Frege's Theorem: An Introduction

By Richard G. Heck, Jr.

1. Opening

WHAT IS THE EPISTEMOLOGICAL STATUS OF OUR KNOWledge of the truths of arithmetic? Are they analytic, the products of pure reason, as Leibniz held? Or are they high-level empirical truths that we know only *a posteriori*, as some empiricists, particularly Mill, have held? Or was Kant right to say that, although our knowledge that '5+7=12' depends essentially upon intuition, it is nonetheless *a priori*? It was with this problem that Gottlob Frege was chiefly concerned throughout his philosophical career. His goal was to establish a version of Leibniz's view (and so to demonstrate the independence of arithmetical and geometrical reasoning), to show that the truths of arithmetic follow logically from premises which are themselves truths of logic. It is this view that we now call *Logicism*.

Frege's approach to this problem had a number of strands, but it is simplest to divide it into two: a negative part and a positive part. The negative part consists of criticisms of the views of Mill and Kant, and others who share them. These criticisms, although they are found in a variety of places, mostly occur in the first three chapters of *Die Grundlagen der Arithmetik*.[1] The positive part consists of an attempt to show (not just that but) *how* arithmetical truths can be established by pure reason, by actually giving proofs of them from premises which are (or are supposed to be) truths of pure logic. There is thus a purely mathematical aspect of Frege's project: it is this on which I intend to focus here.

Frege was not the first to attempt to show how arithmetical truths can be proven from more fundamental assumptions. His approach, however, was more rigorous and encompassing, by far, than anything that had come before. Leibniz, for example, had attempted to prove such arithmetical truths as "2+2=4." His proofs, however, like those of Euclid before him, rest upon assumptions that he does not make explicit: for example, Leibniz make free appeal to the associative law of addition, which says that $(a+b)+c = a+(b+c)$, i.e., allows himself freely to "re-arrange" parentheses. But it is essential to *any* attempt to determine the epistemological status of the laws of arithmetic that we be able precisely to determine upon what

RICHARD HECK, who recently received tenure in the Harvard Philosophy Department, has been at Harvard since 1991, when he received his Ph.D. from MIT. His fields of interest include the philosophies of language, logic, and mathematics, the work of Gottlob Frege, and general issues in metaphysics. Some of his recent publications include "The Development of Arithmetic in Frege's Grundgesetze der Arithmetik" (Journal of Symbolic Logic, 1993) and "The Sense of Communication" (Mind, 1995).

assumptions the proofs of such laws depend. That is to say, it is essential that the proofs be presented in such a way that, once the premises on which they are to depend have been stated, no *additional* assumptions can sneak in unnoticed. Frege's idea was to give the proofs within a "formal system" of logic, in which the permissible inferential steps are explicitly specified, by purely syntactic criteria, so that it becomes no more complicated to determine what assumptions are employed in the proofs than it is, say, to check calculations by long division.

Existing systems of logic, deriving from the work of George Boole (and, ultimately, from Aristotle), were, however, inadequate for this task, for two reasons: first, the systems were ill-suited to the presentation of *proofs*; and, secondly, they were inadequate even to represent the sentences which were contained in those proofs. In particular, it is impossible, in Boole's system, to express sentences containing multiple expressions of generality, such as "Every horse's head is an animal's head" or, more to the point, "Every number has a successor." In a *sense*, these sentences can be represented within Boole's logic, but not in such a way that one can see, on the basis of that representation, why the former follows from, but does not imply, "Every horse is an animal"—and so, not in such a way that proofs even of such simple facts can be carried out within it.

It was for this reason that Frege was forced to develop a new system of formal logic, which he first presented in *Begriffsschrift*.[2] His system is, as he frequently points out, adequate for the representation of sentences like those mentioned above; its rules of inference, though limited in number, allow us to carry out the sorts of proofs that could not be formalized in Boole's system. And once a system adequate to the representation of actual mathematical argumentation has been developed, the question of the epistemological status of arithmetic is transformed. It would be too strong to say that it *reduces* to the question from what assumptions the laws of arithmetic can in fact be proven—for, once such assumptions have been identified, we will be left with the question what the epistemological status of those assumptions is. But nonetheless, the philosophical question takes on a purely mathematical aspect, and mathematical techniques can be brought to bear on it.

Frege's idea that, through formal logic, the resources of mathematics can be brought to bear upon philosophical problems pervades contemporary analytic philosophy.

It would, I think, be impossible to over-emphasize the importance of the contribution implicit in Frege's approach here. Frege's idea that, through formal logic, the resources of mathematics can be brought to bear upon philosophical problems pervades contemporary analytic philosophy, having come down to us through the work of Russell, Wittgenstein, Carnap, and others. But its influence is not limited to philosophy. Mathematical logic, as we now have it, is obsessed with the question what can or can not be proven from particular assumptions, and it is only against the background of Frege's system of formal logic—or, at least, a system of formal logic that meets the conditions his was the first to meet—that this question can even be stated in a way that makes it mathematically tractable.

It is in that sense, then, that Frege's approach was more rigorous than existing

ones. There is also a sense in which it was more general. When Leibniz attempts to prove the laws of arithmetic, he focuses on such claims as that 2+2=4. Frege, however, is interested in more fundamental arithmetical truths, and for good reason. If it is to be shown that *all* truths of arithmetic are provable from logical laws, then, since there are infinitely many of these, this can not be established by literally proving all of them. Rather, some *basic* arithmetical truths need to be identified, from which all others plausibly follow, and then these basic truths need to be proven. That is to say, arithmetic needs to be *axiomatized*, just as Euclid had axiomatized geometry, and then proofs need to be given of the axioms. Frege was not the first to publish such axioms: that honor is typically accorded to Giuseppe Peano, though the historical record makes it clear that Richard Dedekind was actually the first to formulate them.[3] Nonetheless, Frege does state axioms for arithmetic (which are interestingly different from Dedekind's) and his proofs are directed, in the first instance, at these. He also proves, as Dedekind does, that the axioms are sufficient to characterize the abstract, mathematical structure of the natural numbers (up to isomorphism, as it is said, this being a standard test of the sufficiency of an axiomatization).[4]

When Leibniz attempts to prove the laws of arithmetic, he focuses on such claims as that 2+2=4. Frege, however, is interested in more fundamental arithmetical truths, and for good reason.

