1

Frege's Theorem: An Overview

Frege was incredibly single-minded. Almost all of his intellectual work is devoted, in one way or another, to the attempt to establish logicism: the view that arithmetical truths are ultimately logical truths. As is now widely appreciated, thanks largely to Paul Benacerraf (1995), this project had both philosophical and mathematical aspects. The philosophical part involved analyzing the basic arithmetical concepts, so as to arrive at definitions of them that could be used in the mathematical part, which consisted of an attempt to prove the central theorems of arithmetic from those definitions, using nothing but logical means of inference.

The project was first announced in *Begriffsschrift*, and the innovations for which that book is celebrated are all in service of the logicist project. As Frege writes:

> My initial step was to attempt to reduce the concept of ordering in a sequence to that of *logical* consequence,[1] so as to proceed from there to the concept of number. To prevent anything intuitive from penetrating here unnoticed, I had to bend every effort to keep the chain of inferences free of gaps. In attempting to comply with this requirement in the strictest possible way, I found the inadequacy of language to be an obstacle; no matter how unwieldy the expressions I was ready to accept, I was less and less able, as the relations became more and more complex, to attain the precision that my purpose required. This deficiency led me to the idea of the present begriffsschrift. (Bg, pp. 5–6, emphasis in original)

The required logic is put in place in the first two parts of the book. Frege devotes the third to developing the mentioned account of "ordering in a sequence", which is what we now call the theory of the ancestral.

Things must then have moved quickly. Just three years later, Frege says in a letter to the philosopher Anton Marty that he has "nearly completed a book in which [he] treat[s] the concept of number and demonstrates that the first principles of computation... can be proved from definitions by means of logical laws alone..." (PMC, pp. 99–100). Carl Stumpf, who was a colleague of Marty's, writes back (either on his behalf, or else in response to a similar letter) and urges Frege first to publish a prose version, rather than to publish another book filled with strange symbols (PMC, p. 172). Two years later, *Die Grundlagen der Arithmetik*

[1] It is off our main line, so I will not pursue this question, but does Frege really mean what he says here? What one would suppose he meant is: that he attempted to give a purely logical account of the notion of ordering in a sequence. But that is not at all what he says.

appears. This contains the philosophical work, in particular, Frege's analyses of the concept of cardinal[2] number, of the number zero, of the relation between a number and the one following it (known as predecession or, conversely, succession), and of the concept of a natural (or finite) number. It also contains sketches of proofs, from definitions based on those analyses, of several fundamental facts about the natural numbers, for example, that the relation of predecession is one-one and, crucially, that every natural number has a successor. This last is the most difficult and most important since, given the rest, it implies that there are infinitely many natural numbers, and the need to prove the infinity of the number-series is what makes the project hard in the first place.

The formal presentation of these proofs would not appear for another nine years, when the first volume of *Grundgesetze der Arithmetik* was published, in 1893. I discuss the reasons for the long delay elsewhere (Heck, 2005); they involve changes in the underlying logic itself. But, with one notable exception, discussed in Chapter 3, the formal proofs in *Grundgesetze* follow the sketch in *Die Grundlagen* quite closely. So Frege might naturally have regarded the logicist project as more or less complete at that point.

Unfortunately, just as a second volume was about to appear—it was intended to tie up some loose ends and begin extending the project to real analysis—Frege received a now famous letter from Bertrand Russell. The gist was that (what we know as) Russell's Paradox can be derived in Frege's logic from the 'Basic Laws' with which the proofs all begin. The culprit is Basic Law V (five), which, for present purposes, may be stated in the simplified form:

$$\hat{x}(Fx) = \hat{x}(Gx) \;\equiv\; \forall x(Fx \equiv Gx)$$

A term of the form "$\hat{x}(\phi x)$" is to denote the 'extension' of the concept ϕ. The Law then says that two concepts have the same extension just in case the same objects fall under them. That every concept has an extension is treated as a logical truth, one that is a consequence of the fact that, in Frege's logic, function symbols denote total functions.

One of the nice things about Law V is that it makes it particularly easy to define the notion of membership:

$$a \in b \;\equiv\; \exists F[b = \hat{x}(Fx) \wedge Fa]$$

That is: a is in b just in case b is the extension of a concept under which a falls. And so we have immediately:

$$a \in \hat{x}(Fx) \;\equiv\; Fa$$

[2] I shall omit this qualifier except when necessary, as cardinal numbers will be the main focus of our discussion.

which is (a simplified form of) Theorem 1 of *Grundgesetze*. But then Russell's Paradox is almost immediate. Just take Fx to be: $x \notin x$, and a to be $\hat{x}(x \notin x)$. Then we have:

$$\hat{x}(x \notin x) \in \hat{x}(x \notin x) \equiv \hat{x}(x \notin x) \notin \hat{x}(x \notin x)$$

which is straightforwardly contradictory.

Frege saw the problem right away and attempted to respond to it in an appendix to the second volume. Having offered a solution, he notes that "... it will be necessary to check thoroughly all propositions discovered up to this point" (Gg, v. II, p. 265) to make sure that the changes to Law V do not undermine the proofs. Busy with other work, however, Frege does not seem to have made this check for a few years. But, on or about 5 August 1906 (WMR), he seems to have realized that the effort was hopeless. He would publish nothing for the next ten years. The three further papers he did eventually publish are concerned more with philosophy of language and mind than with philosophy of mathematics and logic.

Logicism, at least Frege's version of it, was thus a casualty of Russell's Paradox. The End.

And that was the end of the story until not very long ago. But the roots of a very different story were planted already in 1955. In "Class and Concept", which is generally devoted to clarifying the relation between those two notions, Peter Geach insists that, for Frege, "identifying numbers with certain extensions was both open to question and ... of altogether secondary importance". Indeed, Geach denies that numbers should be defined as classes (or extensions). More interestingly, though, for our purposes, Geach claims to be able to prove the infinity of the series of natural numbers without relying upon "any special set theory" (Geach, 1955, p. 569) and says that he hopes elsewhere "to explain how the infinite series of natural numbers is generated" (Geach, 1955, p. 570). So far as I know, however, no such paper ever appeared.[3]

Similar but much more explicit remarks are contained in Charles Parsons's paper "Frege's Theory of Number", which was published in 1965.[4] Parsons pays a good deal of attention to a principle he labels "(A)" but

[3]There are two other places that Geach comments on this matter. In his review of Austin's translation of *Die Grundlagen*, he remarks that "rejection of [Frege's view that numbers are classes] would ruin the symbolic structure of his *Grundgesetze*, but not shake the foundations of arithmetic laid down in the *Grundlagen*" (Geach, 1951, p. 541). But it is just not clear whether he meant Frege's proof of the infinity of the number series or, instead, the philosophical work done elsewhere in the book, since the latter is Geach's main focus. In his essay on Frege in *Three Philosophers*, Geach similarly remarks: "... Frege explicitly states that the identification of numbers with certain extensions is only a secondary and doubtful point, and in stating his theory of numbers I shall ignore extensions altogether" (Anscombe and Geach, 1961, p. 158). But he says almost nothing there about Frege's proofs.

[4]Parsons cites both "Class and Concept" and *Three Philosophers*, and he has remarked recently that he was much influenced by Geach's discussions.

which we shall be calling HP:[5]

$$\mathrm{N}x : Fx = \mathrm{N}x : Gx \;\equiv\; \mathrm{Eq}_x(Fx, Gx)$$

Here "$\mathrm{N}x : Fx$" abbreviates: the number of Fs; "$\mathrm{Eq}_x(Fx, Gx)$" abbreviates: the Fs and the Gs are 'equinumerous', that is, can be put into one-one correspondence. So HP says that the number of Fs is the same as the number of Gs if, and only if, the Fs and Gs are equinumerous. Parsons, even more so than Geach, emphasizes how central HP is to Frege's philosophy of arithmetic: It is the core of his analysis of number. And Parsons is the first, so far as I know, explicitly to note that, although extensions of concepts are needed for Frege's definition of numbers, they are not needed for the derivation of axioms for arithmetic from HP. Parsons remarks, almost in passing, that "... the argument could be carried out by taking [HP] as an axiom" (Parsons, 1995a, p. 198).

Though the name itself is due to George Boolos (1998k, p. 268), Parsons was thus the first explicitly to state what we now know as Frege's Theorem: Axioms for arithmetic can be derived, in second-order logic, from HP and Frege's definitions of zero, predecession, and natural number. Parsons does not, however, give a proof of Frege's Theorem. I thought until recently that Parsons simply deferred to Frege at this point, claiming that Frege's own proofs in *Die Grundlagen* have that structure already: Frege himself dervies HP from his explicit definition of numbers in terms of extensions, and then derives the arithmetical axioms from HP without making any further use of extensions. But Parsons makes no such claim; I had misremembered. And, for reasons we shall explore later, one certainly could not just have gestured at the formal arguments in *Grundgesetze*. Indeed, Goran Sundholm flatly dismisses Frege's "pious hope in [*Die Grundlagen*] to avoid the use of" extensions as "unrealistic" (Sundholm, 2001, p. 61), on the ground that at least some of his proofs—in particular, that of Theorem 263—require reference to extensions. Sundholm is wrong about this—see the postscript to Chapter 2—but showing that he is wrong requires a good deal of work.

The first published proof of Frege's Theorem would not appear until 1983, in Crispin Wright's book *Frege's Conception of Numbers as Objects* (Wright, 1983). This point alone bears some emphasis. Although initiates can nowadays rehearse the proof from memory, and we understand very well how it works, it has taken us some years to reach this stage. Indeed, Wright's own struggles illustrate this point. The idea that we might derive the axioms from HP had occurred to him, independently, by 1969.[6] But, as Wright wrote to me recently, when he attempted to write up the proof, he ran into trouble. It is not so easy to formalize the proof-sketches in *Die Grundlagen*. There is at least an apparent reference to

[5]See the editorial notes for some remarks on the terminology.

[6]*Frege's Conception* developed from Wright's B. Phil. thesis, which was submitted that year.

extensions that must be excised, though that is not hard to do. What is more problematic is the fact that Frege's proof of the crucial theorem that every number has a successor is flawed, and, as is argued in Chapter 3, there is no faithful adaptation of Frege's argument that will prove what he wanted to prove. A mere gesture towards *Die Grundlagen* therefore would not have been sufficient, even if Parsons had made it, and any claim that so-and-so had a proof at such-and-such a time has to be regarded as provisional, absent evidence that the *i*s really had been dotted and the *t*s really had been crossed.

In fact, the proof in *Frege's Conception* is not quite complete, and Wright comes harrowingly close to making the same mistake that Frege makes in *Die Grundlagen*. Let "$\mathbb{N}x$" mean: x is a natural number; and "Pab" mean: a immediately precedes b in the natural series of numbers (i.e., in the finite case, $b = a + 1$). Then, adapting his notation to ours, Wright takes induction to be:

$$F0 \wedge \forall x \forall y(Fx \wedge Pxy \rightarrow Fy) \rightarrow \forall x(\mathbb{N}x \rightarrow Fx)$$

rather than:

$$F0 \wedge \forall x \forall y(\mathbb{N}x \wedge Fx \wedge Pxy \rightarrow Fy) \rightarrow \forall x(\mathbb{N}x \rightarrow Fx)$$

and the former is, in principle, weaker.[7] However, due to an otherwise minor difference between their proofs, this oversight does not infect Wright's proof of the existence of successors, as it does infect Frege's.[8] What is worth noting, however, is that, once one has appreciated this difference, one can no longer say, as Wright does (and as I occasionally have, too), that Frege's "account of the ancestral has made it possible... to *define* the natural numbers as entities for which induction holds..." (Wright, 1983, p. 161, emphasis in original). What Frege's definition yields is just the weaker principle, whereas induction is the stronger one, and the proof of the stronger principle, though not technically difficult, requires significant logical resources. In particular, it requires impredicative comprehension (see Chapter 12).

Who first proved Frege's Theorem is not, however, a terribly interesting question, except in so far as we are asking whether Frege himself

[7]Saying in precisely what sense it is weaker is not so easy. But here is an example of what I have in mind. In "Ramified Frege Arithmetic" (Heck, 2011), I show that, in more or less Frege's way, we can prove the existence of successors in ramified predicative second-order logic plus HP. We have induction in the first form quite trivially, due to how $\mathbb{N}$ is defined. We cannot prove induction in the second form, since that would give us full PA, which we certainly cannot have. Formally, the crucial point is that interpreting induction means interpreting it with all quantifiers relativized to "$\mathbb{N}$".

