[20] In particular, associative co-reference must be solidly available to mathematicians if it is to be legitimately used in the vesion of Leibniz's proof under consideration. It can only be solidly available, though, in virtue of the associativity of addition on the number-objects being solidly available.[21] And given the infinity of numbers and the fact that addition is an extensional function, the associativity of addition is not solidly available. The associative feature of the operation of addition is grounded in the infinite extension that defines addition, so if the associative property is solidly available, the infinite extension must also be. That, of course, is impossible for creatures like us with finite minds. Perhaps we could hypothesize the associative nature of addition, but such a hypothesis would not suffice for use in a rigorous arithemetical proofs.

Thus, Leibniz's proof that $2 + 2 = 4$ cannot be successful under the assumption that the operation of addition is defined extensionally. If the proof is aimed at showing something about mathematical objects, it is superfluous. If the proof is aimed at showing something

[19]Perhaps I should be more careful when saying something about *every* proper use of the sign, because presumably it has other meanings in non-arithmetical contexts. But forgive my sloppiness; I only mean to discuss arithmetical contexts here.

[20]I distinguish between *assumptions* and *suppositions*. The former must be solidly available; the latter need not be (and may be reduced to absurdity as part of a *reductio* argument).

[21]Suppose not. Then associative co-reference is solidly available in the absence of the solid availability of the associativity of addition. Then either it is not available that the '+' sign is used properly when associative co-reference is at issue, or facts about the proper use of the '+' sign are solidly available without the solid availability of facts about whether or not the proper uses of the '+' sign reflect features of the operation it represents. The latter option is absurd. As for the former option, a linguistic law is not available for the purposes of mathematical proof if it is not even available whether or not that linguistic law is in keeping with the proper use of the central term guiding its application.

about language, one of the crucial assumptions required for it is not available. So, assuming that Leibniz's proof is of some value (an assumption I am happy to make), addition cannot be defined extensionally.

4.3.3 Fregean Functions

Since addition can neither be an extensional function nor any kind of physical operation, we might think that addition is, in the Fregean sense, a function of two arguments. On Frege's view, functions are fundamentally incomplete:

> . . . a function by itself must be called incomplete, in need of supplementation, or 'unsaturated'.[22]

Functions are so to speak gappy—a gap in a function can be filled with an *argument*, which is an object (a complete entity) that completes or saturates the function, thereby yielding a *value* (assuming this is a function of one argument). The value is a complete entity, such as a number or a truth-value:

> . . . the argument does not belong with the function, but goes together with the function to make up a complete whole . . .[23]

An example might help elucidate this account. On Frege's view, concepts turn out to be functions:

> . . . a concept is a function whose value is always a truth-value.[24]

So, for example, a (Fregean) concept like 'human()' will be a (Fregean) function that yields the value True when it goes together with an object like Socrates or Frege as argument, but yields the value False when it goes together with an object like the Moon or the Sun as argument.

[22] See "Function and Concept."

[23] Ibid.

[24] Ibid p. 15.

Given this account, addition is not a concept in the Fregean sense. Taken as a Fregean function, it has the form $() + (\_)$, where the two arguments must be numbers.[25] Once saturated, this function will not yield a truth-value; its value is always a number-object.

This account has some intuitive appeal. It retains the compositional feature of expressions of the form '$x + y$'; the referent of the expression is determined by the composition of the function denoted by '$() + (\_)$' with certain arguments, viz. the objects denoted by x and y. It also explains why truths of the form '$x + y = z$' can be informative: while the same object gets denoted by '$x + y$' and by 'z', the two expressions can have different senses. One of these senses involves the composition of two arguments, x and y, with the addition function. The other sense does not have the same compositional structure. Furthermore, this account can explain the substitution of equals required by Leibniz's proof. The arguments of a function are *objects*, not *expressions* or *senses*. Whether one of the arguments is denoted using the numeral '2' or the expression '$1 + 1$', to use a particular example, the composition of function and arguments will be exactly the same.

Though Frege's account of functions can explain some of the key features of the addition function and Leibniz's proof that $2 + 2 = 4$, like compositionality and the substitution of equals, it leaves some questions unanswered. One unanswered question, which we shall set aside for now, is the question of what a function *is*. Frege describes some features of functions, such as their being unsaturated and combining together with argument-objects to yield values, but these features do not seem to suffice for a satisfying account of the nature of functions.

Another unanswered question has to do with the content of the addition function, and what distinguishes this particular function from other functions. Presumably a satisfactory account of the nature of this particular function will explain the associativity of addition, which is of particular interest to us right now because of its role in Leibniz's proof that $2 + 2 = 4$.