Of course, one cannot, in the strictest sense, prove the axioms of arithmetic within a system of pure logic: none of the expressions in the language in which the system is formulated even purport to refer to numbers, so the axioms of arithmetic can not even be written down in such a system. What are required are definitions of the basic notions of arithmetic in terms of logical notions; proofs of the axioms of arithmetic will then become proofs of their definitional translations into the formal, logical system. The mathematical project is thus, in contemporary terminology, to "interpret" arithmetic in a system of formal logic: to interpret one theory—call it the target theory—in another—call it the base theory—is to show that definitions can be given of the primitive vocabulary of the target theory (in this case, arithmetic) in terms of the primitives of the base theory (in this case, some formal theory of logic), so that definitional transcriptions of the axioms of the target theory become theorems of the base theory.[5] At the very least, such a result shows that the target theory is *consistent*, if the base theory is: for, if there were a proof of a contradiction to be had within the target theory, that proof could be replicated within the base theory, by proving the needed axioms of the target theory within the base theory, and then appending the derivation of a contradiction, in the target theory, to them. Hence, if there is no proof of a contradiction to be had within the base theory, there will be none to be had within the target theory, either.

Such interpretability results were well-known to geometers working in Frege's time—and, indeed, Frege was a geometer by training. Such techniques were among the most frequent used to establish the consistency of various sorts of non-Euclidean geometries—geometries which reject the parallel postulate. Proofs of the consistency of such geometries usually amounted to proofs that they could be inter-

preted within other sorts of theories, whose consistency was not in doubt (e.g., within Euclidean geometry itself).

It is worth emphasizing, however, that interpreting arithmetic within a system of formal logic will not necessarily help us to discover the epistemological status of arithmetic—*even if* we are agreed that the "system of formal logic" really is a system of *logic*, i.e., that all of its theorems are analytic truths. The problem can be illustrated as follows. It was Frege's view that, not just arithmetic, but also analysis (that is, the theory of real numbers), is analytic. If so, given a definition of ordered pairs, the theory of Euclidean geometry can be interpreted in analysis, by means of Cartesian co-ordinates. Does it then follow that, on Frege's view, Euclidean geometry must be analytic? That would be unfortunate, for Frege explicitly agreed with Kant that the laws of Euclidean geometry are synthetic *a priori*. But, in fact, there is no inconsistency in Frege's view here. What the interpretability result establishes is just that what *look like* the axioms of Euclidean geometry can be proven within analysis. The question is whether what *looks like* the parallel postulate really does mean what the parallel postulate means. That is, the question just begs to be asked whether the "definitions" of fundamental geometrical notions in fact capture the meanings of those notions, as we ordinarily understand them—for example, whether "a point is an ordered triple of real numbers" is a good definition of the word "point," as it is used in geometry. If it is not, it has not been shown that the truths of Euclidean geometry can be proven in analysis: not if we identify "truths of Euclidean geometry" by what they *mean* and not just by their orthographic or syntactic structure.

It was Frege's view that, not just arithmetic, but also analysis (that is, the theory of real numbers), is analytic.

A corresponding question will arise in connection with Frege's definitions of fundamental arithmetical notions. It will, that is to say, be open to a Kantian to question whether the definitions Frege gives do in fact capture the meanings of arithmetical notions as we ordinarily understand them. If they do not, then Frege will not have shown that the truths of arithmetic can be proven within logic, but only that sentences *syntactically* indistinguishable from the truths of arithmetic can be so proven. And that is not sufficient. So another large part of Frege's project has to be, and is, to argue that the fundamental notions of arithmetic are *logical* notions, that his definitions of them in logical terms are good definitions, in this sense.

Let me summarize the discussion to this point. Frege's philosophical project, to show that the laws of arithmetic are analytic, that they can be known on the basis of reason alone, has a mathematical aspect. The primary goal is to define the fundamental arithmetical notions in terms of notions of pure logic, and then, within a formal system of logic, to prove axioms for arithmetic. That is how Frege will identify "the basic laws of arithmetic," i.e., the fundamental assumptions on which arithmetical reasoning is founded. As important as this part of the project is, however, it does not answer the epistemological question on its own, for two sorts of questions remain open. First, there is space for the question whether Frege's definitions of the basic arithmetical notions really do capture their meaning—and so whether we have really succeeded in proving the axioms of *arithmetic* and so in

identifying *its* basic laws. And second, even if that question is answered affirmatively, we will have to ask what the epistemological status of the basic laws is: only if they are themselves truths of logic, analytic truths, will the proof show that the axioms of arithmetic are analytic truths. As we shall see, both questions have been prominent in the literature generated by Frege's project.

2. Frege's System of Formal Logic

THE FORMAL LOGICAL SYSTEM WHICH FREGE DEVELOPS IN *Begriffsschrift* is, in essence, what we now know as full, impredicative second-order logic: it allows for quantification over "concepts"—the extralinguistic references of predicates—and relations, as well as over objects. The system as it is presented in *Begriffsschrift* does not, however, meet the demands of rigor Frege imposes. In particular, one of its most important rules of inference is never explicitly stated, the rule in question being a rule of substitution. One might think that Frege's omission here is inconsequential: isn't it just obvious that, if one has a proof of some formula, then the result of substituting various other expressions for the variables which occur in that theorem should also be a theorem? Maybe so, but the claim that substitution is a valid form of inference is, in the context of second-order logic, an extremely powerful one.

In Frege's system, the rule of substitution plays the role played in more modern formulations by the so-called axioms of comprehension: the comprehension axioms characterize the concepts and relations over which the variables of the theory are supposed to range; that is, each of them asserts that a particular concept or relation *exists*. In full second-order logic, one has such a comprehension axiom for every formula of the theory, and the comprehension axioms then jointly assert that every formula $A(x)$ defines a concept, namely, that true of the objects of which the formula is true; that every formula $B(x,y)$ defines a two-place relation; and so forth. And without comprehension axioms, second-order logic is very weak. Indeed, even if we allow ourselves so-called predicative comprehension axioms—that is, even if we assume that formulae which do not *themselves* contain second-order quantifiers (which do not quantify over all concepts and relations) define concepts and relations—the resulting logic is still weak, in a well-defined technical sense.