[8]Instead of trying to show that $\mathrm{N}y : P^{*=}yx$ is always the successor of x, as Frege does, Wright shows that $\mathrm{N}y : (\mathbb{N}y \wedge P^{*=}yx)$ is. (See his Lemma 5.) But then, in the proof of Lemma 512, we can assume that $\mathrm{N}y : (\mathbb{N}y \wedge P^{*=}yx)$ is non-zero, since otherwise we are in Lemma 511. And if it is non-zero, then, for some z, $\mathbb{N}z \wedge P^{*=}zx$, and then $\mathbb{N}x$, by transitivity. This very move is made, not quite explicitly, at the top of p. 163. That therefore gives us yet another way to patch Frege's proof, though it is, again, clearly not what Frege had in mind.

proved it: a question to which we shall turn shortly. But, whoever first proved it, Wright is undoubtedly responsible for the activity that has surrounded Frege's Theorem for the last quarter century or so. If Russell's Paradox, or some other contradiction, is forthcoming from HP itself, then who cares if we can replace Law V with HP? Parsons does not raise the question whether HP is consistent. Wright, however, argues in some detail that nothing like Russell's Paradox can be derived from HP[9] and then proceeds to conjecture that HP is, in fact, consistent (Wright, 1983, pp. 154–8). Once the conjecture had been made, it was quickly proved, independently, by several different people.[10] Indeed, Geach (1976, pp. 446–7) had observed a few years earlier that it has a simple model: Let the domain comprise the natural numbers plus $\aleph_0$, which is the number of natural numbers. This is a countably infinite set, and it is well-known that every subset of a countable set is countable and that countable sets are equinumerous if and only if they have the same cardinality. We can thus take the extension of the cardinality operator $\mathrm{N}x : \phi x$ to be the set of pairs $<S, n>$, where S is a subset of the domain and n is its cardinality. That verifies HP. The argument can easily be formalized in, say, Zermelo set-theory, so the consistency of HP is thereby assured, if we accept the axioms of Zermelo set-theory.

That is enough to establish Frege's Theorem as a legitimate piece of mathematics, but the mathematical result was never Wright's goal. Rather, Frege's Theorem was to be the basis on which Frege's logicism might be resurrected, and it was Wright's insistence on the philosophical significance of Frege's Theorem that caused all the excitement. The project had already been announced in Wright's B. Phil. thesis:

> I wish... to vindicate Arithmetical logicism in the following form—not Frege's conception, it should be noted, but yielded by points implicit in Frege's work which in my view deserve acceptance: starting out with logical notions only, an explanation of the concept of Natural Number can be achieved (following a familiar pattern) of such a kind that foundational arithmetical truths (Peano's axioms) can be seen to be logical consequences of the explanation we adopt. It may be that this is sufficient to satisfy Frege's contention that the truths of arithmetic are analytic, even if they are not definitional equivalents of logical consequences of logical truths, as logicism is usually understood. (Wright, 1969, p. 92)

This is essentially the view labeled Number-theoretic Logicism [III], in *Frege's Conception* (Wright, 1983, p. 153). It is nowadays known as Neo-Logicism.[11]

[9] A similar observation, though concerning a slightly different paradox, is made by Geach (1951, p. 542).

[10] These included Harold Hodes (1981, p. 138), John P. Burgess (1984), and Allen Hazen (1985). Burgess gives the model I am about to describe; Hazen claims, but does not prove, that HP is interpretable in second-order arithmetic; the first published proof of that fact was by Boolos (1998b); Hodes claims the consistency (and even truth) of HP, but without proof.

[11] Or, sometimes, Scottish Neo-Logicism, since there are other forms. The modifier "Scot-

So the idea was roughly this. We may regard HP as "explaining" the concept of number. Such explanations are sufficiently akin to definitions that HP may be regarded as analytic of the concept of number and so as analytic in roughly the sense in which definitional truths are. But then Frege's Theorem shows that the axioms of arithmetic are logical consequences of an analytic truth and so are analytic themselves. Surely that is enough to vindicate the epistemological core of Frege's project, if not to vindicate logicism strictly so called.

Every aspect of this program has proved controversial.

Wright is not explicit, in *Frege's Conception*, about the background logic he is using in his proof of Frege's Theorem. The logic is obviously second-order, in that we are quantifying over relations and the like. But Wright does not specify a particular version of second-order logic as his background logic, preferring instead to identify the results we need our "logic of relations" to deliver (see, e.g., p. 163). It quickly became common practice, however, to assume that the proof was being given in full second-order logic, perhaps because that was the logic with which Frege himself worked. But it is famously controversial whether second-order 'logic' is properly so-called, and that means that it is controversial whether the proof of Frege's Theorem shows that the arithmetical axioms really are *logical* consequences of HP. If not, then we have no reason to regard the axioms as analytic. Some of the work collected in the present volume is concerned with this issue, especially Chapter 12, and to a lesser extent Chapter 7. We'll return to the matter below.

What has attracted the most attention, however, is Wright's claim that HP itself may be regarded as 'analytic'. It has never been entirely clear what notion of analyticity is supposed to be in play here. But what has been clear, or so I have generally supposed, is that the notion is supposed to be broadly epistemological in character, rather than metaphysical. In particular, the notion of analyticity in play is not that familiar from positivism and from Quine's criticisms thereof: a notion of 'truth in virtue of meaning'. Whereas the variety of logicism one finds in positivism is in part motivated by ontological doubts about mathematical objects, Wright's version, like Frege's own, was intended to reveal those doubts as unfounded. We are to imagine a subject, conventionally known as "Hero",[12] who is capable of second-order reasoning, but who is otherwise innocent of arithmetical concepts. Now imagine that Hero stumbles upon HP in a dream—how he discovers it is of no epistemic significance—and, impressed by its beauty, decides henceforth to use expressions of the

---

tish" reflects the fact that both Wright and Bob Hale, who soon became the second most vigorous defender of the view, were at the University of St Andrews when things were heating up. No other form of logicism will be under discussion here, though, so I shall speak simply of Neo-Logicism, meaning no disrespect to other forms.

[12]This sort of story first appears in *Frege's Conception* (Wright, 1983, pp. 141ff) and then re-appears in "On the Harmless Impredicativity of $N^{=}$ (Hume's Principle)" (Wright, 2001c). The name "Hero" is borrowed from a similar tale told by Gareth Evans (1985b).

form "$\mathrm{N}x : \phi x$" as subject only to it. Then the claim is that Hero thereby acquires a concept of number. And, crucially, Wright claims that Hero can commit himself to HP, and to using names of numbers in accord with it, without epistemological presupposition. In particular, Hero does not need antecedently to make sure that there are such objects as numbers for such terms as "$\mathrm{N}x : \phi x$" to denote. Indeed, as Wright emphasizes repeatedly, the demand is incoherent: Prior to his committing himself to HP, Hero has no concept of number and so cannot make sense of the question whether there are such objects.[13] Very roughly speaking, then, the thought is that HP is self-justifying: We need no justification to believe it.[14]

Most of *Frege's Conception*, and much of Wright's work on these issues since, is devoted to defending this conception of reference to abstract entities. I have been deeply sympathetic to it since I first encountered it in 1986, when Sir Michael Dummett, who was then my B. Phil. supervisor, suggested we read *Frege's Conception* in tutorial. By the time I arrived at MIT in 1987, I was actively working on the defense of the Fregean account of reference to abstracta, and two of the three papers that comprised my Ph. D. dissertation were directed at such issues.[15] And by luck, or manifest destiny, George Boolos just happened to be at MIT. Boolos had taken a substantial interest in both the mathematical and the philosophical aspects of Frege's Theorem before I arrived,[16] and that of course fueled my own interest. Eventually, he became my primary Ph. D. supervisor.

Boolos, in those days, was particularly concerned about what has come to be called the 'bad company' objection.[17] At that time, my own discussions of the Fregean approach to abstracta, like Wright's, was focused on what are now called 'first-order' abstraction principles, the most familiar of these being the abstraction for directions that Frege discusses in §§64–8 of *Die Grundlagen*:

$$\mathbf{dir}(a) = \mathbf{dir}(b) \;\equiv\; a \parallel b$$

First-order abstractions are, at least potentially, ontologically innocent: Nothing in the abstraction principle itself prevents one from identifying the direction of the line a with some representative line that has that

[13]To put it differently: It may be an intelligible question whether we, or Hero, should make use of a concept of number, but we cannot understand that question as: Are there such things as numbers? If we are not going to make use of that concept, then we cannot intelligibly ask that question. This sort of point of course goes back to Carnap (1950).

[14]Or, to put the point in the sort of language Wright might now prefer: We need to do nothing at all to earn an epistemic entititlement to believe HP (Wright, 2004).

[15]Versions of those papers appear here as Chapter 8 and Chapter 9.

[16]Indeed, I first met Boolos in the spring of 1987, while I was still in Oxford. He presented "The Consistency of Frege's *Foundations of Arithmetic*" as a lecture in 10 Merton Street.

[17]Early in his career, Boolos had been a logicist sympathizer (Boolos, 1972). Moreover, he was by then well-known for his vigorous defense of second-order logic's claim to that title (Boolos, 1998g, l). So Boolos was profoundly intrigued by Wright's position, even though he was not at all sympathetic to it.

direction, perhaps the line parallel to a that passes through some arbitrarily chosen origin. The possibility of such an identification guarantees the consistency of first-order abstractions. Second-order abstraction principles, however, can be ontologically inflationary, HP being a case in point: If, before committing himself to HP, Hero was capable of thinking of only finitely many objects, then there is no way that he can identify all the objects to which his new cognitive resources give him access with ones he could previously apprehend. What is worse, second-order abstractions can be inconsistent, Frege's own Basic Law V being but one of many salient examples.

The bad company objection, then, is that the general defense of the Fregean apporach to abstracta, if intended to include second-order abstraction, must prove too much. Surely Hero cannot simply commit himself to Basic Law V and thereby acquire a warranted belief in its truth. And, Boolos asked me one day, does Hero even have a way to tell if the principle to which he is committing himself is consistent? My very first publication, which appears here as Chapter 10, was my answer to that question. It develops a virulent form of the bad company objection.

But it was a different question Boolos asked me that has most shaped my work on Frege's Theorem.

1.1 Frege on Frege's Theorem

Dummett's long-awaited *Frege: Philosophy of Mathematics* was published in the spring of 1991. I remember seeing it in a bookstore in Amherst, Massachusetts, when I was there for a conference, and buying it excitedly. I set to reading it as soon as I got home. Many of my conversations with Boolos were soon focused on *FPM*. One day, Boolos had a particularly pressing question concerning the following passage:

> Crispin Wright devotes a whole section of his book... to demonstrating that, if we were to take [HP] as an implicit or contextual definition of the cardinality operator, we could still derive all the same theorems as Frege does. He could have achieved the same result with less trouble by observing that Frege himself gives just such a derivation of those theorems. He derives them from [HP], with no further appeal to his explicit definition. (Dummett, 1991b, p. 123)

Dummett seems to be claiming that Frege himself had given a proof of Frege's Theorem. But where is that proof? Boolos had shown that the proof in *Die Grundlagen* can be reconstructed as a proof of Frege's Theorem (Boolos, 1998b), but the reconstruction is not trivial.[18] As I have already said, then, dismissing Wright's contribution with but a gesture at *Die Grundlagen* would have been worse than uncharitable. In any event, the context of Dummett's remark makes it plausible that the "derivation"

[18] We did not know then just how non-trivial: See Chapter 3 for that story.

of which he speaks is not that in *Die Grundlagen* but the one in *Grundgesetze*. So it looks as if Dummett is claiming that Frege proves Frege's Theorem in *Grundgesetze*. Boolos's question was: Is that true?

My dissertation had just been finished, so I had some time on my hands and quickly set to reading *Grundgesetze*. I soon discovered that, if Dummett's claim could be defended at all, it was going to take work. It was easy enough to verify that, after proving HP, Frege makes, as Dummett had said, "no further appeal to his explicit definition", but that simply does not show that Frege proved Frege's Theorem. The crucial question is whether Frege makes no further appeal *to Basic Law V*, and, strictly speaking, he most certainly does.