[25] I use $()$ and $(\_)$ to designate that the gaps need not be filled by the same argument.

A Primitive Function

One possibility is that the addition function is *primitive*. That is, it has no components; it is not a composite entity built from smaller or simpler parts. Assuming this to be the case, there must be some way to epistemically access this function, this denotation of the '+' symbol, in a way that allows us to successfully execute Leibniz's proof. This could occur in one of two ways. Either we could have primitive knowledge of the addition function, or we could develop the requisite knowledge of the function through our knowledge of other objects or functions. This latter alternative might take any of many different forms, such as knowledge of the pairings of arguments with values, or knowledge of other functions and objects that can be used to define the function.

If the former option holds, and our knowledge of the addition function is in some way immediate or direct, our definitions of numerals like '4' are solid. Assuming we have such knowledge of the function represented by '+', and we also have knowledge of the referents of '1', '2', and '3', we are in a position to define the numeral '4' in terms of its coreference with '$3 + 1$'.

There will, however, be a couple unsettling results. First, it might be unclear why a proof of $2 + 2 = 4$ is required at all. If we have unmediated knowledge of the addition function, it is not obvious why a proof is required to secure our knowledge of this arithmetical truth. It might seem that our knowledge of the way the addition function works is no more privileged when the number one is one of the arguments than it is when both arguments are other numbers. To be honest, it strikes me as intuitively compelling that this kind of direct knowledge of addition would render arithmetical proofs unneccessary for the purposes of securing mathematical truth. But I am far from convinced by this intuitive tug. It seems that a primitive-knowledge account will be compatible with many explanations of why the addition of smaller numbers would be more epistemically secure than the addition of larger numbers.[26]

[26]Let me give an example of one potential explanation to handle objections that the primitive knowledge

The second result is more unsettling. One of the fundamental tools we need to complete Leibniz's proof is the associativity of addition. But if our knowledge of the addition function is direct or immediate, we are unable to explain our knowledge of associativity. We are supposing that addition is a primitive function, and as a function it takes two objects as arguments. Since the arguments of the addition function must be numbers, any structure that the arguments have is structure that the numbers have. Thus, unless the numbers that serve as arguments themselves have a particular compositional structure, viz. that of the addition function completed with two number-object arguments, compositional expressions like '$(a + b) + c$' will not reflect the actual composite structure of addition.[27] Rather, the referent of the expression '$(a+b)+c$' is the value of the addition function together with two number-object arguments. One of those arguments is the referent of 'c', and the other is the referent of '$a + b$'—that is, the number-object that is the value of the addition function together with number-object arguments a and b. Surely this composition is not the same as that expressed by '$a + (b + c)$'. This latter expression refers to the composition of the addition function together with arguments quite distinct from the ones mentioned before.

account renders proof unneccessary. I do not intend to endorse this explanation, but it strikes me as fairly plausible. This explanation depends on the idea that we have a better cognitive grip on smaller number-objects than on larger ones. As a result, the addition of smaller numbers might be more secure than the addition of larger numbers. This might lead us to have a better cognitive grip on the composition of a related function that we understand through the addition function, viz. the single-argument function $() + 1$, than we do on the addition function itself when it is composed with larger numbers as arguments.

[27]Even if the numbers did have the relevant compositional structure, the addition function would not structurally suggest associativity—at least, not immediately or in a way that would allow us to use it as a foundational truth. For the function-argument structure of $(a + b) + c$ to reveal its identity with $a + (b + c)$, the overall structure of the addition function composed with arguments a and $b + c$ must not take the main addition connective in the expression of the function to have any priority over the other addition components in the actual function; regardless of parentheses in the expressions, '$(a+b)+c$' and '$a+(b+c)$' must denote the same function (unsaturated, without particular arguments yet assigned). But then if, for example, $a+b = d+e$, and the arguments identified in the expression '$(a+b)+c$' are just the referents of '$a+b$' and 'c', then the former of these arguments must also just be the referent of '$d + e$' also. So '$(a + b) + c$' and '$(d + e) + c$' must represent the exact same compositions of function and argument. Assuming that $a \neq d$ and $a \neq e$, then, associativity will only be structurally obvious if the numbers are sums of ones, and the central '+' in the expression refers to an instance of the addition operation that has no structural priority over any of the others. It no longer seems that addition would be a primitive function; rather, it would seem to be a composite function given in terms of addings of 1. And furthermore, it seems that the associativity of the general case of $(a + b) + c = a + (b + c)$ would be derived from the associativity of the particular case $(1 + 1) + 1 = 1 + (1 + 1)$. It no longer seems to be an obvious property, fundamental enough to be used without explanation in a proof.