Many philosophers have worried that appeal to impredicative axioms of comprehension introduces a kind of circularity into the characterization of the concepts and relations over which the variables of the theory range. The worry, as it appears in Russell, for example, is that it must be circular to characterize the concepts the theory talks about by quantifying over the concepts the theory talks about. Although Frege never discussed this problem, his response, I think, would have been to say that he does not propose to *say* what concepts lie within the domain of the theory by means of the comprehension axioms: the domain is to contain *all* concepts, and the question how any one of them might be *defined* should not be allowed to determine whether it *exists*; the comprehension axioms simply *assert* that every formula defines a concept.[6] But let me not pursue the matter further here.

Frege's failure to state a rule of substitution in *Begriffsschrift* is, thus, an important omission. But it is one he remedies in his later presentation of his formal theory in *Grundgesetze der Arithmetik*. In fact, his presentation of the system there is as rigorous as any before Gödel mathematized syntax in the early 1930s.

The axioms of arithmetic can not, however, be proven within full second-order logic alone. It can easily be shown that even the claim that there are two distinct objects is not a theorem of second-order logic; but arithmetic posits the existence, not just of two, but of infinitely many natural numbers. It is *this* claim that is the central obstacle to any Logicist development of arithmetic. Indeed, it is hard to see that there is any more difficult problem facing one who would answer the epistemological question with which we are concerned than to explain the genesis of our knowledge that there are infinitely many numbers, whatever sort of answer she might ultimately want to give.

3. Three Lessons and a Problem

AS SAID ABOVE, FREGE ARGUES, IN THE FIRST THREE CHAPters of *Die Grundlagen*, that Kantian and empiricist philosophies of arithmetic will not do. He comes away with three lessons on which he proposes to base his own view, and with one very large problem. The first lesson is that the natural numbers are to be characterized, as Leibniz suggested, by defining "zero" and "increase by one." This is clear enough, and we shall see below how Frege intends to define these notions. The second lesson is harder to understand: Frege puts the point by saying that "the content of a statement of number is an assertion about a concept" (*GI* §55). What he means is that that to which number is ascribed is not, strictly speaking, *objects*. Suppose I were to say, pointing to a pile of playing cards, "How many?" You might answer by counting the cards and telling me that there are 104. But then again, you might answer 2. It all depends upon how you take my question, which contains an ambiguity that can be resolved by asking: How many *what*? How many cards? Or how many complete packs? But then, it would seem, the number 104 is ascribed not to the pile of cards, their aggregate (or "fusion"), that somewhat scattered object whose parts are all and only the parts of the cards: for the number 2 can as justly be ascribed to that aggregate (for the packs have the same parts the cards do). One might conclude from this that ascriptions of number are subjective, that they essentially depend upon our way of *regarding* the aggregate to which number is assigned. But there is an alternative: to say, with Frege, that number is ascribed to a *concept*, in this case, either to *card on the table* or to *complete pack of cards on the table.*

That having been said, it is overwhelmingly natural to suppose that numbers are a kind of higher-order property, that they are *properties of concepts.*[7] For example, 0 would be the property a concept has if nothing falls under it (so the concept *disco album in my collection* has the property zero, since I do not own any disco albums). A concept will have the property 1 if there is an object which falls under it, and every object which does fall under it is identical with that one (so the concept *object identical with George Clinton* would have the property 1, since George Clinton falls under it, and every object which does is identical with him). And a concept will have the property $n+1$ if there is an object, x, which falls under it and the concept *object which falls under the original concept, other than x*, has the property n. Thus, the concept *object that is identical with either George Clinton or James Brown* will have the property 1+1 (i.e., 2), for there is an object falling under it, namely, the Godfather of Soul, such that the concept *object that is identical with either George Clinton or James Brown, other than the Godfather of Soul*, has the prop-

erty 1, for this is just the same concept as *object identical with George Clinton*, described in other words.

Frege discusses this sort of proposal in §§55-61 of *Die Grundlagen*. It is not entirely clear why he rejects it, but it seems plausible that his reason, ultimately, is that such an account will not enable us to prove the axioms of arithmetic—not, that is, without some further assumptions. Suppose that there were exactly two objects in the world, call them George and James. Then there will be a concept which has the property 0; there will be others which have the property 1; and there will be one which has the property 1+1. But there will be no concept which has the property 1+1+1: for there to be such a concept, there would have to be one under which three objects fell and, by hypothesis, there are only two objects in existence. Nor would any concept have the property 1+1+1+1, for the same reason. Just how one wants to describe the situation here is a delicate question: one can either say that there is no number 1+1+1 or one can say that 1+1+1 and 1+1+1+1 turn out, in this case, to be the same. Either way, though, there will be only finitely many numbers, and the laws of arithmetic cannot be proven. (For example, depending upon how we choose to describe the situation, either 2+2 will not exist at all, or it will be the same as 2+1.)

Of course, there are not just two objects in existence: but a similar problem will arise if there are only finitely many. The situation can be remedied if we simply assume, as an axiom, that there are infinitely many objects: this is the course Russell and Whitehead take in *Principia Mathematica*. Frege would not have had any interest in this sort of "solution," though. As said, the really hard problem facing the epistemologist of arithmetic is to account for our knowledge that there are infinitely many numbers. It is hard to see how assuming that we know that there are infinitely many *other* sorts of things is supposed to help. Proving the laws of arithmetic from such an assumption simply leaves us with the question of its epistemological status, and that is no advance, since the epistemological status of just this assumption was the hard problem with which we started. Moreover, if the objects we assume to exist are supposed to be *physical* objects like George Clinton, it seems unlikely that the claim that there are infinitely many of these is one *logic* could establish: indeed, one might well wonder whether it is even *true*.

The third lesson, then, is supposed to be that, despite the fact that an ascription of number makes an assertion about a concept, numbers themselves are *not* properties of concepts: they are *objects*. To put the point grammatically, number-words are not *predicates of predicates*, but *proper names*. (It should, all along, have seemed strange to speak of zero as being a property of a concept, to say such things as that the concept *disco album in my collection* has the *property* zero!) One might wonder how these two doctrines can be jointly held: the answer is that we need only insist that the most fundamental sort of expression that names a number is one of the form 'the number belonging to the concept *F*,' and that this is a proper name. Ascriptions of number, such as 'There are 102 cards on the table,' then get re-cast

The really hard problem facing the epistemologist of arithmetic is to account for our knowledge that there are infinitely many numbers.

as identity-statements, e.g.: the number belonging to the concept *card on the table* is identical with 102. This does, as it were, ***contain*** an assertion about a concept, but the number 102 does not appear as a ***property*** of a concept.