In *Grundgesetze*, Frege speaks not of extensions of concepts but, more generally, of the value-ranges of functions. The value-range of a function can be compared to its graph, in the set-theoretic sense: a set of ordered pairs of arguments and values. Frege does not, however, define value-ranges in terms of ordered pairs.[19] Rather, he regards the notion of a value-range as primitive, and he characterizes value-ranges simply by saying that two functions will have the same value-range just in case they have the same values for the same arguments. Where "$\grave{\epsilon}F\epsilon$" means: the value-range of the function $F\xi$, then, what Frege says is that "$\grave{\epsilon}F\epsilon = \grave{\epsilon}G\epsilon$" will be true just in case "$\forall x(Fx = Gx)$" is true (Gg, v. I, §3). Clearly, this leads quickly to Basic Law V:

$$(\grave{\epsilon}F\epsilon = \grave{\epsilon}G\epsilon) \;=\; \forall x(Fx = Gx)$$

It is the identity-sign that occurs in the middle here, rather than "$\equiv$", because, for Frege, the truth-values are objects like any others. Consequently, since concepts are functions from objects to truth-values, concepts too have value-ranges, and it is easy to see that two concepts will have the same value-range just in case they have the same extension, that is, have the value Truth for all the same arguments.

Terms denoting value-ranges appear throughout Frege's proofs. Almost every theorem in *Grundgesetze* depends upon the previously mentioned Theorem 1, which, as said earlier, leads directly to Russell's Paradox.

Nevertheless, it was clear from the outset that many of the uses Frege makes of value-ranges can easily be eliminated. Frege almost never quantifies over concepts, for example, preferring instead to quantify over their extensions.[20] So we find things like:

$$\forall f(\ldots a \in f \ldots)$$

[19] It's an interesting question whether, perhaps, he once did do so. As Sundholm (2001, pp. 60ff) notes, pairs seem to have played an important role in early versions of *Grundgesetze*.

[20] Is it a mere curiosity that Frege quantifies over concepts themselves in the definition of the ancestral? Is it just more convenient, technically speaking, to do so? Or is there more to be said on this point?

rather than things like:

$$\forall F(\ldots Fa \ldots)$$

But this is easily remedied, and it seemed likely that other uses of value-ranges would prove to no be more problematic. I therefore set out—in part because reading Frege's notation was then still a chore—to produce a complete translation of Frege's proofs into modern notation, including these sorts of mechanical 'corrections'. When I was done, I could answer Boolos's question: Frege had proved Frege's Theorem in *Grundgesetze*, *modulo* this sort of mechanical elimination of the use of extensions.

This reconstruction of Frege's proofs would appear in "The Development of Arithmetic in Frege's *Grundgesetze der Arithmetik*", which is reprinted here as Chapter 2. In particular, Section 2.1 contains a detailed discussion of how Frege uses value-ranges and Law V in his proofs of the arithmetical axioms; the central parts of the paper contain a commentary on those proofs. But the mere fact that Frege's proofs can be mechanically translated into ones in 'Frege Arithmetic'—second-order logic plus HP—does not, of course, show that Frege *knew* that arithmetic could be derived from HP. I mention some reasons to think he did in Section 2.7, but a more sustained discussion of the question was always necessary, and it appears here in Section 4.3. In brief, the philosophical significance Frege attributes to his proof of the axioms requires that those proofs depend only upon HP and not also upon Law V. It was important to Frege, for philosophical reasons, that his proof of the arithmetical axioms did not depend essentially either upon his explicit definition of numbers in terms of extensions or upon any other sort of appeal to extensions, either.

If Frege did know that arithmetic can be derived from HP, however, then the next question is: Why didn't Frege himself consider retreating from Law V to HP when confronted with Russell's Paradox? In fact, it turns out that he did consider this option. In one of his letters to Russell, Frege writes:

> We can also try the following expedient, and I hinted at this in my *Foundations of Arithmetic*. If we have a relation $\Phi(\xi, \eta)$ for which the following propositions hold: (1) from $\Phi(a, b)$ we can infer $\Phi(b, a)$, and (2) from $\Phi(a, b)$ and $\Phi(b, c)$ we can infer $\Phi(a, c)$; then this relation can be transformed into an equality (identity), and $\Phi(a, b)$ can be replaced by writing, e.g., "$\S a = \S b$". If the relation is, e.g., that of geometrical similarity, then "*a* is similar to *b*" can be replaced by saying "the shape of *a* is the same as the shape of *b*". This is perhaps what you call "definition by abstraction".

What Frege is discussing here is precisely the possibility of replacing the explicit definition of numbers as extensions with HP. He concludes:

> But the difficulties here are [][21] the same as in transforming the generality of an identity into an identity of value-ranges. (PMC, p. 141)

[21] At this point, the translation inexplicably contains the word "not", which is not found in Frege's original letter.

Note how this language echoes the language earlier in the passage, where Frege speaks of the relation Φ's being "transfomed into an equality". There is thus an explicit comparision between "transforming the generality of an identity into an identity of value-ranges"—that is, moving from "$\forall x(Fx = Gx)$" to "$\grave{\epsilon}F\epsilon = \grave{\epsilon}G\epsilon$"—and the transition from "$\mathrm{Eq}_x(Fx, Gx)$" to "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$". The same "difficulties" affect both moves.

Frege certainly is not saying here that HP too is inconsistent.[22] So these "difficulties" are independent of the contradiction, and they apply as much to HP and numbers as to Law V and value-ranges. What are they? Frege mentions them right before the two passages quoted in the last paragraph:

> I myself was long reluctant to recognize the existence of value-ranges and hence classes; but I saw no other possibility of placing arithmetic on a logical foundation. But the question is, How do we apprehend logical objects? And I have found no other answer to it than this, We apprehend them as extensions of concepts, or more generally, as value-ranges of functions. I have always been aware that there were difficulties with this, and your discovery of the contradiction has added to them; but what other way is there? (PMC, pp. 140–1)

The question Frege raises here, how we apprehend logical objects, is the same question that opens §62 of *Die Grundlagen*: "How, then, are numbers to be given to us, if we cannot have any ideas or intuitions of them?" And the answer to which he here commits himself, that "[w]e apprehend [logical objects] as extensions of concepts", is the answer he had reached in §68, where numbers had been defined as extensions of concepts. He had settled on that definition because a very different proposal, made in §63, had been found wanting.

That proposal was that we apprehend numbers as 'abstracts', that is, as the referents of expressions introduced by abstraction, in this case, by HP. And what was the fatal problem with that proposal? It was the "third doubt" introduced in §66. It is known, because of the example used when it is first raised, in §56, as the "Caesar problem":

> In the proposition
>
> "the direction of a is identical with the direction of b"
>
> the direction of a plays the part of an object, and our definition affords us a means of recognizing this object as the same again, in case it should happen to crop up in some other guise, say as the direction of b. But this means does not provide for all cases. It will not, for instance, decide for us whether England is the direction of the Earth's axis—if I may be forgiven an example which looks non-sensical. Naturally no one is going to confuse England with the direction of the Earth's axis; but that is no thanks to our definition of direction. (Gl, §66)

[22] Frege is talking about abstraction principles quite generally, and the examples he gives are first-order and provably consistent. The direction abstraction, in particular, is equivalent to the introduction of points at infinity into Euclidean geometry. That construction, which was well-known to Frege (Tappenden, 1995; Wilson, 1995), amounts to a consistency proof.

As said, it is Frege's inability to resolve this problem that forces him to abandon the view that numbers are abstracts in favor of the view that numbers are extensions.

At the time he wrote *Die Grundlagen*, Frege held open the possibility that there might be some other way to respond to the Caesar problem, so that reference to extensions would not be necessary (Gl, §107). In the end, however, as he wrote to Russell, he could find no other response, and so he was, by his own lights, forced to acknowledge the existence of extensions and to define numbers in terms of them.

To understand why Frege did not himself abandon logicism for Neo-logicism, we must thus understand the Caesar problem. Chapters 4–6 are devoted to that task. In the next section, I will try to synthesize their conclusions.

Before I continue, however, I want to emphasize again the epistemological character of Frege's discussion. What forces Frege to abandon the account of numbers as abstracts for the explicit definition of numbers as extensions is, in the first instance, an *epistemological* problem: Frege thinks that the question how we might have a kind of cognitive access to numbers that is neither sensible nor intuitive cannot be answered in terms of abstraction principles. We shall return to the importance of this point.

1.2 The Caesar Problem

1.2.1 Caesar and Value-Ranges

As we saw above, Frege explicitly compares the move from "$\forall x(Fx = Gx)$" to "$\grave{\epsilon}F\epsilon = \grave{\epsilon}G\epsilon$" to the move from "$\mathrm{Eq}_x(Fx, Gx)$" to "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$", saying that the same "difficulties" affect both moves. The central difficulty with the latter, we have seen, is the Caesar problem. In what sense is it supposed to impinge upon the former?

It is easy to get confused about this matter, and I fear that I have succumbed to that confusion from time to time. At the very least, I have not been as clear as I should have been about how the Caesar problem bears upon the question how we apprehend value-ranges as objects. Let me begin by trying to do better.

The most important thing to remember about the Caesar problem is that it arises as an objection to a particular view about how we apprehend logical objects: the view introduced in §62 of *Die Grundlagen*. This view, which I shall call *abstractionism*, is a view about what is required for singular thought about certain sorts of abstract objects.[23] The objects in question are what I have already called "abstracts", but they are more

[23] Compare Hale and Wright's remarks about why it is important to regard "the number of Fs" as a referring expression rather than a definite description (Hale and Wright, 2009a, §1).

commonly known as *types*. Types here are supposed essentially to be 'of' objects of another sort: the tokens of that type. This relationship is reflected in how we most fundamentally refer to a type: as the type instantiated by a given token. So, for example, directions are essentially the directions *of* lines, and the fundamental way to refer to a direction is as the direction of a given line. Abstractionism then comprises two claims: First, that the capacity for singular thought about objects of a given type T derives from and is constituted by an appreciation of the truth-conditions of identity judgments about Ts, where these identity judgments involve the fundamental way of referring to Ts; Second, that the truth-conditions of such identity judgments may be given in terms of an equivalence relation on the tokens, that is, in terms of an *abstraction principle*, a statement of the form: $\phi(a) = \phi(b) \equiv Rab$.

Let me pause to insist that this is the correct way to formulate the view. The more usual formulation is that our ability to refer to (think about, apprehend) types rests upon our acceptance of an abstraction principle for them. But then the question must immediately arise what is so special about abstraction principles, which are just sentences of a particular syntactic form. Why shouldn't our capacity to refer to numbers derive from our acceptance not of HP but of the Dedekind-Peano axioms (Boolos, 1998d, p. 311)? The best answer[24] to this question is that abstraction principles are *not* special. What underlies our capacity for thought about numbers is not our acceptance of HP but our appreciation of the truth-conditions of numerical identities involving the most fundamental means of reference to numbers. What we must 'accept' is not HP itself but a specification of the truth-conditions of such identities that looks like a metalinguistic version of HP: "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$" is true iff "$\mathrm{Eq}_x(Fx, Gx)$" is true. Given that, of course, HP's truth is immediate.[25] But acceptance of HP is not what underlies our capacity for thought about numbers, and there is no similar story to be told about the Dedekind-Peano axioms.

Now, it is clear that abstractionism could as well offer an answer to the question how we apprehend value-ranges. Indeed, Frege almost seems to endorse this view. Here is how he introduces value-ranges in *Grundgesteze*:

> I use the words "the function $\Phi(\xi)$ has the same *value-range* as the function $\Psi(\xi)$" generally to denote the same as the words "the functions $\Phi(\xi)$ and $\Psi(\xi)$ always have the same value for the same argument". (Gg, v. I, §3)

It looks very much as if Frege is here explaining the truth-conditions of identities between value-ranges in terms of the co-extensiveness of the corresponding functions, and taking that to be sufficient to make refer-

[24] Hale and Wright (2009a) discuss this worry in a reply to a paper by MacFarlane (2009), but they do not make the present point.