These arguments are the referents of 'a' and '$b+c$', rather than of '$a+b$' and 'c'; unless $a = c$, these are different argument-objects. Thus, unless the addition function—the unsaturated thing—is built from components, the structure of the addition function gives us no reason to believe that $(a + b) + c = a + (b + c)$. The structure of the addition function therefore provides no explanation of the associativity of addition.

In order to use associativity in Leibniz's proof, we must have some explanation or knowledge of it. But if the addition function is primitive, and if our grasp on the function is direct and does not come through our understanding of other objects or functions, then grasping the addition function is sufficient neither for us to explain why addition is associative, nor to immediately recognize that addition is associative.

Knowledge through Other Objects or Functions

Let us then consider the latter alternative: that, though the addition function is primitive, our knowledge of it comes through knowledge of other objects or functions. If so, the relevant objects or functions that allow us to have knowledge of the addition function must suffice for us to also have knowledge of the key features of this primitive function, and the key features of expressions used to pick it out (that is, expressions involving the '+' symbol). In particular, for the purposes of Leibniz's proof, these functions or objects must provide the grounds for our knowledge of associativity, the substitution of equals, and the denotations of the particular expresssions used in the definitions of the numerals (e.g. the denotation of '$3 + 1$' in the definition of the numeral '4').

Let us suppose that our knowledge of the addition function only arises through our knowledge of other functions or objects. It seems that there are four ways in which this might occur. First, our knowledge of the addition function might arise as a result of our pairing arguments with values; 2 and 2 get paired with 4, 5 and 7 with 12, and so on. Second, it might arise as a result of placing some sort of restriction on some more general function, perhaps by fixing one or more arguments, or by restricting the available arguments to those that stand in some relation to one another. For example, there might be a function of three arguments that gives us the addition function when two of the argument places must be filled by the same argument. Third, our knowledge of the addition function might arise as a result of our knowledge of one or more restrictions of the addition function. We might, for example, come to know it through our knowledge of the single-argument function that we express as '() + 1', or perhaps knowing many other such functions too, such as those expressed as '() + 2', '() + 3', '() + 4', and so on. And fourth, our knowledge of addition might arise as a result of describing the intension, or meaning, of the function. This might happen, for example, through an implicit or explicit definition of the function, or through an account of the complement of the function—an account of what the function is not.

Setting aside what Frege might have thought,[28] none of these four approaches seem to provide plausible explanations of how we could come to know a primitive addition function. The first option, pairing arguments with values, is just one way to think about extensional

---

[28]Frege's view on functions is a tricky matter, and it is not clear whether or not some of these options straightforwardly conflict with his view. Given some of the things he says in "Function and Concept," he might be understood to believe that there are distinct functions with identical argument-value pairings:

> There are functions, such as $2+x-x$ or $2+0\cdot x$, whose value is always the same, whatever the argument; we have $2 = 2+x-x$ and $2 = 2+0\cdot x$. Now if we counted the argument as belonging with the function, we should hold that the number 2 is the function. But this is wrong. Even though here the value of the function is always 2, the function itself must be distinguished from 2; for the expression for a function must always show one or more places that are intended to be filled up with the sign of the argument. (p. 7)

This quotation might be taken to suggest that, on Frege's view, argument-value pairings are not sufficient to uniquely identify a function. But it is unclear whether that is a fitting interpretation of Frege. After all, he identifies $2+x-x$ and 2 as distinct, but the former is a function (it takes arguments) while the latter is an object (it is saturated, and does not take arguments). He does not explicitly say that $2+x-x$ and $2+0\cdot x$ are distinct functions. Thus, he has not unambiguously given a case in which the *pairings* of arguments with values are identical and the functions are not.

He later adds a comment that might be taken to suggest again some resitance to the idea that the argument-value pairings uniquely identify a function:

> If we write
> $$x^2 - 4x = x(x-4),$$
> we have not put one function equal to the other, but only the values of one equal to those of the other. (p. 9)

But this might simply be a caution about jumping to identity conclusions in cases of function; he does not explicitly claim that x^2-4x and $x(x-4)$ are distinct functions. Let us, then, set aside the complexities of Frege scholarship and just focus on what might be true.

functions. As we have already seen, understanding addition through its extension will not position us to know the associativity of addition securely enough to use it as an assumption in Leibniz's proof.