This, however, leaves Frege with a problem. He has denied that arithmetic is either synthetic ***a priori*** or ***a posteriori***, partly on the ground that numbers are not given to us either in perception or in intuition. How then *are* they given to us? What, so to speak, is the mode of our cognitive access to the objects of arithmetic? Frege's way of answering this question is subtly to change it, writing:

> Only in the context of a proposition does a word mean something. It will therefore suffice to explain the sense of a proposition in which a number-word occurs.... In our present case, we have to define the sense of the proposition "the number which belongs to the concept *F* is the same as that which belongs to the concept *G*"; that is to say, we must reproduce the content of this proposition in other terms, avoiding the use of the expression "the number which belongs to the concept *F*"(*GI* §62).

It is here that Frege makes "the linguistic turn." It is not, of course, that no philosopher before him had ever been concerned with language; Locke, for example, was obsessed with it and argued, repeatedly, that various sorts of philosophical problems were the results of illusions fostered by misunderstandings of language. (Perhaps Locke was the first logical positivist.) What is original about Frege's approach is that the epistemological problem with which he begins is transformed into a problem *in the philosophy of language*, not so that it can be discarded, but so that it can be *solved*.

The answer to the epistemological question that is implicit in Frege's treatment of it is that our cognitive access to numbers may be explained in terms of our capacity to *refer* to them, in terms of a capacity to denote them by means of expressions we understand. Of course, if the capacity to refer to an object by means of a proper name itself depended upon our having, or at least being able to have, perceptions or intuitions of it (or of other objects of its kind), no actual benefit would accrue from reconceiving the problem in this way. But it is precisely to deny that the explanation must proceed in such terms that Frege invokes the "context principle," the claim that the meaning of an expression can be explained by explaining the meanings of complete propositions in which it occurs. The goal, in this case, will be to give the definition of "the number which belongs to the concept *F* is the same as that which belongs to the concept *G*" in terms of concepts which themselves belong *to pure logic*. If that should be possible, it will follow that there is, so to speak, a purely logical route to an understanding of expressions which refer to numbers, that is, to a capacity to refer to them—and so to a capacity for cognitive access to them.

Frege notes, quoting from Hume, that a criterion for sameness of number is ready to hand. Suppose, for example, that I want to establish that the number of plates is the same as the number of glasses. One way, of course, would be to count them, to assign numbers to the concepts *plate on the table* and *glass on the table* and then to see whether these numbers are the same. But there is another way: I can pair off the plates and the glasses, say, by putting one and only one glass on each plate, and then see whether each plate ends up with a glass on it and whether there are any glasses left over. That is to say, I can attempt to establish a "one-one corre-

lation" between the plates and the glasses: if there is a one-one correlation to be had, the number of plates is the same as the number of glasses; otherwise not.

Indeed, as Frege is fond of pointing out, the process of counting itself relies upon the establishment of one-one correlations.[8] To count the plates *just is* to establish a one-one correlation between the plates and an initial segment of the natural numbers, beginning with 1; the last number used is then assigned to the concept *plate on the table* as its number. Why, indeed, does the fact that the same number is assigned by this process to the concepts *plate on the table* and *glass on the table* show that the number of plates is the same as the number of glasses? Because, says Frege, if there is a one-one correlation between the plates and the numbers from 1 to n, and another between the glasses and the numbers from 1 to m, then there will be a one-one correlation between the plates and the glasses if, and only if, n is m.

The notion of one-one correlation can itself be explained in logical terms (assuming, that is, that we accept Frege's claim that the general theory of concepts and relations, as developed in second-order logic, does indeed count as logic). For a relation R to be a one-one correlation between the Fs and the Gs is for the following two conditions to hold:

1. The relation is one-one, that is, no object bears R to more than one object, and no object is borne R by more than one object
2. Every F bears R to some G and every G is borne R by some F

In contemporary symbolism:

$$\forall x \forall y \forall z \forall w[Rxy \,\&\, Rzw \rightarrow (x{=}z \equiv y{=}w)] \,\&$$

$$\forall x[Fx \rightarrow \exists y(Rxy \,\&\, Gy)] \,\&\, \forall y[Gy \rightarrow \exists x(Rxy \,\&\, Fx)]$$

And we can now say that *there is* a one-one correspondence between the Fs and the Gs if *there is* a relation R which satisfies these conditions. Say that the Fs and Gs are "equinumerous" if so. The definition on which Frege settles is then:

> the number of Fs is the same as the number of Gs if, and only if, the Fs and the Gs are equinumerous

This has come to be called *Hume's Principle*, since, as said, Frege introduces it with a quotation from Hume (*not* because anyone thinks Hume really had this in mind).

4. The Cæsar Problem and Frege's 'Solution'

THE STATUS OF THIS EXPLANATION OF 'THE NUMBER OF Fs' is one of the major open problems with which Frege's philosophy of arithmetic leaves us. It is worth emphasizing, however, that the problem is more general than whether it provides us with a *purely logical* route to an apprehension of numbers. Frege's idea is that a capacity for reference to abstract objects can, in general, be explained in these sorts of terms. Thus, he thinks, our capacity to refer to *directions* can be explained in terms of our understanding of names of directions, these in turn explained by means of a principle analogous to Hume's Principle, namely:

> the direction of the line a is the same as the direction of the line b if, and only if, a is parallel to b

This does not promise a *purely logical* route to an apprehension of directions, on Frege's view, since lines are given to us only in intuition. But that does not matter: the *directions* of lines are *not* given in intuition (or so he claims), and so there is a corresponding problem about how they are given to us, a problem to which he offers a corresponding solution (see *Gl* §§64-5). This sort of answer to the question how we are to explain our capacity to refer to abstract objects thus generalizes in a natural way, and the exploration of its strengths and weaknesses has occupied a number of philosophers in recent years.[9]

Oddly enough, however, Frege ultimately rejects the claim that Hume's Principle *does* suffice to explain names of numbers. The stated reason is that the definition "will not decide for us whether [Julius Cæsar] is the same as the [number of Roman emperors]—if I may be forgiven an example which looks nonsensical."[10] Hume's Principle tells us what statements of the form "the number of *F*s is the same as the number of *G*s" mean; but it utterly fails to tell us what statements of the form "*q* is the number of *G*s" mean, except when "*q*" is of the form "the number of *G*s" (see *Gl* §66). It is far from obvious, though, why this is supposed to be a problem. I myself have come to the conclusion that it is not one problem, but many. First of all, as Frege notes, we do not seem to be in any doubt about Cæsar: whatever numbers might be, he's not one of them; *a fortiori* he's not the number of Roman emperors. And it's hard to see what the basis of this knowledge could be, if it was not somehow contained in our understanding of names of numbers. (It isn't an *empirical* fact that Cæsar isn't a number!) But if it is, then the offered explanation of names of numbers must, at least, be incomplete, since it does not capture *this element* of our understanding of them.