[25] I do not think this is quite right, actually—see Section 1.2.3—but I am here just outlining the simplest form of abstractionism.

ence to them possible. That would be an abstractionist view about value-ranges. But Frege explicitly disowns that view later in *Grundgesetze*:[26]

> ...[W]e said: If a (first-level) function (of one argument) and another function are such as always to have the same value for the same argument, then we may say instead that the value-range of the first is the same as that of the second. We are then recognizing something common to the two functions, and we call this the value-range of the first function and also the value-range of the second function. We must regard it as a fundamental law of logic that we are justified in this recognizing something common to both, and that accordingly we may transform an equality holding generally into an equation (identity). (Gg, v. II, §146)

Frege is here insisting that the stipulation from §3 is inadequate, on its own, to ground reference to value-ranges. To move from "$\forall x(Fx = Gx)$" to "$\grave{\epsilon}F\epsilon = \grave{\epsilon}G\epsilon$", we must be "justified in... recognizing something common to the two functions", but Frege is here denying that we can earn a right to this recognition simply by changing the shapes of ink marks.[27] If we could, no more would need to be said in defense of the claim that we may transform "$\forall x(Fx = Gx)$" into "$\grave{\epsilon}F\epsilon = \grave{\epsilon}G\epsilon$". But, as a quick look at the broader context will show, Frege desperately wishes he could say more. It is because he can't that he can only insist on the "fundamental law of logic" he needs:[28]

> If there are logical objects at all—and the objects of arithmetic are such objects—then there must also be a means of apprehending, or recognizing, them. This service is performed for us by the fundamental law of logic that permits the transformation of an equality holding generally into an equation. Without such a means a scientific foundation for arithmetic would be impossible. (Gg, v. II, §147)

And it is important to appreciate that the "fundamental law of logic" for whose acceptance Frege is arguing here is not Law V itself. It is, rather, something that is a law of logic in a quite different sense and that serves to *justify* Law V. This law is what justifies our "recognizing something common", so that "accordingly we may transform an equality holding generally into an equation" (Gg, v. II, §146).

There is a sense, then, in which the Caesar problem is still in play in *Grundgesetze*, and there is a sense in which it is not. The Caesar problem *is* in play in the sense that it continues to frustrate abstractionism, so that it prohibits Frege from adopting an abstractionist account of reference to value-ranges, just as it forced him to abandon the abstractionist account of reference to numbers. But, in a different sense, the Caesar problem is *not* in play: Frege does not hold any view to which the Caesar

[26]This passage is from volume II, which was published a decade later. But there is no serious doubt, it seems to me, that it reflects Frege's view in 1893, as well.

[27]It is sometimes claimed that Frege regards "$\forall x(Fx = Gx)$" and "$\grave{\epsilon}F\epsilon = \grave{\epsilon}G\epsilon$" as synonymous, the two being mere stylistic variations of one another, like active and passive. The passage we are discussing is not consistent with such an interpretation.

[28]Note the similarities to the letter to Russell from which I quoted earlier.

problem is an objection, so he does not need to solve it. What he does need is an answer to the question how we apprehend value-ranges as objects, and the sad truth, which he himself recognizes, is that he simply does not have one.

If so, then it does seem fair, in the end, to regard the explicit definition of numbers as extensions as hopeless, quite independently of the inconsistency of Law V. Why, then, did Frege insist upon painting himself into this corner? The answer at which I arrive in Chapter 5 is that, in Frege's day, at least, there were significant dialectical differences between the two positions we have been discussing. The dominant logical tradition against which Frege had himself struggled, the Boolean tradition, regarded extensions of concepts as *the* fundamental materials of logic.[29] It would thus have been entirely reasonable for Frege to expect agreement with his "fundamental law of logic" permitting the move from "$\forall x(Fx = Gx)$" to "$\grave{\epsilon}F\epsilon = \grave{\epsilon}G\epsilon$". For him simply to insist on a law permitting the move from "$\mathrm{Eq}_x(Fx, Gx)$" to "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$", however, would have been for him to beg the very question logicism was meant to answer.

One might object, however, that the Caesar problem must play more of a role in *Grundgesetze* than I am allowing. After all, a form of it seems to arise in §10, where Frege discusses the question whether the truth-values are value-ranges and, if so, which ones they are. But the correct conclusion, for which I argue in Section 5.2, is that the problem under discussion in §10 is not the Caesar problem, not if the 'Caesar problem' is the problem discussed in the central sections of *Die Grundlagen*.

So what is the Caesar problem?

1.2.2 The Caesar Problem as Epistemological

I wish I could say, after umpteen years of thinking about it, that I now understand the Caesar problem and why Frege thought it a fatal objection to abstractionism. Unfortunately, I don't. What I've got are a handful of thoughts about how the Caesar problem bears upon abstractionism, and how it does not. I suspect that Frege himself saw many aspects of the problem only obscurely.

Let me begin by emphasizing again that the Caesar problem arises as an objection to a certain view. Frege famously held, at least in his mature period, that every well-formed sentence must have a determinate truth-value, lest (at least) one part of that sentence not have a reference. The Caesar problem can sound as if it is just an application of this more general view, but, for reasons discussed in Section 6.1, that is wrong. This 'principle of complete determination' does indeed imply that we need to fix a sense for identities like "$0 =$ Caesar", but not any way of doing so counts as resolving the Caesar problem. Frege might have

[29] Might Frege have been led to his acceptance of extensions in part through his reading of Boole?

stipulated truth-values for such statements, much as he stipulates in §10 of *Grundgesetze* that Truth and Falsity are to be their own unit classes, but no such stipulation will save abstractionism. Nor will rejecting the principle of complete determination.

Second, it is important to appreciate that the Caesar problem is less about deciding the truth-values of certain statements than it is about guaranteeing that those statements have a sense. On Frege's view, again, if a statement does have a sense, then that sense is a mode of presentation of a truth-value, and the statement must therefore have a truth-value as well. But the worry is not that we do not know whether Caesar is 0. It is that we do not so much as understand the question whether he is. Or, rather, the problem is that we *do* understand that question, and we *do* know the answer. But the abstractionist account of reference to numbers utterly fails to explain how we might even make sense of it. It explains identity statements of a very particular form, "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$", and no others. Whence our understanding of statements of the form "$t = \mathrm{N}x : Gx$" when t itself is not of the form "$\mathrm{N}x : Fx$"?

This is one of the few things of which I am sure here: that one main lesson of the Caesar problem was supposed to be that there is more to our apprehension of numbers as objects than abstractionism can explain. I do not claim to be the only, or even the first, philosopher to appreciate this fact. The way Hale and Wright approach the Caesar problem, at least in their more recent discussions (Hale and Wright, 2001c, 2008), seems to involve at least an implicit recognition of it. But I do not think the point has been appreciated as widely as it should be. What makes a proper appreciation of it so difficult is that it is hard to keep the epistemological dimensions of the problem firmly in view. It is easy to slip back into thinking that the problem is to fix the truth-values of certain sorts of identities, using whatever resources might be available. But if the question is one about *our* apprehension of numbers as objects and abstractionism's failure to capture it, then the resources used in that account have to be ones of which ordinary folk have some sort of grasp.[30] (If one is inclined to deny that the question is about ordinary folks' apprehension of numbers as objects, see Section 1.3.)

I have been careful here *not* to say that what the Caesar problem requires is that we explain the sense of such claims as "Caesar is zero". It is, after all, a common reaction to the Caesar problem to insist that even

[30]Sider (2007, pp. 220ff) emphasizes a similar point, but then misplays it, or so it seems to me. Sider seems to think that Neo-logicism is a quasi-Cartesian enterprise whose goal is to "dispel doubts about mathematics" and so make it "epistemically secure". If that were the project, then, indeed, everything that went into the argument that abstract entities exist, including the "underlying metaontology", would be part of the justificatory basis for our knowledge of arithmetic, and Sider's complaint that this doesn't look epistemically secure would be justified. But that is just not the project, not as I understand it, anyway, though Sider's charge might well apply if the project were understood the way MacBride (2000, p. 158) suggests it should be. We'll discuss this issue below.

to ask whether Caesar is zero is to commit some kind of category mistake. So it is a possible view that the question whether Caesar is zero is actually unintelligible, and that our denial that he is zero involves something other than an assertion. Saying that Caesar is not a number is thus comparable, in relevant respects, to saying that the concept *horse* is not a concept. Saying that Caesar is not a number is a confused attempt to express as a substantive claim what must ultimately be understood in terms of the differences between number-words and people-words, be these logical, syntactic, or semantic.

Many people probably still think, as I once did, that this view simply isn't available to Neo-logicists. It is widely believed that Frege's proofs of the axioms of arithmetic rest essentially upon his view not just that numbers are objects, but that numbers are objects of the same logical category as concrete objects such as people. Very roughly, the thought is that, when we speak of "the number of Fs", it must be permissible for F to be true both of people and of numbers. Of people (and the like), because our initial understanding of what Frege calls "ascriptions of number" cannot presuppose a familiarity with numbers themselves. The basic instances of HP must concern things like "the number of Roman emperors". But we must also be able to speak of the number of numbers having some property, because the idea behind the proof of the existence of successors is to consider the sequence: $\mathrm{N}x : x \neq x$, $\mathrm{N}x : x = 0$, $\mathrm{N}x : (x = 0 \vee x = 1)$, ... If so, however, then it looks as if the numbers and the people need to be in the same domain, and that will force the Neo-logicist to regard the question whether Caesar is zero as at least intelligible.

This line of thought, perhaps bound up with various sorts of formal considerations,[31] may have played a role in leading Frege to the Caesar problem. But the lesson of Chapter 6, as I now see it, is that there is no extra problem here. Contrary to the common wisdom, Frege's proofs of the axioms of arithmetic—in particular, of the existence of successors—do *not* require there to be a single, universal domain over which all first-order variables range. The proofs do not even require that numbers be objects. If so, then the view that the question whether Caesar is zero commits a category mistake *is* available to Neo-logicists.

Contrary to what Hale and Wright (2001c, p. 346) seem to think, however, I never meant to suggest that simply making this view available somehow solved the Caesar problem in all its manifestations.[32] On the contrary, Chapter 6 opens by identifying the epistemological aspect of the Caesar problem we have been discussing and a semantical aspect we shall discuss shortly, and then suggesting that, *besides* these general

[31] Tappenden (2005) discusses the mathematical background of the Caesar problem.

[32] In footnote 6 of Chapter 6, I say: "Adopting the view discussed below, that numbers are simply a different *sort* from people, will not relieve one of the obligation to explain the origins of speakers' knowledge of this fact." In footnote 25, I insist that the semantic form of the Caesar problem still arises when HP is treated predicatively. These notes were contained in the original paper (Heck, 1997).

problems, which Frege raises against both the direction abstraction and HP, the Caesar problem might "raise quite specific problems in the case of numbers" (this volume, page 130). The point of Chapter 6 is to argue that those more specific problems can be resolved, and so to make room for the kind of response to the epistemological aspect of the Caesar problem just outlined. Resolving the "quite specific problems" was never intended to relieve abstractionists of the need to address the more general ones.

So I agree with Hale and Wright (2001c, p. 350) that "... we cannot avoid confronting something like the question whether numbers are or aren't of the same Sort as people". My goal, as I have been saying, was simply to put that question on the table and thereby to make a particular kind of answer to the Caesar problem available to Neo-logicists: that numbers and people are just radically different sorts of objects. The point of making this view available is that a defense of it, or so it seemed to me, would be able to appeal to quite different resources than views that assume that people and numbers must occupy a single domain. I never claimed to have provided a defense of that view, however, and I certainly was never under any illusion that it somehow didn't need one. But the view cannot just be dismissed with the suggestion that, if numbers are not of the same sort as people, then we might just as well regard "Caesar is zero" as intelligible but false (Hale and Wright, 2001c, p. 351). As Frege well knew, the problem is in no way limited to mixed identity statements: If "Caesar" and "zero" are intersubstitutable *salva significatione*, then "$0 +$ Caesar" must also have a sense and, unless one thinks that "$+$" fails to refer, a reference. Which, pray tell? And if one is inclined to reply that "$+$" is only partially defined, then I ask: Why can't "$=$" be partially defined, too? And what's the difference between saying that these signs are only partially defined and saying that numbers and people are of different sorts?

1.2.3 The Caesar Problem as Semantical

Consider the formulae "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$" and "$\mathrm{Eq}_x(Fx, Gx)$". According to abstractionism, they are supposed to have the same truth-condition. Yet the former is supposed to involve reference to numbers in a way that the latter does not: It is, in particular, supposed to be by our coming to understand the former as equivalent to the latter that we acquire a capacity for singular thought about numbers. But if we really do understand "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$" as an identity statment containing expressions that refer to objects, then that understanding must comprise (or at least make available) an understanding of the complex predicate "$\xi = \mathrm{N}x : Gx$". However, treating "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$" as having the same truth-condition as "$\mathrm{Eq}_x(Fx, Gx)$" does not, by itself, give us the ability to understand the question whether the predicate "$\xi = \mathrm{N}x : Gx$" is true or false of a given object. Of course, if this object is given to us as the number of ϕs, for some ϕ, then all is well. But if it is not given to us in that form, then we are completely lost, and the example of Caesar simply illustrates this point.[33] Indeed, even if the object we are given *is* a number, we cannot understand the question whether "$\xi = \mathrm{N}x : Gx$" is true or false of it unless it is given to us as the number of ϕs. But then the lesson of the Caesar problem appears to be this: If our understanding of names of numbers were adequately explained by abstractionism, then we could not understand the question whether "$\xi = \mathrm{N}x : Gx$" is true or false of *objects* at all, which means that we could not understand the predicate itself, which means that we could not understand "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$" as an identity statement. So, again, abstractionism would fail fully to capture our understanding of names of numbers.