The second option, according to which we reach addition by putting restrictions on the arguments of a more general function, will not suffice for Leibniz's proof either. The resulting restricted function, which we suppose to be addition, will be a composite of one function with another function, or with an argument. We have also supposed, however, that addition is a primitive function. So, either the composite including the more general function positions us to recognize another function (addition) that we grasp primitively, or we grasp addition through the composite. In the former case, regardless of the inspiration, we have a primitive grasp of addition. We have already covered the arguments against this option. In the latter case, our full grasp of the addition function is through this composite that includes the more general function. But given the lack of candidates for such a role, it is prima facie implausible that we grasp addition this way. Furthermore, if we can only grasp addition through this composite, it seems that the postulation of a distinct and primitive addition function is superfluous; the composite function alone provides all our knowledge of the structural features of addition, and presumably shares those structural features. Let us use Ockham's razor to shave the primitive addition function from the account. Soon we will consider the possibility that addition is a composite function.

Largely similar concerns will undermine the third option, according to which we can come to know addition through our knowledge of restrictions of the addition function. Either the function $() + 1$, for example, is a composite consisting of the addition function partially saturated with one argument (the number 1), or it is its own primitive function. In the former case, we must already know the addition function component of the composition in order to know the composite $() + 1$ function. In the latter case, the situation is again one in which our full grasp of the addition function is through this other function. Once again, it seems fitting to take Ockham's razor to the primitive addition function; the function $() + 1$, along with the construction we use to build the addition function from it, will suffice alone for use in Leibniz's proof.

The fourth option covers a range of possibilities: explicit definitions of the intension of the function, implicit definitions, descriptions of the complement, etc. These possibilities will fall into one of the two problems we have already seen. One possible problem is that the function will be defined or described in a way that attributes some kind of composite structure to the function, which we have assumed to be non-composite. The other is that our definition or description will not uniquely determine the addition function, either because a non-composite function is identified through compositional means, or because it is introduced by an implicit definition, which possibility we found unsatisfactory in the second chapter.

None of these four ways of using other functions to grasp a primitive addition function seems to be a satisfactory approach. So, let us consider the possibility that the function is composed or defined using other functions. Let us consider the possibility that the meaning of the '+' is constructed.

4.4 The Successor Function

The obvious candidate for constructing the meaning of the '+' from more basic components should be familiar from Peano Arithmetic. This is the recursive definition of addition in terms of the successor function (S):

- $a + 1 = S(a)$
- $a + S(b) = S(a + b)$[29]

There are two things to notice about such a definition. First, the first clause can define in either direction, so to speak. One possibility is that the '+1' operation is assumed,

[29] This definition of addition relies on the assumption that 1, rather than 0, is the smallest natural number. If we prefer to begin with 0 instead, as is commonly done in Peano Arithmetic, we can modify the definition by replacing '$a + 1 = S(a)$' with '$a + 0 = a$'.

and the successor operation is defined in terms of it. Notice that this approach would not obviously run into all the same sorts of problems that we have seen before, many of which arose from worries about adding potentially large numbers. The second clause would then use the newly-defined successor operation to define addition on larger numbers. The other possibility is that the successor operation is assumed, and all cases of addition are defined in terms of it.

It seems that the definition fits the latter description; the successor operation is assumed, rather than the '+1' operation. Were the '+1' operation assumed, the first clause wouldn't be necessary. The addition of larger numbers could be defined directly in terms of addition of ones in the following way: '$a+(b+1) = (a+b)+1$'. A further assumption would be required, of course, viz. that for every number $a > 1$, there exists a number b such that $a = b + 1$. But an analogous assumption is also required when we define addition in terms of successor.

Even though the particular definition provided above seems to take the successor function to be basic, nothing important rides on whether we base the definition on the adding of ones or on the successor function. In either case, we identify one operation with the other: $S(a) = a + 1$. The successor of a number is identical with the result of adding one to that number. We start with an operation on one number, which we may label with 'successor' or with 'plus one'. Ultimately, the operation in question is given both labels. In what follows, I will take successor to be the more basic form.

Second, assuming that the successor function is taken to be primitive, this definition clearly defines addition in terms of *ordinal* number instead of *cardinal* number. Regardless of whether the definition of addition assumes the successor function or the +1 operation, the assumed operation treats numbers as having an *order*—the order provided by the successor or +1—and not simply a size in terms of one-to-one correspondences. The successor (or +1) function puts the numbers in order, and relies on that order to define the operation of addition.[30]

Once addition is treated as a defined function, we can put this definition to use in Leibniz's proof. Recall that accouting for the associativity of addition was the most significant problem we had with this proof when we took addition to be a primitive function. Fortunately, associativity becomes a rather straightforward affair once addition is treated as a defined operation. Using the '$a+(b+1) = (a+b)+1$' version of the definition of addition, we get:

$2+2 = 2+(1+1)$ (by the definition of '2' and the substitution of equals)

$= (2+1)+1$ (by the definition of addition)

$= 3+1$ (by the definition of '3' and the substitution of equals)

$= 4$ (by the definition of '4').