The second worry can be seen as arising out of this one. When I said above that Cæsar is not a number, I used *the concept of number*: if we understood the concept of number, it would be easy to answer the question whether some object, call it *a*, is the number of *F*s. For either *a* is a number or it is not. If it is not, it certainly isn't the number of *F*s; and, if it is, then it's the number of *G*s, for some *G*, so Hume's Principle will tell us if it is the number of *F*s. Moreover, it seems clear that what Frege needs to explain is not just our capacity to refer to individual numbers, but our understanding of the concept of number itself. Given Hume's Principle, the natural way to try to do so is to say that something is a number if there is a concept whose number it is. But if we spell that out, we find that what we have said is that *a* is a number if, and only if, there is a concept *F* such that *a* is the number of *F*s. And as said, Hume's Principle simply fails to explain what "*a* is the number of *F*s" is supposed to mean: so, unless there is some other way to define the concept of number, we will be without any understanding even of *what it means* to say that Cæsar is not a number.

Hume's Principle simply fails to explain what "a is the number of Fs" is supposed to mean: so, unless there is some other way to define the concept of number, we will be without any understanding even of what it means to say that Caesar is not a number.

All of this having been said, I think we can now begin to see what the problem with Hume's Principle—considered as a definition or explanation of names of numbers—is supposed to be. The object, recall, was to explain our cognitive access to numbers by explaining our capacity to refer to them; and to do that by defining, or explaining, names of numbers in purely logical terms. But now consider Hume's Principle again

> the number of *F*s is the same as the number of *G*s if, and only if, the *F*s and the *G*s are equinumerous

and focus especially on its left-hand side. This *looks like* an identity-statement: but what reason have we to think it really is one? Why think that "the number of *F*s," as it occurs here, is a *proper name* at all? That is, that the sentence has the *semantic*, as opposed simply to *orthographic*, structure of an identity-statement? Why not say instead that "the number of *F*s is the same as the number of *G*s" is just an incredibly misleading way of *writing* "the *F*s and the *G*s are equinumerous?" If, in fact, the sentence *did* have the semantic structure of an identity-statement, it would have to be legitimate to replace "the number of *F*s" with names of other kinds, for example, "Cæsar," and so to consider such sentences as "Cæsar is the number of *G*s." Or again, it must be permissible to replace "the number of *F*s" with a variable and so to consider such "open" sentences as "*x* is the number of *G*s," and to ask whether they are true or false when the variable takes various objects, e.g. Cæsar, as its value. But what the above observations show is that Hume's Principle, on its own, simply does not explain what sentences like this are supposed to mean. And that calls into doubt whether it really does explain *identity-statements* containing *names of numbers*.

This problem is of particular importance within Frege's philosophy. I mentioned earlier that one of his central goals is to explain the genesis of our knowledge that there are infinitely many numbers, that one of his central reasons for insisting that numbers are objects is that only then can this claim be proven. Frege's strategy for proving it is to argue thus: let 0 be the number belonging to the concept *object which is not the same as itself*; it is clear enough that 0 is indeed that number. Let 1 be the number belonging to the concept *object identical with 0*; 2, the number belonging to the concept *object identical with either 0 or 1*; and so forth. Again, it is clear enough that 1 and 2 are indeed these numbers; and only if we treat numbers as objects can we speak of such concepts as these. But, when we formalize Frege's argument within second-order logic, the expression "the number belonging to the concept *F*" will be cashed as: the number of *x* such that *x* is an *F*. And thus, the definitions of 0, 1, and 2 become:

> 0 is the number of *x* such that $x \neq x$
> 1 is the number of *x* such that $x = 0$
> 2 is the number of *x* such that $x = 0$ or $x = 1$

And so forth. And when we replace '0' with its definition in the second line, we get:

> 1 is the number of *x* such that: x = the number of *y* such that $y \neq y$

So Frege's definition of 1 contains an open sentence of the form "*x* is the number of *F*s"—precisely the sort of sentence Hume's Principle is impotent to explain.

It is a nice question—a question which has been the focus of much research in

recent years—whether there is a way around this problem. For present purposes, however, let me just record it. For, whatever its ultimate resolution, Frege himself declared it insoluble and so abandoned the attempt to use Hume's Principle as an explanation of names of numbers. In its place, he installs an *explicit* definition of names of numbers:

> the number belonging to the concept *F* is: the extension of the concept *concept equinumerous with the concept F*

Extensions here can be thought of as sets: the number of *F*s is thus the set of all concepts which are equinumerous with the concept *F*. So 0 will turn out to be the set of all concepts under which nothing falls; 1, the set of all "singly instantiated" concepts; 2, the set of all "doubly instantiated" concepts; and so on and so forth.

This move, however, turned out to be disastrous. In order to give this definition within his formal system of logic, Frege requires some axioms which tell us what we need to know about the extensions of concepts. And, if Logicism is to be established, the axioms need, at least plausibly, to be logical truths. There is an eminently natural axiom to hand. Talk of the "extensions" of concepts is governed by a principle of extensionality, that the concepts *F* and *G* will have the same extension only if every *F* is a *G* and every *G* is an *F*—if, that is to say, the *F*s just are the *G*s. And surely, one might say, every concept must *have* an extension. So the necessary axiom can be taken to be the following:

> the extension of the concept *F* is the same as the extension of the concept *G* iff every *F* is a *G* and every *G* is an *F*

Taken as a *definition* or *explanation* of names of extensions, this principle would suffer from problems similar to those afflicting Hume's Principle. Frege therefore does not so take it, but regards it as an axiom, saying in *Die Grundlagen*, simply that he "assume[s] it known what the extension of a concept is" (*Gl* §68 note).