The question whether abstractionism can provide a sense for complex predicates of the form "$\xi = \mathrm{N}x : \phi x$" is particularly pressing for Neologicists, since such predicates play a critical role in Frege's proof of the existence of successors. Frege's idea is to consider the series $\mathrm{N}x : x \neq x$, $\mathrm{N}x : x = 0, \ldots$. But the latter is really:

$$\mathrm{N}x : (x = \mathrm{N}y : y \neq y)$$

and the argument of the outer occurrence of "N" is thus:

$$\xi = \mathrm{N}y : y \neq y$$

If this predicate has not been provided with a sense, Frege's proof of the existence of successors consists of a series of uninterpreted squiggles.

I have introduced this issue in broadly epistemological terms since, as I have emphasized, abstractionism is fundamentally an epistemological doctrine. But the present issue is not epistemological but semantic. What this form of the Caesar problem threatens to show is that we cannot simultaneously both regard "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$" as having the same truth-condition as "$\mathrm{Eq}_x(Fx, Gx)$" and treat "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$" as having the 'logical form' it overtly appears to have.[34]

In the case of first-order abstractions, we can make some progress, if we accept the view, discussed in the previous section, that types are of a different logical category from their tokens. If so, then we will have a special style of variable that ranges over the types, and these variables can be eliminated in favor of variables ranging over tokens. Consider, for example, lines and their directions, and write variables ranging over directions in boldface. Then, quite generally, we can transform "$\ldots \mathbf{d} \ldots$" into "$\ldots \mathrm{dir}(d) \ldots$", trading (both free and bound) occurrences of "**v**" for those of a corresponding variable "v". Repeat as necessary. The end result will then be something from which all occurrences of "$\mathrm{dir}(\xi)$" can be

[33] For defense of that interpretive claim, see p. 129.

[34] Sider (2007, p. 204) voices a similar concern, though he does not connect it to the Caesar problem.

eliminated *via* the abstraction principle (and related principles concerning predicates of directions).

The procedure can be extended to 'predicative' second-order abstractions, where we do not permit embeddings like: $\mathrm{N}x : (x = \mathrm{N}y : \phi y)$. That there be no such embeddings is a form of the requirement that the types introduced by the abstraction be of a different logical sort from their tokens. In this case, however, the tokens—what numbers are 'of'—are concepts, so the requirement becomes: the objects introduced by abstraction must be of a different sort from those of which these concepts are true or false. In this case, then, we can transform "$\ldots \mathbf{n} \ldots$" into "$\ldots \mathrm{N}x : Fx \ldots$" and eliminate the cardinality operator via HP.[35,36]

This approach can easily feel like the most blatant cheating, but there is actually something very natural about it. The basic idea is just that, since directions are fundamentally and essentially *of* lines, specifying the domain over which direction-variables range shouldn't be any harder than pointing at the lines and saying: the directions of those.[37] To the extent that this explanation of variables ranging over types works, however, it works too well. Traditionally, the thought would have been that, by showing us how statements that quantify over directions can be translated into ones that only quantify over lines, the procedure shows us how to *eliminate* quantification over directions. For the reasons given in Section 8.3, I do not think that is the right way to put the point. We have to allow that abstractionism can explain our understanding of a class of abstract *terms*, by which I mean: terms that more or less look like they ought to refer to abstract objects, to types. But what the availability of the translation does show is that abstractionism cannot secure the idea that these terms refer *to abstract objects*. What the availability of the translation shows, that is to say, is that, for all we have said so far, abstractionism is compatible with the view that 'names of types' refer not to types but to representative tokens of the type. For example, a 'name of a direction' might refer to a representative line that has that direction.[38]

So here is another thing I take myself to know: So long as we

[35] Matters are a bit more delicate than I have indicated, since we will need to deal with problems similar to those discussed in Section 8.3. But this can be done in much the way done there.

[36] The restriction to predicative abstractions is essential. If the abstraction is regarded as impredicative, then nothing like this elimination can work. Such a reduction is precisely what Frege attempts in §§29–31 of *Grundgesetze*. Success would have amounted to a consistency proof for Basic Law V (Heck, 1998a). The problem, ultimately, is that we have no way to eliminate occurrences of abstract terms inside second-order variables, like: $G(\mathrm{N}x : Fx)$. In the present case, we can increase the order of "G", so that "$G(\mathrm{N}x : Fx)$" becomes a third-order statement. Clearly, this will not work if we can have something like: $F(\mathrm{N}x : Fx)$, since we will then have no way to keep the two Fs tied together.

[37] This might seem to limit the domain to the types of tokens that actually exist. This is what Hale and Wright (2001a, pp. 422–3) call the "Problem of Plenitude". My own view is that it ceases to be a problem once the strict form of abstractionism we are discussing has been abandoned in favor of a more sophisticated one. We will get to that shortly.

[38] As discussed in Section 9.2, the best version of this view is probably supervaluational.

insist that the truth-condition of "$\mathrm{dir}(a) = \mathrm{dir}(b)$" really is the same as that of "$a \parallel b$", we will be unable to argue that names of directions refer to abstract entities or, to put it in the material mode, that directions are abstract.[39] If so, then abstractionism, in this form, is inadequate as an account of our capacity for singular thought about abstract entities. Thought of the sort abstraction makes possible might just as well be about the concrete. Indeed, this sort of abstractionism looks suspiciously like Berkeley's, and he was no friend of the abstract (see Section 9.1, pages 204–206).

There is another route to much the same conclusion, one that goes through the bad company objection, mentioned earlier. The thought is very simple: If abstraction can make the truth-condition of "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$" the same as that of "$\mathrm{Eq}_x(Fx, Gx)$", then why can't it also make the truth-condition of "$\hat{x}(Fx) = \hat{x}(Gx)$" the same as that of "$\forall x(Fx \equiv Gx)$"? Wright has replied, reasonably enough, that it was never any part of the view that such stipulations must always succeed. In the case of extensions, we might just regard the attempt as a failure (Wright, 2001d, p. 281). The problem then becomes to distinguish the good cases from the bad ones. Anyone familiar with this problem will know, however, that the sorts of conditions that have been proposed are highly complex and, in general, are not ones whose satisfaction is easily determined.[40] That means that we can very easily find ourselves in a position where we do not, and perhaps even cannot, know whether a particular abstraction satisfies the condition proffered and so has succeeded in making it possible for us to refer to objects of whatever kind is at issue.

There are, however, two different ways to understand what failed abstractions fail to do. Wright's various discussions of this issue are most naturally read, it seems to me, as suggesting that failed abstractions do not even specify a *sense* for the expressions they characterize.[41] So, for example, an attempt to introduce extensions by abstraction would, on this view, fail even to give a sense to "$\hat{x}(Fx)$" and so would fail to introduce even the *concept* of an extension. This kind of view is not without precedent. One might compare it to Gareth Evans's views about empty demonstratives. On Evans's view, if one is hallucinating a little green man and attempts to venture the thought *that man is laughing at me*, one fails thereby to think any thought at all (Evans, 1985c). But, for the sorts of reasons given by Gabriel Segal (2000), that view strikes me as indefensible, and similar considerations apply here.

[39] The problem is even worse in the second-order case, since then the representative tokens are concepts. So we cannot even argue that names of numbers refer to objects, that is, that numbers are objects.

[40] See the postscript to Chapter 10 for some of the history.

[41] For example, in one such discussion, Wright (2001d, pp. 281–2) characterizes abstraction as an attempt "to fix a new concept" and says that bad cases "misfire" or "abort". That suggests that no new concept is fixed in those cases.

Change history slightly.[42] Suppose that, instead of abandoning abstractionism, Frege had embraced it for both numbers and extensions but, for reasons of simplicity and convenience, decided to define numbers as extensions anyway. *Grundgesetze* then ends up being pretty much as it is. In particular, the formal arguments are completely unchanged. Question: Are we really to suppose that these arguments would have expressed no thoughts whatsoever? that, when Frege thinks to himself, "So $a \in \hat{x}(Fx) \equiv Fa$", he isn't actually thinking *anything*? Much the same question can of course be raised about the actual Frege and the actual *Grundgesetze*.[43] But the question has special force for the view I have attributed to Wright: the view that abstraction is a legitmate form of concept formation that just happens to misfire in some cases and so fails to introduce any concept at all. The sort of intellectual and cognitive activity abstraction makes possible does not seem to differ between the good cases and the bad ones, at least so long as we do not know that we are in a bad case.[44]

A quite different view would allow that, in both the good and the bad cases, abstraction can fix a *sense* for the introduced names, but that only in the good cases does it manage to provide those names with *reference*. If that is right, however, then it follows immediately that abstraction cannot, in fact, make the truth-condition of "$\hat{x}(Fx) = \hat{x}(Gx)$" the same as that of "$\forall x(Fx \equiv Gx)$" and, by parallel reasoning, cannot make the truth-condition of "$\mathrm{N}x : Fx = \mathrm{N}x : Gx$" the same as that of "$\mathrm{Eq}_x(Fx, Gx)$".

Isn't that conclusion just incompatible with abstractionism? With its original letter, yes, but not necessarily with its spirit. It depends upon how the extra content of "$\hat{x}(Fx) = \hat{x}(Gx)$" is to be understood. The obvious thought is that what distinguishes "$\hat{x}(Fx) = \hat{x}(Gx)$" from "$\forall x(Fx \equiv Gx)$" is that the former is committed to the existence of an extension: the common extension of $F\xi$ and $G\xi$. One way to implement this idea is to regard abstraction as conditioned on the existence of the relevant objects. So, roughly speaking, "$\hat{x}(Fx) = \hat{x}(Gx)$" will mean something like: If $\hat{x}(Fx)$ exists, then $\forall x(Fx \equiv Gx)$. Hartry Field makes a proposal of this sort, and Wright (2001a) objects that it is just incoherent: If the point of the

[42]Another argument here could be based upon the fact that even Law V is not inconsistent unless we accept sufficiently strong comprehension axioms (Heck, 1996). One might respond that the sort of comprehension needed in that case (Π^1_1 comprehension) is also needed for the proof of Frege's Theorem. But it seems plausible that other sorts of examples would require even stronger comprehension axioms (or even choice principles), and nothing stronger than Π^1_1 comprehension is needed for the proof of Frege's Theorem.

[43]Exactly what one might want to say about the actual case depends upon how one understands Frege's introduction of the smooth breathing. There is room, I think, for the view that Frege really does fail to explain it, since he barely gestures at what it is supposed to mean. In so far as we do understand it, perhaps that is because we fall back on the abstractionist explanation Frege himself disowns. If so, that looks like additional evidence in favor of the conception for which I am arguing.

[44]Then, I take it, it's like tic-tac-toe. You know how to play the game, but you can't really play it any more.

abstraction is to introduce (or characterize) the concept of extension, we cannot use that concept in stating the truth-condition of "$\hat{x}(Fx) = \hat{x}(Gx)$". Fair enough. But it would surely be enough to evade this objection if it could be argued that a conception of what it is for extensions to exist was implicit in the abstraction principle itself. And it would be enough, too, if such a conception could be extracted from additional materials similar in character to an abstraction principle.

Let me not try to explain here how that might go. Chapter 9 contains what I have to say about the matter. For present purposes, what matters is just the general idea: that we should seek a view that gives substantial content to the idea that directions, or shapes, or numbers exist as abstract objects (not just as representative tokens), so that the existence of such entities will *not* simply be a consequence of the fact that there are lines, or figures, or concepts. We have to do so, I have argued, if we are to have any hope at all of making sense of how an abstraction principle might fail to be true.

Doing so would also put us in a position to respond to the semantic form of the Caesar problem we were discussing earlier. We saw there that there is a promising strategy by means of which the abstractionist might explain quantification over directions, but that this strategy is equally available to someone like Berkeley, who wants to regard 'thought about directions' not as thought about something abstract but as abstract thought about something concrete. If "$\mathrm{dir}(a) = \mathrm{dir}(b)$" does not have the same truth-condition as "$a \parallel b$", however, then "$\exists d(d = \mathrm{dir}(b))$" does not have the same truth-condition as "$\exists l(l \parallel b)$". Unlike the latter, the former quantifies over, and therefore is committed to the existence of, directions. To say so would be completely unhelpful if we did not understand what such a commitment might involve, and abstractionism in its original form makes such an understanding unavailable in principle: The original view was that one need understand no more than the abstraction principle in order to have a concept of direction, that is, to understand what directions are and so to understand what it means to say that some (or most or all) directions are thus and so. But Chapter 9 is an attempt to provide precisely this missing piece: An account of what else we need to know to understand what it is for directions to exist as abstract objects.