The definitions of the numerals in terms of successor should be obvious, and the proof is essentially the same (but includes more lines for substitutions) when successor is used instead of $+1$. We are thus well-positioned to give Leibniz's proof that $2+2=4$, provided that we can account for successor.[31]

4.4.1 Accounting for Successor

Just as we must understand the expression '$3+1$' compositionally in order to define the numeral '4' using co-reference, so we must also understand '$S(3)$' compositionally in order to define '4' once we reduce addition to the successor function. So, what is the successor function, and how do we understand it?

[30] Frege gives an account of cardinal number using something like the successor function. The reason his account is of cardinal number, and not ordinal number, is that he uses the equinumerocity of concepts in order to define the successor function. Accordingly, he does not take the successor function to be primitive, as we are assuming here that it is.

[31] While the uses of associativity required by Leibniz's proof and others of the kind are immediate from the definition of addition, the general assocativity of addition requires a short proof, by strong induction. I argue in Chapter 2 that we cannot make use of strong induction without already understanding the structure of the sequence of natural numbers.

As with addition, it seems *prima facie* unlikely that successor is merely an extensional function. But let us suppose, for the sake of argument, that it is. Then, assuming that addition is defined over all natural numbers (an assumption we should be happy to make), we must use the full extension of the successor function in order to give a full definition of addition. Were we to limit ourselves to some proper initial segment of the successor extension, we would not succeed in defining expressions of the form '$a + b$' over all of the numbers. In particular, we would not succeed in defining it for sufficiently large a or b. After all, infinitely many extensional functions over all of the natural numbers will agree with the successor function on any of its given finite subsets of argument-value pairings. Thus, in order to fully define addition, we must be able to distinguish the full extension that is the successor function. We cannot do this; our minds are finite. So, if successor is extensionally defined, we cannot use it to fully define addition. The result that we cannot define addition in this way is deeply implausible, so it seems unlikely that successor is an extensional function.

Furthermore, the basic axioms of Peano Arithmetic involve the successor function, and it seems that we must be able to know these axioms. That is, we do know these axioms; we intuitively accept them as these axioms as accurate descriptions of the way the successor function operates on the natural numbers.[32] In particular, I have the following axioms in mind (where n and m are natural numbers):

- $(\forall n)S(n) \neq 1$[33]
- $(\forall n)(\forall m)(S(n) = S(m) \rightarrow n = m)$

If successor is merely an extensional function, we are not in a position to know these axioms. Unless we knew the full extension that is successor, we would not be able to know that it

[32]Recall from Chapter 2 that our knowledge of arithmetic—and in particular, our knowledge of the successor function—is not grounded in axiomatic definition.

[33]Recall that we are assuming that natural numbers begin with 1, not with 0. Were we to begin with 0, the same form would apply: $(\forall n)S(n) \neq 0$. For our purposes, nothing of consequence will ride on this.

includes no pairing of an argument with a value of 1, or that no two distinct arguments are paired with the same value. Since these truths are obvious to us, it seems that successor is not an extensional function.

The successor function also is not a promising candidate for definition in terms of other functions or in terms of more basic concepts. There have been attempts to define it (aside from defining it as +1). But such attempts inevitably require us to treat the natural numbers in non-arithmetical terms. For example, if we were to reduce arithmetic to set theory, we could take zero to be equal to the empty set, and then define successor as the following operation: $S(n) = n \cup \{n\}$. Then we take every natural number to be a set formed in this way. While we can define a successor (or successor-like) function on these sets, we either are identifying natural numbers with sets, or we assume the successor function and merely draw an isomorphism between the natural numbers and the relevant sets.[34] Frege does something similar in *The Foundations of Arithmetic*. He identifies numbers using an operation on things of another kind, viz. second-level concepts, and then describes a relation between second-level concepts that yields the successor function.[35] But it seems that we can understand Leibniz's proof without knowing anything about second-level concepts, or what it is to take the union of a set with its singleton. It seems that this approach could not explain how easy it is to understand Leibniz's proof.

[34]Recall §4.1.