Unfortunately, however, this axiom, which in *Grundgesetze* becomes Frege's Axiom V, suffers from far greater problems of its own: in the context of full second-order logic, it is *inconsistent* since, as Russell showed, paradox arises when we consider the concept *object which does not fall under any concept whose extension it is* and ask whether *its* extension falls under it. If it does, it doesn't; and if it doesn't, it does. Whoops.

5. Frege's Theorem

FIFTEEN YEARS AGO, THE STORY WOULD HAVE ENDED HERE. Frege does show that, given Axiom V and the definition of numbers as extensions of certain concepts, the axioms of arithmetic can all be proven. But this fact is uninteresting, since *anything* can be proven in an inconsistent theory. The axioms of arithmetic can be proven, but then so can their negations! A closer look at the *structure* of Frege's proofs reveals something interesting, however. Although Frege abandons Hume's Principle as a definition of names of numbers, he does not abandon it entirely—he continues to assign it a central role within his philosophy (see *Gl* §107). Frege is not entirely explicit about why, but the reason seems, to me anyway, to be that he did not regard Hume's Principle as *wrong*, but as incomplete. The concept of number really is, according to Frege, inti-

mately bound up with the notion of one-one correlation; but one can not use this observation to *define* names of numbers *via* Hume's Principle. Still, any acceptable definition must be compatible with Hume's Principle, must yield it as a relatively immediate consequence. Completing the definition then takes the form of providing an explicit definition from which Hume's Principle can be recovered. And, indeed, the very first thing Frege proves, once his explicit definition has been given, is that Hume's Principle follows from it.

It was Charles Parsons who first observed, in his paper "Frege's Theory of Number,"[11] that, once Frege has proven Hume's Principle, the explicit definition quietly drops out of sight—and, with it, all further references to extensions in Frege's proof. That is to say, the proof proceeds in two quite separate stages: first, there is a proof of Hume's Principle from Axiom V and the definition of names of numbers in terms of extensions; and then, there is a proof of the axioms of arithmetic from Hume's Principle within pure second-order logic. The observation did not cause much of a stir, however, for no special interest would attach to it unless Hume's Principle were, unlike Axiom V, *consistent* with second-order logic, and Parsons did not so much as raise the question whether it is.

Almost twenty years later, Crispin Wright re-discovered what Parsons had observed and showed in detail how the axioms of arithmetic can be derived from Hume's Principle in second-order logic.[12] He also showed that an attempt to replicate Russell's paradox within the new system fails—and he conjectured that the new theory was in fact consistent. Once formulated, the conjecture was quickly proved.[13]

If second-order arithmetic were ever shown to be inconsistent, that would precipitate a crisis in the foundations of mathematics that would make the discovery of Russell's paradox look trivial by comparison.

Call the second-order theory whose sole "non-logical" axiom is Hume's Principle *Frege Arithmetic*. Then it can be shown that Frege Arithmetic is consistent, if second-order arithmetic is; that is, if Frege Arithmetic is inconsistent, so is second-order arithmetic. But if second-order arithmetic were ever shown to be inconsistent, that would precipitate a crisis in the foundations of mathematics that would make the discovery of Russell's paradox look trivial by comparison. So Frege Arithmetic is (almost certainly) consistent. And *second-order arithmetic can be interpreted within it*: given appropriate definitions, one really can prove the axioms of arithmetic from Hume's Principle, in second-order logic. Let me say that again: *All truths of arithmetic* follow logically from the principle—seemingly obvious, once one understands it—that the number of *F*s is the same as the number of *G*s if and only if the *F*s are in one-one correspondence with the *G*s. It is this surprising and beautiful theorem that is now called Frege's Theorem. And study of the detailed, formal proof Frege gives of the axioms of arithmetic in *Grundgesetze* has shown that, charitably read, it does indeed amount to a proof of Frege's Theorem.[14]

How is Frege's Theorem proven? It would be inappropriate to give a complete proof of it here, but it is worth giving the reader some sense of how the proof goes. Let us write "#x:Fx" to mean "the number of x such that x is an F;" let

"$Eq_x(Fx,Gx)$" abbreviate the formula which defines "the *F*s are equinumerous with the *G*s." Then Hume's Principle can be written:

$$\#x{:}Fx = \#x{:}Gx \text{ iff } Eq_x(Fx,Gx)$$

Frege then defines zero as the number of the concept *non-self-identical*:

$$0 =_{df} \#x{:}x \neq x$$

We also need to define "increase by one." What Frege actually defines is a *relation* between numbers which we may call the relation of predecession: intuitively, a number m is one less than a number n if, so to speak, a concept has the number m whenever it is one object short of being a concept which has the number n. That is, m (immediately) precedes n just in case there is a concept F and an object y such that: the number of *F*s is n; y is an F; and the number of *F*s, other than y, is m. (Compare the definition of "$n+1$" considered in section 3.) Formally, Frege defines:

$$Pmn \equiv_{df} \exists F \exists y[n = \#x{:}Fx \,\&\, Fy \,\&\, m = \#x{:}(Fx \,\&\, x \neq y)]$$

Now, among the axioms of arithmetic will be the following three claims: that zero has no predecessor; that no number has more than one predecessor; and that no number has more than one successor. All of these are now easily proven.

Consider the first, for example. Suppose that zero did have a predecessor, that is, that, for some m, we had $Pm0$. Then, by definition, there would be a concept F and an object y such that:

$$0 = \#x{:}Fx \,\&\, Fy \,\&\, m = \#x{:}(Fx \,\&\, x \neq y)$$

That is to say, 0 is the number of *F*s, y is an F, and m is the number of *F*s, other than y. *A fortiori*, there is a concept F, whose number is 0, under which *some* object falls. But that is impossible. For 0 is the number of the concept *non-self-identical* and so, if 0 is the number of *F*s, the number of *F*s is the same as the number of non-self-identical things; and so, by Hume's Principle, there must be a way of correlating the non-self-identical things one-to-one with the *F*s. But that there can not be, if something, say y, is an F: which non-self-identical thing is y supposed to be correlated with? So nothing precedes zero. The proofs of the other two axioms mentioned are a little harder, but not much.