Some will object that, no matter how successful this attempt, it can do nothing to secure the existence of abstracta. Indeed, it might be thought that the broadly epistemological approach I have been developing flatly misunderstands the nature of the problem, which is ontological or metaphysical: No account of our apparent capacity for singular thought about abstract entities, however successful, can do anything to guarantee that we actually succeed in thinking about anything abstract, because no such account can guarantee that the requisite objects exist. But that charge, I counter, misconstrues both the dialectical situation and what we should regard as the legitimate aspirations of philosophy.

First, it is admitted on all sides that ordinary thought is replete with what looks *prima facie* like singular reference to the abstract, and, if we are not to take it at face value, we need to be given reason not to do so. I know of no reasonably convincing route to skepticism about the abstract other than the epistemological one (see Section 8.1). I am not going to insist that the burden of proof here is on the nominalist; I hate burden of proof arguments. Nonetheless, the first order of business must surely be to answer the question how we should conceive of singular thought about entities with which we have no causal interaction. If we can come up with a plausible answer, then it is unclear what reasons might remain for doubt about the abstract.

Second, there have been several attempts in the last few years to articulate a conception of 'meta-ontology' that might serve the goals of abstractionism and so Neo-logicism. The two most prominent are due to Ted Sider and Matti Eklund. Sider (2007) proposes that Neo-logicists should adopt the view that the meanings of quantifiers are not fixed but can vary and be extended as new sorts of 'objects' are introduced. Eklund (2006) proposes instead that Neo-logicists should be ontological 'maximalists', who believe that everything that can exist does exist. But neither alternative looks plausible as an interpretation of extant Neo-logicists, or so Hale and Wright (2009b) themselves have claimed.

From the very outset, Wright's version of Platonism has been based upon a form of quietism—or, better, small-n naturalism—according to which philosophy simply ought not to be in the business of questioning the results of the sciences. Consider, for example, this remark:

> If... certain expressions in a branch of our language function syntactically as singular terms, and descriptive and identity contexts containing them are true *by ordinary criteria*, there is no room for any ulterior failure of 'fit' between those contexts and the structure of the states of affairs which make them true. So there can be no philosophical science of ontology, no well-founded attempt to see past our categories of expression and glimpse the way the world is truly furnished. (Wright, 1983, p. 52, my emphasis)

The emphasized phrase, "by ordinary criteria", is the key to Wright's thinking here. Whether arithmetical claims really are true is simply no part of what he thinks is at issue philosophically. Here's Wright's account of what is:

> For Frege, the question is... how we get into cognitive relations with the states of affairs which make number-theoretic statements true: a question which he rightly saw as calling for a systematic account of the content of those statements, and to which his logicism was offered as an answer. (Wright, 1983, p. 52)

The philosophical issue, then, is whether arithmetical thoughts really do involve singular thought about numbers, as they appear to do, and, if so, how such singular thought should be understood.

So, if the question is supposed to be how we know that there are any directions, or word- and sentence-types, or numbers, the answer is: Not by doing philosophy. Whereas Wright supposes, however, that the existence of such objects is essentially trivial—a consequence of the reflexivity of the equivalence relation that features in the abstraction principle—my view, which is developed in Chapter 9, is that the question whether such objects exist is ultimately a question about the extent to which substantive theorizing about them is possible.

1.2.4 The Existence of Successors

Where, then, does that leave us as regards the success of Neo-logicism?

As was mentioned earlier, the abstractionist strategy for explaining quantification over types simply does not apply when the abstraction principle is understood impredicatively. That means that the strategy does not apply, in particular, to HP, as it must be understood for the proof of Frege's Theorem: We have to be able to "count numbers" for the proof of the existence of successors to go through. The direct approach therefore does not work.

But an indirect approach might yet. As is argued in Section 6.2, the proofs of the other axioms do go through, even if we formulate HP predicatively. So, if the form of abstractionism to which I have committed myself is in fact defensible, then Neo-logicism can claim significant if limited success: We can explain how singular thought about numbers is possible and how some of the most basic laws of arithmetic are implicit in the very nature of such thought. And, as is argued in Section 6.3, construing HP predicatively does not prevent us from counting numbers. Whereas the original form of HP will apply only to concepts true or false of 'basic objects'—e.g., people, or whatever we can think about prior to becoming familiar with numbers—a new form of HP can be formulated that will apply to concepts true or false of numbers. The question is: How should we understand the relation between the number of basic objects that are F and the number of numbers that are ϕ? If we are going to continue thinking of everything predicatively, then these have to be understood as numbers of different kinds, so that the the question whether they are the same is just unintelligible. Clearly, however, that is not how we think, and it is completely obvious when those numbers will be identical: They will be the same just in case there is a one-one correspondence between the Fs and the ϕs.

It is the status of this last move that is critical. With what right, the question must be, do we here loosen the restriction on the predicativity of abstraction? It should be obvious that such loosening will not always be a good idea: A parallel move involving extensions would lead to inconsistency. So it would not be unreasonable to wonder at this point if any real progress has been made, if the goal was a logicism that would encompass the infinity of the numbers. But that need not be the goal, and I would insist, myself, that the epistemological interest of Frege's

Theorem is quite independent of the fortunes of Neo-logicism. As Frege himself makes wonderfully clear, the point of "looking for the fundamental principles or axioms upon which the whole of mathematics rests" is that, once this question has been answered, "it can be hoped to trace successfully the springs of knoweldge on which this science thrives" (PCN, op. 362). Logicism was where Frege hoped this project would lead us, but it is not intrinsic to the project itself. If HP is in some sense the fundamental principle on which our knowledge of arithmetic rests, then that fact is, it seems to me, of tremendous epistemological significance, whatever the epistemological status of HP itself (or of the more sophisticated articulation of it we have just been considering).

Nor, however, do I think that we should despair of a more encompassing logicism. Perhaps the sort of identification that is needed here, between the numbers of basic objects and the numbers of numbers, could be regarded as embodying some sort of claim that, though substantial, needs no independent warrant, but is somehow a conceptual hostage to fortune. I do not know how to develop that vague suggestion, but I think it is worth a look.

1.3 What Does HP Have To Do With Arithmetic?

"If HP is in some sense the fundamental principle on which our knowledge of arithmetic rests...". What does that mean? In what sense might HP ground arithmetical knowledge?

I do not mean to imply, by the way, that Frege's Theorem can be of no *philosophical* significance unless it is of *epistemological* significance. At the end of Chapter 7, I myself suggest, borrowing some suggestions from William Demopoulos (1998, 2000), that Frege's Theorem might feature in an explanation of why the finite cardinal numbers satisfy the Dedekind-Peano axioms. Even in this case, however, it seems clear that the success of such an explanation depends upon HP's being, in some appropriate sense, more fundamental than the Dedekind-Peano axioms and, indeed, upon HP's being *the* fundamental fact about cardinal numbers. Whether it depends, as Demopoulos (1998, p. 483) writes, upon "the thesis that [HP] expresses the preanalytic meaning of assertions of numerical identity" is not so clear.[45]

I doubt, however, that Frege would have been satisfied with only this much. His focus is clearly on an epistemological issue. This is evident at the very beginning of *Die Grundlagen*, where Frege describes his project in terms of epistemological categories that he borrows from the Kantian tradition (Gl, §3), but one might dismiss or re-interpret this allusion.[46]

[45] A similar claim is made by Benacerraf (1995, p. 46), though his is made on behalf of Hempel.

[46] Benacerraf (1995, pp. 53–7) is undoubtedly correct that Frege's own conception of these

One cannot, however, ignore the centrality of the question how numbers are given to us and how Frege uses that question to motivate HP (Gl, §§62ff). My own thinking about the philosophical significance of Frege’s Theorem has thus always focused on the question what, if any, epistemological significance it has. And for *that* purpose, it has always seemed obvious to me that it *is* essential that HP should capture the ordinary meaning of claims of numerical identity: If the question is how numbers *are* given to us,[47] and if the answer is supposed to be that our capacity for singular thought about them is to be explained in terms of our taking the truth-condition of “$\mathrm{N}x : Fx = \mathrm{N}x : Gx$” to be (almost) the same as that of “$\mathrm{Eq}_x(Fx, Gx)$”, then the view just is that HP (or something very like it)[48] is not just implicit in our ordinary thought about numbers but is partially constitutive of our capacity for such thought.

Even if that could be established, nothing yet follows about the epistemological status of our arithmetical knowledge. Forget about analyticity and focus instead on the question whether our arithmetical knowledge is *a priori*. Even if HP is a conceptual truth in the sense that it is implicit in singular thought about numbers, it does not follow from the fact that the Dedekind-Peano axioms can be derived from HP that anyone knows those axioms *a priori*. To use a now familiar distinction, what would follow, at most, is that anyone capable of singular thought about numbers *has justification* for the Dedekind-Peano axioms,[49] not that anyone *is justified* in believing them, let alone justified *a priori*. For that to follow, HP would have to play the right sort of justificatory role with respect to arithmetical knowledge.[50] What sort of justificatory role is not at all clear to me, in part because I am no expert on epistemology. But the question does need to be asked.

Moreover, there are questions to be asked about Frege’s definitions of zero, predecession, and finitude. The first of these seems pretty rea-

categories is importantly different from that of his predecessors. But Benacerraf sometimes gives the impression that Frege is ultimately uninterested in epistemology—see especially the discussion of “metaphysical” dependence on pages 55-6—and with that I do not agree, for the reasons about to be given.

[47] Frege’s question is *not*: How *might* numbers be given to us in such a way that we *might* have *a priori* knowledge of them?

[48] This caveat, and the previous one, of course reflect my view, discussed in Section 1.2.3, that we do not really take “$\mathrm{N}x : Fx = \mathrm{N}x : Gx$” and “$\mathrm{Eq}_x(Fx, Gx)$” to have the very same truth-condition. The qualification is not critical here, however, and so I will ignore it henceforth, in order to simplify the exposition.

[49] It is a substantial issue whether even this much is true, because it is a substantial issue whether ‘having justification’ is closed under logical consequence. It seems to me, however, that, if we are discussing mathematical and logical knowledge itself, then this kind of closure principle will make the notion useless. So I think there is no defensible thesis in this vicinity.

[50] Linnebo (2004, p. 168) makes an even stronger claim: “For Frege’s Theorem to [establish that arithmetical knowledge is *a priori*], . . . its proof must have at least a reasonable claim to being just an explication of our ordinary arithmetical reasoning”. That is probably too strong.

sonable. The other two, however, are disputable. The defintion of finite, or natural, number has of course been actively disputed since not long after Frege gave it. The most famous criticisms were due originally to Poincaré, and similar concerns have been voiced ever since by those with predicativist leanings.[51] What is perhaps less obvious is that there are questions to be asked about Frege's definition of predecession, questions that focus on its logical complexity. The definition is

$$Pab \equiv \exists F \exists y[b = \mathrm{N}x : Fx \wedge Fy \wedge a = \mathrm{N}x : (Fx \wedge x \neq y)]$$

and the worry, first expressed by Linnebo (2004, pp. 172–3), is that the presence of the existential second-order quantifier here (that is, the fact that P, so defined, is Σ^1_1) makes the content of this definition depend too much on what comprehension axioms we have available: a may precede b if we accept certain comprehension axioms, but not if we do not.

Finally, there are questions to be asked about the logic used in the proof of Frege's Theorem. Traditionally, the question has been asked in the form: Is second-order 'logic' worthy of the name? But this is not the best form in which to raise the question. Rather, the question should be: Does the sort of reasoning employed in the proof of Frege's Theorem preserve whatever nice epistemological property one thinks HP has, in virtue of its being implicit in singular thought about numbers? What must be shown here depends upon what one thinks that nice property is. But it is not unreasonable to suppose that, if the 'logic' needed for the proof of Frege's Theorem does deserve the name, then the nice property will indeed be preserved, and I tend myself to discuss the question in those terms.