[35]See §76. Here is how Frege describes this operation:

> The proposition:
>
> "there exists a concept F, and an object falling under it x, such that the Number which belongs to the concept F is n and the Number which belongs to the concept 'falling under F but not x' is m"
>
> is to mean the same as:
>
> "n follows in the series of natural numbers directly after m."

4.4.2 A Primitive Successor Function

We are left, then, thinking of the successor function as primitive and basic. We do not know it through its extension, nor do we know it through any more basic arithmetical functions. It seems that we do not know it through other functions of any sort. Successor seems to be about as primitive and basic as a function can be.[36]

There is some inclination, of course, to describe the successor function in terms of ordinary language. The successor of a number n is just the next number after n, or the number that follows n. These descriptions fit with our colloquial usage of the term 'successor', when we are not discussing arithmetic or numbers. Elizabeth II is the successor of George IV, because Elizabeth immediately followed George as the reigning monarch of England. But these descriptions do not serve to explicate the successor function. Rather than giving an account of the successor function, we seem to be giving synonyms that apply in a broader class of domains (e.g. English monarchs). We seem to grasp all of these operations with equal immediacy. The only thing that separates them is the domains on which they operate.

Let us suppose that we do, in fact, grasp this primitive successor function, and that we instinctively understand how it operates on the natural numbers. Then we need not know in advance, the full extension of the successor function in order to use it to fix referents using definitions like '$4 = S(3)$' (or '$4 = 3 + 1$'). Given a number n, and the intension of the successor function, '$S(n)$' uniquely picks out a referent, which we can identify through our primitive grasp of the intension of 'S'. Since we have a determinate conception of

[36]Perhaps there is a more primitive function. Perhaps we could think of the successor function as built from the idea of 'a different thing', which may strike us as more basic than 'the next thing'. But then we would have to build the successor function using omething like Peano Arithmetic. One problem with such an account is that it seems to get the order of explanation wrong. But, there are two more striking problems. First, we already established in Chapter 2 that Peano Arithmetic does not uniquely characterize the natural numbers, so it does not uniquely characterize successor (though perhaps it could if we limited successor to the natural numbers). And Peano Arithmetic does nothing to guarantee that $S(2) = 3$, for example, instead of $S(2) = 4$. It would be completely consistent with such a definition of successor to switch the order on the natural numbers.

Second, the successor function would then be a generic and implicitly-defined function. A structure that is implicitly defined does not distinguish between isomorphism classes, as discussd in Chapter 2. So the successor function would not pick out a unique referent for '$S(1)$'. This would make our definition of the numeral '2' an unsuccessful definition.

the successor function we use to define addition, the definition of the addition function presumably covers the addition of any pair of natural numbers—as we would expect.

Assuming that we do have this immediate understanding of the successor function, Leibniz's proof no longer seems to pose a problem. The relevant use of associativity follows immediately from our definition of addition, and the definitions appear to be informative and useful. If we already know the referent of '3', our understanding of the successor function positions us to identify its successor, which we then label '4'.

Our immediate understanding of successor would have further implications as well. This is because we intuitively accept the axioms of Peano Arithmetic.[37] We intuitively recognize that the number 1 is not the successor of any other natural number, and that no two distinct numbers have the same successor.[38] But once we put this together with another fact we intuitively recognize, viz. that every natural number has a successor, we get an infinite sequence of numbers. Our primitive understanding of the successor function, and its operations on the numbers, reveals that we understood something else as well. It reveals that we already, at some level, understand the structure of the sequence of natural numbers.

4.5 Final Thoughts

We ought to be able to successfully execute a standard proof of '2+2=4', and proofs of other similarly simple arithmetical truths. In order to do so, we must be able to make use of a few simple definitions, e.g. '4=3+1'. We also must be able to make use of a few simple arithmetical laws, such as the associativity of addition ($a + (b + c) = (a + b) + c$). As it turns out, our ability to make adequate use of these fairly simple tools depends upon our defining addition from its most simple cases—cases of adding ones. That is, we must have an understanding of the successor function as a function over all the natural numbers in order

[37] I neglect to distinguish here between the first-order axiom schema of Peano Arithmetic and the second-order axioms. But I do not discuss induction here, so the two are effectively equivalent for our purposes.

[38] Recall, once again, that we are assuming that the natural numbers begin at one, rather than zero. This assumption has no real impact for our purposes.

to prove fairly simple arithmetical truths, like '2+2=4'. In order to understand the successor function that well, we must have a grasp of the entire structure of the natural numbers as an infinite, nonrepeating sequence.