There are two more axioms which need to be proven; these axioms make crucial reference to the notion of a *natural* (or *finite*) number, and that has not yet been defined. One of the two axioms is the principle of mathematical induction. Induction is a method for proving that all natural numbers have some particular property: the method is to show (i) that 0 has the property, and then to show (ii) that, if a given number n has it, then, its successor, $n+1$ must also have it. Why, intuitively, does the method work? Well, suppose that both (i) and (ii) hold. Then, certainly, 0 has the property. And so, by (ii), 0+1, i.e, 1, just also have it; hence, by (ii) again, 1+1, i.e., 2 has it; so 3 has it; and so on. So, all natural numbers have it, because *all natural numbers can be "reached" from 0 by adding one.*

Frege's definition of natural number in effect transforms the italicized clause into a definition. Here is something we know about the concept *natural number*: (i′) 0 falls under it; and, (ii′) whenever an object falls under it, so does the successor of that object (if it has one). Now, there are lots of concepts which satisfy conditions (i′) and (ii′): for example, the concept *is self-identical* satisfies them, since *every*

object falls under it. But not every concept satisfies the conditions: the concept *is identical with zero* does not, since 1 is the successor of 0 and does not fall under the concept, whence it fails to satisfy (ii′). So some concepts satisfy (i′) and (ii′) and some do not: call a concept *inductive* if it does. What can we say about the inductive concepts? Well, if a concept is inductive, surely every natural number must fall under it: 0 will, by (i′); so 1 will, by (ii′); and so on. Or, to put the point differently, if there is an inductive concept under which *a* does *not* fall, *a* is not a natural number. Conversely, if *a* is not a natural number, then there is an inductive concept under which it does not fall: namely, the concept *natural number* itself. So *a* is not a natural number if, and only if, there is an inductive concept under which it does not fall. Or, negating both sides of this claim:

> *a* is a natural number if, and only if, it falls under *every* inductive concept

And that, now, can be taken as a *definition*. Note that it will just *fall out* of the definition that proof by induction if valid: for if the hypotheses of the induction, (i) and (ii), are satisfied, then the property in question is inductive; so every natural number will have it.

One might worry that there is some kind of circularity in the last paragraph. It is important to realize, however, that any circularity there might be here does not lie in Frege's definition. We can say what an inductive concept *is* without appealing to the concept of *natural number*. Formally, we have:

$$\mathrm{Nat}(a) \equiv_{\mathrm{df}} \forall F[F0 \,\&\, \forall x \forall y(Fx \,\&\, Pxy \rightarrow Fy) \rightarrow Fa]$$

Whatever circularity there might be lies, rather, in the argument used to motivate the definition, that is, *in the argument given for the claim that it properly defines the concept "natural number"*—and that sort of objection, as I said earlier, is in order here, since Frege does need for his definitions to capture the meanings of the primitive arithmetical notions. This sort of objection was originally pressed by Henri Poincare, but has in recent years been developed by Charles Parsons.[15] Let me not discuss it further, however, except to say that it is bound up with the concerns about impredicativity mentioned earlier in connection with the comprehension axioms for second-order logic.

The only axiom for arithmetic which we have not discussed is the most important: that every natural number has a successor. The axioms so far established do not, on their own, imply that there are infinitely many numbers; in fact, they are consistent with there being only *one* number, namely, zero. In their presence, however, the claim that every number has a successor implies that there are infinitely many numbers: for then, 0 will have a successor, call it 1, from which it must be distinct; if 0 were 1, then 0 would precede itself, so 0 would have a predecessor, which it does not. But then 1 has a successor, call it 2; and 2 must be distinct from 0 (if not, 0 would again have a predecessor, namely, 1) and from 1 (otherwise 1 would precede itself and 0 would also precede it, contradicting the fact that no number has more than one predecessor). And then 2 will have a successor, call it 3, which will be distinct from 0, 1, and 2; and so on.

How, then, are we to prove that every number does have a successor? Frege's argument is far too complex for me to explain it in detail here, but the basic idea

behind it has been mentioned above. Basically, we are to take the successor of 0, 1, to be $\#x{:}x{=}0$; the successor of 1, 2, to be $\#x{:}(x{=}0 \vee x{=}1)$; and so forth. So, in general, the successor of a number n will be the number of x such that x is either 0 or 1 or... or n. The argument can be made rigorous, and it can be shown to work.

6. Closing: The Philosophical Significance of Frege's Theorem

SO, FREGE'S THEOREM SHOWS THAT THE AXIOMS OF ARITHmetic can be proven, in second-order logic, from Hume's Principle. What are we to make of this fact, philosophically? Does it show that Logicism is true? Not, presumably, if Logicism is the claim that the truths of arithmetic are truths of *logic*, for there is no good reason to suppose that Hume's Principle is itself a truth of logic. Indeed, given how the notion of logical truth tends to be understood in contemporary philosophy, so that a truth of logic is something true in all interpretations, Hume's Principle has just been proven *not* to be a truth of logic, since it is not true in any interpretation whose domain is finite.

Still, though, one might think that Hume's Principle is, even if not a truth of *logic*, at least of a similar epistemological status. Wright suggests, for example, that it can, and should, be understood, as embodying an *explanation* of names of numbers—or, to put the point less technically, that we can come to know that it is true simply by reflecting upon what number-words *mean*. If, in that sense, it is *analytic of* the concept of number, then the axioms of arithmetic turn out to follow from a Principle which is analytic, in an extended sense, and that would seem enough to satisfy Frege's epistemological ambitions. And if that is not Logicism, surely it is rightful heir to the name.

This view has been the subject of a great deal of discussion over the last ten years or so. The problems it faces may be divided into two groups: the Cæsar problem and the "bad company" objection. Enough has already been said about the former above: the Cæsar objection, as was said, purports to show that Hume's Principle *cannot* be taken as an explanation of names of numbers. But let me not discuss it further here.

The "bad company" objection is so-called because it has the following form. Even though Hume's Principle is not *itself* inconsistent, it is of a form which can give rise to inconsistency. For example, Axiom V is, as was said earlier, inconsistent, and the two principles are both of the form:

the blah of *F* is the same as the blah of *G* iff the *F*s so-and-so the *G*s

where "so-and-so" stands for some relation between concepts which is, in technical parlance, an equivalence relation (that is, which has the formal properties necessary to guarantee that the definition will not contradict the laws of identity). Now, one might well say, what's the problem? To quote Michael Jackson, "One bad apple don't spoil the whole bunch of girls." But still, if there are problems that affect claims formally similar to a given one, then, even if they do not themselves affect the given claim, it is natural to suspect that those problems will, upon closer examination, be seen to be manifestations of deeper problems that *do* affect the given one.