The questions just mentioned about Frege's logic and definitions are addressed in Chapter 12. It turns out that much less than full second-order logic is needed for the proof of Frege's Theorem. The power of second-order logic derives from the 'comprehension axioms', which are of the form:

$$\exists F \forall x[Fx \equiv \phi]$$
$$\exists F \forall x \forall y[Fxy \equiv \phi]$$

and so forth, where ϕ is some formula not containing F free.[52] Each of these axioms asserts that a given formula defines a 'concept' or 'relation': something in the range of the second-order variables. Sub-systems of second-order logic arise from restrictions on comprehension, that is, on what sort of formula ϕ may be. If, for example, we require ϕ not to contain bound second-order quantifiers, we have predicative second-order logic.

[51] The issue is raised, for example, by Parsons in "Frege's Theory of Number" (Parsons, 1995a, §VIII) and by Hazen in his review of *Frege's Conception* (Hazen, 1985, pp. 252–3).

[52] It generally will contain the indicated first-order variables free, and it may also contain additional free variables as parameters.

If we require ϕ to be of the form $\forall F \ldots \forall G\phi$, where ϕ contains no second-order quantifiers, then we have Π^1_1 comprehension. And so forth.

The natural question to ask is then: What is the weakest natural logic in which Frege's Theorem can be proven? The answer turns out to be that we need no more than Π^1_1 comprehension. By itself, that does not help very much, but there is a very different sort of logic that, since it has the same proof-theoretic strength, is also adequate for the proof of the axioms. This logic, which I call "Arché logic", has a stronger claim to count as 'logical' than full second-order logic does, because the standard challenges to second-order logic's right to the name do not apply to it. Moreover, Frege's definitions both of predecession and of the ancestral can be stated in such a system in an extremely natural way. The apparent complexity of the definition of predecession is then revealed as illusory.[53] The definition of the ancestral has an elegance that the usual definition lacks and that goes a long way towards making it plausible that this definition really does capture something fundamental about the notion of finitude.

I want to focus here, however, on whether there really is any sense in which an acceptance of HP is implicit in arithmetical thought. The issue seems to me to be critical, but it is poorly understood. Before we turn to that question, however, we need first to address a different one, namely, whether a Neo-logicist really does need to claim that HP is implicit in arithmetical thought. Wright has strenuously resisted this claim. Consider, for example, these remarks:

> Grant that a recognition of the truth of [HP] cannot be based purely on analytical reflection upon the concepts and principles employed in *finite*[54] arithmetic. The question, however, surely concerned the reverse direction of things: it was whether access *to* those concepts and validation *of* those principles could be achieved via [HP], and whether [HP] might in its own right enjoy a kind of conceptual status that would make that result interesting. (Wright, 2001b, p. 321, emphasis original)

Wright goes on to list "four ingredient claims" he says constitute Neo-logicism, the last of which reads: "[HP] may be laid down without significant epistemological obligation: ... it may simply be stipulated as an explanation of the meaning of statements of numerical identity..." (Wright, 2001b, p. 321).

[53]A similar treatment can be applied to addition and multiplication, by the way. In the case of addition, for example, the definition we want is:

$$\neg\exists x(Fx \wedge Gx) \rightarrow \mathrm{N}x : Fx + \mathrm{N}x : Gx = \mathrm{N}x : (Fx \vee Gx)$$

Just as in the case of predecession, it is only to formulate this for the general case of "$a + b$" that we need existential second-order quantifiers.

[54]The finitude of the arithmetic does not seem to play any significant role in this particular objection, so far as I can see, but there is another line of objection in which it might figure. See p. 36, below.

The emphasis on stipulation, which one finds in many of Wright's discussions (with and without Hale), can, to some extent, be interpreted charitably. In a way, it is but a familiar sort of idealization, one that abstracts from psychological contingencies deemed philosophically irrelevant.[55] But Wright's appeal to our conceptual freedom does more substantial work. In the quote above, Wright is insisting that the important question is not whether HP *does in fact* play some central role in ordinary arithmetical thought, but whether someone *might* acquire the sort of arithmetical knowledge we have by "stipulat[ing HP] as an explanation of the meaning of statements of numerical identity" and then proving the Dedekind-Peano axioms. The interesting questions are therefore supposed to be: Could one gain access to arithmetical concepts by committing oneself to using expressions of the form "$\mathrm{N}x : Fx$" as HP instructs and then, by rehearsing the proof of Frege's Theorem, validate the Dedekind-Peano axioms? If so, we are told, then HP "might... enjoy a kind of conceptual status that would make that result interesting". But my problem is that I do not see what sort of interest *any* result of that sort might have.

Grant that someone previously innocent of numerical concepts might stumble upon HP and decide to regard it as explanatory of the concept of cardinal number. Grant that such a person might then discover the proof of Frege's Theorem and so arrive at *a priori* knowledge of the Dedekind-Peano axioms. So what? As Kripke (1980, pp. 34–5) famously pointed out, "*a priori*" is an epistemic adverb. Truths are not what are *a priori*. It is this or that person's belief that is or is not justified *a priori*. At best, then, Wright's approach leads to the conclusion that it is possible for someone to have *a priori* knowledge of the basic laws of arithmetic. But did no one know those laws *a priori* before 1983? Do more than a handful of people now? So if, as MacBride (2000, p. 158) insists, the Neo-logicist "project was never to uncover *a priori* truth in what we ordinarily think, but to demonstrate how *a priori* truth could flow from a logical reconstruction of arithmetical practice", I find myself wondering why we should care. The point here, which is essentially Quine's (1969), emerges from skepticism about the very idea of 'rational reconstruction'.[56]

And it just isn't clear that HP *can* be "stipulated as an explanation of the meaning of statements of numerical identity" (Wright, 2001b, p. 321). Statements of numerical identity already have perfectly good meanings. If one wants to stipulate HP in an effort to fix the meanings of statements of 'gnumerical' identity, then that is a different matter, but the knowledge one might then develop from HP will not be numerical knowledge but gnumerical knowledge; it will not, that is, include any knowledge even of

[55]In particular, it allows us easily to divorce the 'context of justification' from the 'context of discovery'.

[56]I have elsewhere voiced similar concerns about the hypothetical character of radical interpretation, if its point is understood epistemologically (Heck, 2007, §2).

numerical identities. The point, of course, is that, if we are interested in what we know and how we know it, we must individuate the objects of knowledge finely: These are questions at the level of sense, not of reference (though we shall see that there are problems even at that level).

Wright is not, of course, unaware of this point:

> ...[I]t is one thing to define expressions which... behave as though they express [arithmetical] notions, another to define those notions themselves. And it is the latter point, of course, that is wanted if [HP] is to be recognized as sufficient for a theory which not merely allows of pure arithmetical interpretation but to all intents and purposes *is* pure arithmetic. (Wright, 2001b, p. 322, his emphasis)

But Wright clearly thinks that little needs to be done to defend the thesis that "$\mathrm{N}x : \ldots x \ldots$" really does mean: the number of ...s, or something close enough, and that none of the necessary work involves anything like conceptual analysis. Wright claims that "any doubt on the point has to concern whether the definition of the arithmetical primitives which Frege offers... [is] adequate to the ordinary *applications* of arithmetic". And to dispose of that doubt, Wright says, it will suffice to establish all instances of the following schema, which Hale once called "Nq":

$$n_f = \mathrm{N}x : Fx \;\equiv\; \exists_n x(Fx)$$

Here, "n" is a schematic variable for a numeral; $\exists_n$ is the numerically definite quantifier, "There are exactly n"; "n_f" abbreviates Frege's definition of the number that n denotes (so, e.g., "0_f" abbreviates: $\mathrm{N}x : x \neq x$). After observing that all instances of Nq can indeed be proven—the proof is by induction on n and is not difficult—Wright then remarks: "That seems to me sufficient to ensure that [HP] itself enforces the interpretation of Fregean arithmetic as genuine arithmetic, and not merely a theory which can be interpreted as such" (Wright, 2001b, p. 322).

But this cannot possibly be sufficient. No such argument can establish any more than that the Frege-ese version of "The number of Fs is n" is *provably equivalent* to the ordinary version, and that is far weaker than showing that they have anything close to the same meaning,[57] or even that "$\mathrm{N}x : Fx$" actually denotes a cardinal number. Just what would count as 'getting the meaning close enough to right' is a difficult question, to which we shall return, but the examples I am about to give do not split hairs.

Fix two geometrical points A and B. Now consider the following recursive definition:[58]

[57] To be clear, however, I am *not* assuming that there must be strict identity of sense between analysans and analysandum. Something weaker is obviously meant to be sufficient.

[58] Of course, this will only work for the finite case, but Wright too is only talking about the finite case. One might actually wonder why he thinks he can restrict attention to that case, since HP concerns infinite cases, too. This is less of a problem for me, since my own view is that what is implicit in ordinary arithmetical thought is HP restricted to the finite case. But Wright has not exactly been sympathetic to that view.

1. If $\neg\exists x(Fx)$, then $\mathrm{N}x : Fx = A$

2. If $\neg Fa$, then $\mathrm{N}x : (Fx \vee x = a) =$ the point bisecting the line between $\mathrm{N}x : Fx$ and B

Relying of course upon geometrical axioms, we can then prove HP (restricted to finite concepts),[59] define "0", "1", and the like exactly in Frege's way, and prove all instances of Nq. Does that show that "$\mathrm{N}x : \ldots x \ldots$", so defined, means: the number of . . . s, or something close enough?

Similarly, as W. W. Tait (1996) reminded us, the finite cardinals can be defined in terms of the finite ordinals, which is how Dedekind in fact proceeded. Indeed, the point is utterly general. So long as we are assured of the existence of (Dedekind) infinitely many objects, second-order logic will allow us to interpret the Dedekind-Peano axioms and then define the finite numbers in such a way that we can prove Nq. To give just one other example:[60]

1. If $\neg\exists x(Fx)$, then $\mathrm{N}x : Fx =$ '|'

2. If $\neg Fa$, then $\mathrm{N}x : (Fx \vee x = a) = \mathrm{N}x : Fx \frown$ '|'

Now the finite numbers are strings of strokes, as they were for Hilbert, and the proofs of HP and Nq will depend upon the laws of syntax.

It is really quite plausible that our grasp of the notion of infinity, and of a recursive process, emerges somehow from our competence with language. But surely the question whether cardinal numbers can be defined as strings of strokes is not answered by establishing Nq. Similarly, it is a serious question whether the ordinals or cardinals are more fundamental, and it is no small virtue of Dedekind's approach that, as Cantor showed, it can be extended into the transfinite. Dummett (1991b, p. 293) insists, on roughly these grounds, that, ". . . if Frege had paid more attention to Cantor's work, he would have understood. . . that the notion of an ordinal number is more fundamental than that of a cardinal number". My own view is different: I doubt that either is more fundamental. And I am not about to let Dummett off the hook by conceding that the provability of all instances of Nq shows that the definition of cardinals in terms of ordinals gets the meanings of cardinal identities close enough to right.

The short version, then, is that there are just too many ways to define "$\mathrm{N}x : Fx$" that make Nq provable, and very few of them have any claim at all to count as defining a 'genuinely arithmetical' notion.

So it looks as if we are left with the question what conditions a definition of "$\mathrm{N}x : Fx$" must satisfy if it is to count as introducing a genuinely

[59]This will rely upon a specification of what a finite concept is. One natural specification would use the (third-order) ancestral to define finitude recursively: Empty concepts are finite, and so are the results of adjoining an object to a finite concept. Then the proof of HP just uses the form of induction this definition makes available.

[60]Here, the frown, $\frown$, denotes concatenation.

arithmetical notion. I want to urge, however, that this way of putting the question, and the entire emphasis on our freedom to stipulate abstraction principles, is at best extremely misleading. It makes it seem as if the crucial questions are hypothetical: how reference to numbers, and *a priori* knowledge of their properties, *might* be possible. But if these things are possible, then that is because they are *actual*. We really do refer to numbers in our ordinary use of arithmetical language, and many of us we really do have *a priori* knowledge of some of their properties. The really interesting question is not how that might work but how it *does* work.

Abstractionism, as I understand it, offers an answer to part of this question. It is the view that our capacity for singular thought about types rests upon our appreciation of the truth-conditions of identity statements concerning them, where those truth-conditions can (almost) be given in the form of an abstraction principle. Maybe no form of abstractionism is true. But if, as I think, and as Wright has spent a lot of time arguing, some form of abstractionism is true, then our *actual* capacity for singular thought about numbers rests upon our *actual* appreciation of the (near) truth of some abstraction principle concerning numbers, and then the question becomes: Which one?