4.6 Conclusions

In the first chapter of this dissertation, it became clear that we have a number of desiderata for theories of mathematical knowledge. One of these desiderata is that an account of mathematical knowledge ought to be able to explain all of our mathematical knowledge. Another is that an account of mathematical knowledge ought not to presuppose unexplained antecedent mathematical knowledge. Any account of mathematical knowledge that fails to satisfy these two desiderate is, quite simply, inadequate as a theory of knowledge. Any account of mathematical knowledge that is inconsistent with the satisfaction of these two desiderata is unacceptable.

In the remaining chapters of this dissertation, I considered three ways in which philosophers have attempted to use definitions to account for our mathematical knowledge. I hope to have shown that these accounts are, at best, radically incomplete. It could not be the case that all of our mathematical knowledge is grounded in axiomatic definitions because we have arithmetical knowledge that could not emerge from axioms. It could not be the case that our ability to think about the natural numbers is grounded in a stipulation of Hume's Principle as an implicit definition; first, such a stipulation could not uniquely fix referents for number-phrases, and second, ancient mathematicians could not have stipulated it in a suitable way. And while I think that the use of explicit definitions in arithmetical proofs helps us develop secure knowledge of arithmetical truths, our ability to understand those proofs presupposes an understanding of the structure of arithmetic. Our most basic arithmetical knowledge is not grounded in these explicit definitions, but is cognitively prior to those definitions.

The upshot of all of this is, I believe, that we must have some very basic and primitive understanding of the structure of the natural numbers if we are to have any mathematical knowledge at all. The rest of mathematics presupposes arithmetic, and definitions alone cannot ground all of our arithmetical cognition.

Bibliography

[1] Aristotle (1984). *Metaphysics* in Barnes 1984, Vol. I. pp. 431-599

[2] Aristotle (1984). *Physics* in Barnes 1984, Vol. II. pp. 436-660

[3] Barnes, Jonathan, Ed. (1984). *The Complete Works of Aristotle: The Revised Oxford Translation, Vol. I-II.* Princeton: Princeton UP

[4] Benacerraf, Paul (1973). "Mathematical Truth." *Journal of Philosophy* Vol. 70, No. 8, pp. 661-680

[5] Benacerraf, Paul (1965). "What Numbers Could Not Be." *Philosophical Review* Vol. 74, No. 1, pp. 47-73

[6] Benacerraf, Paul and Hilary Putnam, Eds (1983). *Philosophy of Mathematics: Selected Readings*. Second Ed. Cambridge: Cambridge UP

[7] Benacerraf, Paul and Crispin Wright (1985). "Skolem and the Skeptic." *Proceedings of the Aristotelian Society, Supplementary Volumes* Vol. 59, pp. 85-115, 117-137

[8] Blanchette, Patricia (1996). "Frege and Hilbert on Consistency." *The Journal of Philosophy* Vol. 18, No. 7, pp. 317-336

[9] Cottingham, John, Robert Stoothoff, and Dugald Murdoch, Eds. (1984). *The Philosophical Writings of Descartes, Vol. I-III.* Cambridge: Cambridge UP

[10] Dedekind, Richard (1924). "The Nature and Meaning of Numbers" in *Essays on the Theory of Numbers*. Translation by W.W. Beman. Chicago: Open Court Publishing

[11] Descartes, Rene (1641). *Meditations on First Philosophy*. English translation by J. Cottingham in Cottingham, Stoothoff, and Murdoch 1984 Vol. II. pp. 1-62

[12] Descartes, Rene (1644). *Principles of Philosophy*. English translation by J. Cottingham in Cottingham, Stoothoff, and Murdoch 1984 Vol. I. pp. 193-292

[13] Euclid (1956). *The Thirteen Books of Euclid's Elements*. Volumes I-III. Translated/Edited by T.L. Heath. 2nd Revised Ed. New York: Dover

[14] Ewald, William (1996). *From Kant to Hilbert: A Source Book in the Foundations of Mathematics*. Vol II. Oxford: Clarendon

[15] Frege, Fottlob (1879). *Begriffschrift: eine der arithmetischen nachgebildete Formalsprache des reinen Denkens*. English translation by S. Bauer-Mengelberg as "Concept Script" in Van Heijenoort 1967. pp. 1-82

[16] Frege, Gottlob (1884). *Die Grundlagen der Arithmetik*. English translation by J.L. Austin as *The Foundations of Arithmetic*. 2nd Revised Ed. Evanston, IL: Northwestern UP, 1980

[17] Frege, Gottlob (1892). "Über Sinn und Bedeutung." English translation by M. Black as "On Sense and Reference." *The Philosophical Review* Vol. 57, No. 3, 1948. pp. 209-230