One way to develop this objection is as follows. The claim that Hume's Principle is *true* is one that commits us, minimally, to its consistency; it is hard to

see how we can claim to *know* it to be true unless we are also prepared to claim that we *know* it to be consistent. Not that we do not know this: we do, as surely as we know much else we claim to know in mathematics. But the proof that it is consistent is one that can be formalized only within an exceptionally strong mathematical theory, one (by Gödel's theorem) essentially stronger than Frege Arithmetic (or second-order arithmetic) itself. And it seems to be asking us to swallow a lot to suppose that we can come to know anything like *that* simply by reflecting upon the concept of number.

This objection, however, is difficult to evaluate. For one thing, anyone attracted to the claim that arithmetic is analytic is already asking us to swallow a lot. The argument would seem to show, at best, that Wright ought to claim not just that we know Hume's Principle to be true, but that we know it to be consistent, which is more. And what's a little more? Moreover, there are worries about the form of argument that was deployed in the last paragraph: it appears to be of a sort that drives many kinds of skeptical arguments. Compare: if I know that there is a computer on my desk, then I must also know that there is an external world. Is it supposed to follow that my claim to know that there is a computer on my desk, as it were, depends upon my knowing *antecedently* that there is an external world? If so, we're in trouble, because it is hard to see how, if I suspend belief in all my particular items of knowledge about the world—i.e., subject that knowledge to Cartesian doubt—I could ever recover my senses. And it is a now-familiar move in epistemology to deny that, just because some claim A implies some other claim B, a claim to know A must be *supported by* a claim to know B: we may well know that B because we know that A, and not conversely. But then a similar move would appear to be available in this case: to grant that, if we know that Hume's Principle is true, we know it is consistent, but to deny that our claim to know that it is true need rest upon independent knowledge of its consistency.

Anyone attracted to the claim that arithmetic is analytic is already asking us to swallow a lot.

However good this response might be, though, there are other worries. If the view is supposed to be that *every* consistent principle of the general form mentioned above is "analytic," it is refutable. Principles of that form are a dime a dozen. And some of the consistent ones are inconsistent *with each other*. Thus, for example, it is possible to write down such principles which, though consistent, imply that there are only finitely many objects. Since Hume's Principle implies that there are infinitely many objects, it is inconsistent with these ones. Worse, choose any proposition you like: it is possible to write down a principle of the form mentioned above which implies it; and, by the same token, it is possible to write down another which implies its negation.[16] So not all such consistent principles can even be *true*, let alone analytically true. What is needed, then, is some way of distinguishing the "good" principles from the "bad" ones. But this problem has only begun to be studied, and it is not yet clear whether success is to be had. Even if it is, the question will remain whether there is any ground on which to claim that the good ones are all analytic.[17]

It is worth saying, however, that one should not over-emphasize the importance of this question. Let it be granted that Hume's Principle is not analytic. It

might nonetheless be that it has a role to play in a story about the genesis, or foundation, of our knowledge of the truths of arithmetic. The Principle does, after all, have powerful intuitive appeal, and the claim that it is, in some deep sense, integral to our understanding of the concept of number that concepts have the same number if, and only if, they are equinumerous could be true, even if it is *not* true that Hume's Principle is analytic, in any sense that will rescue Frege. φ

Endnotes

[1] Gottlob Frege, *The Foundations of Arithmetic*, tr. J.L. Austin (Evanston IL: Northwestern University Press, 1980). Further references are in the text, marked "*GI*", with a section number.

[2] Gottlob Frege, *Begriffsschrift: A Formula Language Modeled Upon That of Arithmetic, For Pure Thought*, in J. van Heijenoort (ed. and tr.), *From Frege to Gödel: A Sourcebook in Mathematical Logic* (Cambridge, MA: Harvard University Press, 1967), pp. 5-82.

[3] See Richard Dedekind, "The Nature and Meaning of Numbers," in *Essays on the Theory of Numbers*, tr. by W. W. Beman (New York: Dover Publications, 1963), pp. 44-115.

[4] For discussion of Frege's axioms and his proof of this theorem, see my "Definition by Induction in Frege's *Grundgesetze der Arithmetik*," in W. Demopoulos (ed.), *Frege's Philosophy of Mathematics* (Cambridge MA: Harvard University Press, 1995), pp. 295-333.

[5] Interpretations can take a more complicated form, but we need not consider such matters here.

[6] See Gottlob Frege, *Grundgesetze der Arithmetik* (Hildesheim: Georg Olms Verlagsbuchhandlung, 1966), vol. I, section 66.

[7] For a detailed discussion of this proposal, see my "The Julius Cæsar Objection," in R. Heck (ed.), *Language, Thought, and Logic: Essays in Honour of Michael Dummett* (Oxford: Oxford University Press, 1997).

[8] See *Grundgesetze*, vol. I, section 108.

[9] See Crispin Wright, *Frege's Conception of Numbers as Objects* (Aberdeen: Aberdeen University Press, 1983), and Bob Hale, *Abstract Objects* (Oxford: Blackwell, 1988).

[10] Frege is actually discussing the definition of directions here, but it is clear that it is meant to apply, *mutatis mutandis*, to the case of numbers and Hume's Principle.

[11] In his *Mathematics in Philosophy* (Ithaca NY: Cornell University Press, 1983), pp. 150-75.

[12] In his *Frege's Conception of Numbers as Objects.*

[13] This observation was made, independently, by George Boolos, John Burgess, Allen Hazen, and Harold Hodes. For a proof, see the second Appendix to George Boolos and Richard G. Heck, Jr., "*Die Grundlagen der Arithmetik* §§82-3," in M. Schirn, ed., *Philosophy of Mathematics Today* (Oxford: Oxford University Press, 1997).

[14] See my "The Development of Arithmetic in Frege's *Grundgesetze der Arithmetik*," *Journal of Symbolic Logic* **58** (1993), 579-601; reprinted, with a postscript, in Demopoulos (ed.), pp. 257-94.

[15] See, for example, "Frege's Theory of Number."

[16] See my "On the Consistency of Second-order Contextual Definitions," *Noûs* **26** (1992), pp. 491-4.

[17] See Crispin Wright, "The Philosophical Significance of Frege's Theorem" and George Boolos, "Is Hume's Principle Analytic?" in Heck (ed.), *op. cit.*