This question is, broadly speaking, empirical and psychological, but that makes it no less philosophical, and much of the work collected here is directed at it, in one way or another. I argue in Chapter 7 that, if we think of HP itself as restricted to the finite case, an appreciation of the connection between cardinality and equinumerosity that it reports really is fundamental to thought about cardinality. Unfortunately, that positive contribution, of which I am personally quite fond, has generated much less attention than a negative argument given in Chapter 11, where I claim to show that HP, if *not* understood as restricted to the finite case, cannot be what underlies arithmetical knowledge.

The worry, which Wright (2001b, p. 317) calls the "concern about surplus content", was first articulated by Boolos. After mentioning that HP is logically stronger than the Dedekind-Peano axioms, Boolos (1998d, p. 304) asks, "Faced with [such] results, how can we really want to call HP analytic?" It is not entirely clear why there should be any special problem for the Neo-logicist here. I suggest in Section 11.2, however, that what was really bothering Boolos rests upon a vague but nonetheless intelligible thought to the effect that the 'foundation' for a given discipline should not outstrip the discipline itself. For example, if someone said that our knowledge of elementary arithmetic rested upon our knowledge of Zermelo-Frankel set theory (ZF), one might reasonably reply that ZF is just way stronger than is needed, and so, while arithmetic might be founded on *part* of ZF, it surely isn't founded upon all of it.

In this particular case, we know how to proceed. ZF has lots of axioms, so we can look at which ones are used in interpreting arithmetic and which are not. It turns out that we can isolate a natural fragment of ZF, known as hereditarily finite set theory, and show that it is equi-interpretable with PA. That makes the view that arithmetic is founded on hereditarily finite set theory much more reasonable than the view that it is founded on ZF. The situation with HP is more difficult. HP is a single axiom, so it is less clear how we might weaken the foundation in this case. As it happens, there is a way forward: We can restrict HP's claims about identity of cardinality to the case of finite concepts, and then show that this new principle is equi-interpretable with something very close to (second-order) PA.[61]

Wright (2001b, pp. 317–18), however, expresses doubts about this entire line of argument. But I, in turn, have trouble understanding his doubts. The thought was just that, if HP really is at work in ordinary arithmetical thought, then there ought to be evidence of *its* being at work there, rather than some weaker principle that can do the same job.[62] Here is the way I put the point in Chapter 11:

> If Frege's Theorem is to have the kind of interest Wright suggests, it must be possible to recognize the truth of HP by reflecting on fundamental features of arithmetical reasoning—by which I mean reasoning about, and with, *finite* numbers, since the epistemological status of arithmetic is what is at issue. For what the logicist must establish is something like this: That there is, implicit in the most basic features of arithmetical thought, a commitment to certain principles, the (tacit) recognition of whose truth is a necessary precondition of arithmetical reasoning, and from which all axioms of arithmetic follow. (below, p. 245)

I then claim, on the basis of Boolos's results, that "... no amount of reflection on the nature of arithmetical thought could ever convince one of HP, nor even of the coherence of the concept of cardinality of which it is purportedly analytic" (below, p. 246).

Part of the argument here involves my insisting, in the second half of the long quote, on a concern with our *actual* arithmetical knowledge. Wright, as we saw, is uninterested in our actual knowledge, but we have already discussed that issue. The insistence on the restriction to *finite* arithmetic can be questioned, however. But before we get to that, let me clarify and correct my view.

As Wright is interpreting it, the argument from surplus content is precisely that: It is based upon a prohibition of surplus content.[63] I think

[61] I suspect that the formal arguments given for this claim could be greatly simplified if the definition of finitude mentioned in footnote 59 were used instead of the one I actually use in Chapter 11.

[62] Compare the sort of argument Evans (1982, §2.4) gives against relativizing reference: If reference were relativized to a world, say, then that would make certain readings of sentences possible that we never in fact get, so it would be a mystery why not.

[63] So "no surplus content" is supposed to be a necessary condition for successful identification of the foundation of a discipline. Wright also discusses the question whether lack of surplus content is *sufficient* for some sort of analyticity. I am not sure why that question is raised, however. I do not, so far as I can see, commit myself to such a sufficiency claim, and it does not seem very plausible.

it is an interesting question whether such a principle can be sustained, and we shall return to this question at the end of this section. But the argument was never meant to be so easy. I do not draw the conclusion that HP cannot ground our arithmetical knowledge simply from the logical facts established by Boolos. That "no amount of reflection on the nature of arithmetical thought could ever convince one of HP" is meant to follow from the fact of surplus content. But the crucial point was supposed to concern a closely related *conceptual* gap between finite arithmetic and the theory of infinite cardinality first introduced by Cantor. The reason HP has surplus content is that it answers the question when *infinite* concepts have the same cardinality. It is the fact that HP answers this question that is the problem, not the fact of surplus content to which it gives rise: Because HP answers that question the way it does, the concept of cardinality it characterizes is the Cantorian one, and what I argue in Chapter 11 is that the concept of cardinality that Cantor introduced cannot be what underlies our knowledge of finite arithmetic, because plenty of people have the latter who do not have the former.

This part of the argument can be challenged. MacBride (2000) suggests, in particular, that I overlook evidence that the Cantonian concept of cardinality is already implicit in ordinary arithmetical thought. For the reasons given in the Postscript to Chapter 11, however, I disagree, though I was no doubt too quick to draw the conclusion I did. I also think that any suggestion I might have made that the implicit commitments of ordinary arithmetical thought must be extracted by purely *philosophical* reflection should be rejected, since psychological phenomena of the sort discussed in Chapter 7 are also important.

A different way to question the argument, and one that seems implicit in some of Wright's reflections, is to ask whether finite arithmetic is really the right focus of investigation,[64] the alternative being something like the "general theory of cardinality". The idea, I take it, would be that *finite* arithmetic is not, so to speak, a natural epistemological kind; the natural kind is cardinal arithmetic generally. So what we would need to know to resolve this kind of dispute is how we should circumscribe what Frege would have called a "branch of knowledge" in such a way that its foundations might sensibly be investigated separately from those of other "branches of knowledge". Why, for example, is it at least plausible to isolate arithmetic from geometry (or syntax) and offer different accounts of our knowledge of each? Why should arithmetic, in the now familiar sense, be regarded as separable from real (or even complex) analysis? I suspect that good answers to these questions would have to draw upon psychology as well as philosophy, and I do not myself have any clear idea what such answers might be like. Nonetheless, I am reasonably confident that finite arithmetic does constitute an isolable body of knowledge in

[64]This is the other line of argument mentioned in footnote 54.

this sense, and the history of foundational studies, both before and after Cantor, seems to provide a good deal of evidence for this claim.

That said, I think we can now see why a prohibition on surplus content might seem reasonable. If the 'basic laws' that allegedly underlie a given, *properly circumscribed* body of knowledge were sufficient to provide for a sort of knowledge that we simply do not find exhibited, that would be reason to doubt that those really were the basic laws of that discipline. This would not be conclusive reason. We are dealing, in effect, with the distinction between competence and performance, and we might have reason to think that the competence, though there, was for some reason not manifested in performance. But one would have to tell a story about why not.[65] There is not, then, a prohibition on surplus content, but its presence should serve as a warning that something is amiss. In the present case, it is just such a warning, and there is something amiss.

1.4 Logicism and Neo-Logicism

Much of the literature on Neo-logicism has been concerned with the question whether its treatment of arithmetic can be mimicked in other areas, such as real analysis and set theory.[66] The question is not uninteresting, but my sense has long been that it has been accorded much too much significance. Casual presentations of logicism often present it as the view that "mathematics is logic". But Frege would have disagreed, since he regarded geometry—the field in which he was actually trained as a mathematician—as synthetic. By the same token, one's logicism need not encompass set theory or even analysis. Even if all of the various attempts to identify abstraction principles that are sufficient for these theories were to fail, that would in no way undermine the claim that arithmetic itself is analytic. What it would show is that other branches of mathematics depend upon "sources of knowledge" different from the ones upon which arithmetical thought draws. But maybe arithmetic just is special in that sense.

Yet another possibility is that not even all of arithmetic is analytic—that is, known on the basis of a principle constitutive of arithmetical thought—though some of it is. Perhaps no more than Robinson arithmetic is analytic. Maybe only the theory of successor is analytic. In "Ramified Frege Arithmetic" (Heck, 2011), I show that the basic axioms concerning successor can all be proven from HP in ramified predicative

[65] The beginnings of a story can certainly be told in the case of HP. One might want to say, much as MacBride (2000, p. 155) does, that HP was the principle guiding use of finite cardinals, but that it conflicted with an intuition regarding parts and wholes that was wrongly extrapolated from finite to infinite arithmetic. But I think there are independent problems with this line of argument. See the Postscript to Chapter 11.

[66] Several papers on these sorts of issues have been reprinted together in *The Arché Papers on the Mathematics of Abstraction* (Cook, 2007).

second-order logic, and in the context of predicative logic, there is no problem of bad company, since even Basic Law V is consistent in a predicative setting (Heck, 1996).[67] The argument has its limitations, however. Not only do we lose induction, for the sorts of reasons mentioned above, but there is also no clear way to define addition and multiplication so that the existence of sums and products can be proven. (Uniqueness is easy.)

However, it can be shown that, even in simple predicative second-order logic, we can, using standard definitions of cardinal addition and multiplication,[68] straightforwardly interpret a purely relational version of the theory known as R. The usual form of R has as axioms all true instances of the following formulae:

$$\begin{aligned} n + m &= k \\ n \times m &= k \\ m &< k \\ m &\neq k \end{aligned}$$

where m, n, and k are schematic variables for numerals. In the relational version, we do not have function symbols S, $+$, and $\times$, but relations $P(a, b)$, $A(a, b, c)$, and $M(a, b, c)$ and have as axioms all true instances of:

$$\begin{gathered} P(n, m) \\ A(n, m, k) \\ M(m, n, k) \end{gathered}$$

plus the assertions that these values are unique.[69] The resulting theory interprets R,[70] so it is sufficient for the numeralwise representability of all recursive functions and is therefore essentially undecidable (Tarski et al., 1953). What "Ramified Frege Arithmetic" shows, then, is that, if we ramify, then we do get the existence of successors and so can replace "$P(a, b)$" with a function symbol again. It would be nice if something similar could be done for sums and products, but, while I have not given up

[67] This paper is not reprinted here, as it consists mostly of intricate argumentation conducted in a barely intelligible formalism.

[68] For these definitions, see Burgess's book *Fixing Frege* (Burgess, 2005).

[69] Of course, the numerals cannot now be defined in the usual way—0, $S0$, $SS0$, etc.—since we do not have S. But we can define them via Russellian descriptions:

$$\begin{aligned} 1 &\stackrel{df}{=} \iota x(P(0, x)) \\ 2 &\stackrel{df}{=} \iota x(\exists y(P(0, y) \wedge P(y, x)) \end{aligned}$$

and so forth. Since we can prove both existence and uniqueness, the descriptions are proper. The issue is less pressing in the present context, since the numerals will be defined in Frege's way.

[70] Thanks for this information to Albert Visser, whose wonderful paper on R (Visser, 2009) got me thinking about this matter.

hope, and do not have a proof that it cannot be done, my various experiments have left me skeptical.[71]

What we get in the predicative case is therefore non-trivial—essential undecidability is as good a test for non-triviality as I can imagine—but it is not very much. Still, it would be wrong to dismiss these results on that ground. Even these weak results are capable of grounding significant philosophical conclusions, it seems to me, for the simple reason that it is arithmetic's commitment to an infinity of numbers that has always seemed to set it apart from the logical. If reason itself can provide us only with access to an infinity of numbers, while we must draw upon resources from elsewhere to establish much knowledge about them, then, well, that is how things are, and reason will still have proven capable of rather more than empiricists have generally supposed.

The lesson with which I should like to close, then, is one I have already announced but shall now re-iterate: If HP, or something like it, is indeed the fundamental principle on which all arithmetical thought is founded, then that is an epistemological result of great significance, whether or not HP itself enjoys any special epistemic virtue, and whether or not our knowledge of the Dedekind-Peano axioms derives directly from our knowledge of HP.

[71]One *can* get the existence of sums and products by restricting the domain to numbers for which sums and products exist. There is a nice presentation of the techniques for doing so, which are originally due to Robert Solovay, in Burgess's book *Fixing Frege* (Burgess, 2005, ch. 2). The difficulty, in the present context, is that this amounts to redefining the notion of natural number. As Visser (2011) notes, there may be a coherent philosophy of arithmetic to be founded on this idea, but it is one that would need to be developed, and it is very unFregean.