[18] Frege, Gottlob (1893). *Grundgezetse der Arithmetik*. English translation by M. Furth as *Basic Laws of Arithmetic*. Los Angeles: U of California, 1967

[19] Frege, Gottlob (1891) . "Funktion und Begriff: Vortrag, gehalten in der Sitzung." English translation by M. Black as "Function and Concept" in Geach and Black 1980, pp. 21-41

[20] Frege, Gottlob (1892) . "Uber Begriff und Gegenstand." English translation by M. Black as "Concept and Object" in Geach and Black 1980, pp. 42-55

[21] Frege, Gottlob (1980). *Philosophical and Mathematical Correspondence*. Translation by H. Kaal. Eds. G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, A. Veraart, B. McGuiness. Oxford: Blackwell

[22] Frege, Gottlob (1971). *On The Foundations of Geometry and Formal Theories of Arithmetic*. Translation by E.-H. W. Kluge. New Haven: Yale UP

[23] Frege, Gottlob (1980). *Translations from the Philosophical Writings of Gottlob Frege*. Third Ed. Eds. P. Geach and M. Black. Totowa, NJ: Rowman & Littlefield

[24] Gödel, Kurt. "What is Cantor's Continuum Problem?" in Benacerraf and Putnam 1983. pp. 470-485

[25] Hale, Bob (1984). "Frege's Platonism." *The Philosophical Quarterly* Vol. 34, No. 136, pp. 225-241

[26] Hale, Bob and Crispin Wright (2002). "Benacerraf's Dilemma Revisited." *European Journal of Philosophy* Vol. 10, No. 1, pp. 101-129

[27] Hale, Bob and Crispin Wright (2001). *The Reason's Proper Study: Essays Towards a Neo-Fregean Philosophy of Mathematics*. Oxford: Clarendon

[28] Hale, Bob and Crispin Wright. "To Bury Caesar ..." in Hale and Wright 2001. pp. 335 - 396

[29] Heck, Richard G., Jr. (2002) "Cardinality, Counting, and Equinumerosity." *Notre Dame Journal of Formal Logic* Vol. 41, No. 3, pp. 187-209

[30] Heck, Richard (2003). "Frege on Identity and Identity Statements: A Reply to Thau and Caplan." *Canadian Journal of Philosophy* Vol. 33, No. 1, pp. 83-102

[31] Hilbert, David (1899). *The Foundations of Geometry.* Translation by E.J. Townsend. 3rd Edition. Illinois: Open Court Publishing, 1938

[32] Hilbert, David (1926). "On The Infinite." *Mathematische Annalen* Vol. 95, pp. 161-190. Translation by E. Putnam and G. Massey printed in Benacerraf and Putnam 1983. pp. 183-201

[33] Leibniz, Gottfried Wilhelm (1704). *Nouveaux Essais sur l'Entendement Humain.* English translation by A.G. Langley as *New Essays Concerning Human Understanding.* Chicago: Open Court Publishing, 1916

[34] Parsons, Charles (2008). *Mathematical Thought and Its Objects.* Cambridge: Cambridge UP

[35] Plato (2002). *Meno.* English translation by G.M.A. Grube in *Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo.* Indianapolis: Hackett. pp. 58-92

[36] Plato (2002). *Phaedo.* English translation by G.M.A. Grube in *Five Dialogues: Euthyphro, Apology, Crito, Meno, Phaedo.* Indianapolis: Hackett. pp. 93-154

[37] Putnam, Hilary (1980). "Models and Reality." *The Journal of Symbolic Logic* Vol. 45, No. 3, pp. 464-482

[38] Spinoza, Baruch (1677). *Ethics.* English translation by S. Shirley in *The Essential Spinoza: Ethics And Related Writings*, Ed. by Michael L. Morgan. Indianapolis: Hackett, 2006. pp. 1-306

[39] Tarski, Alfred (1944). "The Semantic Conception of Truth: and the Foundations of Semantics." *Philosophy and Phenomenological Research*, Vol. 4, No. 3, pp. 341-376

[40] Thau, Michael and Ben Caplan (2001). "What's Puzzling Gottlob Frege?" *Canadian Journal of Philosophy* Vol. 31, No. 2, 00 159-200

[41] Van Heijenoort, Jean (1967). *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*. Cambridge, Mass: Harvard UP

[42] Wright, Crispin (1983). *Frege's Conception of Numbers as Objects*. Aberdeen: Aberdeen UP.

[43] Zermelo, Ernst (1930). "On Boundary Numbers and Domains of Sets: New Investigations in the Foundations of Set Theory" in Ewald 1996. pp. 1219-1233