30–49
10.1093/philmat/nkv038
Philosophia Mathematica
Advance Access publication December 1, 2015

The Breadth of the Paradox†

Patricia Blanchette

Department of Philosophy, University of Notre Dame, Notre Dame, Indiana 46556, U.S.A.
E-mail: Patricia.Blanchette.1@nd.edu

ABSTRACT

This essay examines Frege's reaction to Russell's Paradox and his views about the grounding of existence claims in mathematics. It is argued that Frege's strict requirements on existential proofs would rule out the attempt to ground arithmetic in (HP). It is hoped that this discussion will help to clarify the ways in which Frege's position is both coherent and significantly different from the neo-logicist position on the issues of: (i) what's required for proofs of existence; (ii) the connection between models, consistency, and existence; and (iii) the prospects for a logical grounding of arithmetic in the wake of the paradox.

1. INTRODUCTION

Frege's reaction to the disastrous letter from Russell (22 June 1902) was as follows:

> Your discovery of the contradiction has surprised me beyond words ... it has rocked the ground on which I meant to build arithmetic. It seems accordingly that the transformation of the generality of an identity into an identity of ranges of values is not always permissible ... It is all the more serious as the collapse of my Law V seems to undermine not only the foundations of my arithmetic but the only possible foundations of arithmetic as such. [Frege, 1979, p. 132]

One of the enduring questions we face in trying to understand Frege's logicist project is the question of why he found the paradox quite this devastating.

†An early version of this essay was presented at the Munich Center for Mathematical Philosophy in January 2013; many thanks to audience members for comments at that time, and to Roy Cook for comments on the essay. Thanks as well to two anonymous referees for this journal.

Why does the contradiction threaten not just the specific way in which Frege had tried to ground arithmetic in logic, but the 'only possible' way of doing so?

We get some information from the letter Frege sent to Russell six weeks later. On July 28 1902, Frege writes:

> I myself was long reluctant to recognize ranges of values 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 ranges of values of functions. I have always been aware that there are difficulties connected with this, and your discovery of the contradiction has added to them; but what other way is there? [Frege, 1979, pp. 140–141]

This passage gives us a lot to go on. By 'logical objects', Frege generally means those objects whose existence is given via principles of logic. The purely logical status of *Grundgesetze*'s comprehension principle, essential to the claim that it really is logic, at bottom, to which we will have reduced arithmetic, is equivalent to the claim that value ranges are logical objects. The letter continues:

> We can also try the following expedient, and I hinted at this in my Foundations of Arithmetic. If we have a relation $\Phi(\xi, \zeta)$ 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". But the difficulties here are the same as in transforming the generality of an identity into an identity of ranges of values.[1] [Frege, 1979, p. 141]

The general procedure Frege refers to here is that of introducing a term-forming operator *the f of* ... by stipulating that statements of the form

(ID) $$\text{The } f \text{ of } a = \text{the } f \text{ of } b$$

are to have the same content as corresponding instances of

(EQU) $$\Phi(a, b),$$

[1] Frege to Russell 28 July 1902. Note the mistaken 'not' in the English translation, final sentence quoted. The usual way to present the abstractionist strategy is to require in addition that Φ be reflexive. That extra requirement is redundant in the cases Frege has in mind, in which we are given the assertion that $\Phi(a, b)$, from which by symmetry and transitivity, Frege's (1) and (2), it follows that $\Phi(a, a)$.

where Φ is an already recognized equivalence relation. Given such an equivalence between (ID) and (EQU), the associated abstraction principle,

$$\text{(ABS)} \qquad \text{The } f \text{ of } a = \text{the } f \text{ of } b \text{ iff } \Phi(a, b)$$

is, presumably, guaranteed to be true simply by the intended meaning of the introduced term-forming operator.

Russell's paradox shows immediately that not every attempt to introduce a term-forming operator in this way is successful, since there are equivalence relations (*e.g.*, the relation of coextensiveness) such that the required objects, the referents of the newly introduced terms, do not exist. In short, not every abstraction principle is true, and not every statement of the form $\Phi(a, b)$, for Φ an equivalence relation, 'can be replaced by' one of the form *the* f *of* a = *the* f *of* b, on pain of contradiction.

The paradox does not show, however, that all abstraction principles are inconsistent, and it would seem to leave the way open to affirming some of them. Especially important here is the principle (HP), the abstraction principle that characterizes the identity of cardinal numbers in terms of the equinumerosity of associated concepts:

$$\text{(HP)} \qquad \forall F \forall G\ (\mathrm{N}xFx = \mathrm{N}xGx \text{ iff } Fx \text{ equ } Gx),$$

where '$\Phi(x)$ equ $\Psi(x)$' abbreviates the statement that there is a 1-1 function mapping the Φs onto the Ψs.

As Crispin Wright has shown, following a suggestion of Charles Parsons and of Frege himself, (HP) suffices for the proof of those core principles on which Frege meant to ground the whole of arithmetic.[2] And as Richard Heck [1993] has shown, Frege clearly knew that (HP) would suffice in this way. This raises the natural question of why Frege didn't react to the paradox by simply beginning with (HP), avoiding all appeal to his inconsistent value ranges.

The neo-logicist strategy, the strategy just described of grounding arithmetic in (HP), has proven an important and fruitful approach to arithmetical foundations and to abstract objects more generally. The purpose of this essay is to pursue the question of why Frege took Russell's paradox to have the broad scope that he took it to have, which is in part the question of why he found the neo-logicist path impassable. The reason to pursue this question is partly historical: we cannot understand Frege's conception of abstract objects, and of our capacities to know and to refer to them, if we don't understand why he took the paradox to undermine abstractionism in general. Perhaps more importantly, as I shall argue below, the neo-logicist and the Fregean responses to the paradox stem from significantly different conceptions of what is required for a proof of mathematical existence. And while the Fregean approach to mathematical objects entails that logicism is undermined by the paradox, nevertheless that

---

[2] [Parsons, 1965; Wright, 1983]. On the consistency of the theory obtained by adding (HP) to second-order logic, see [Boolos and Heck, 1998], esp. Appendix 2.

approach, to which Frege was more deeply wedded than he was to logicism, offers us a compelling account of the nature of existence claims in mathematics.

2. TRUTH AND CONSISTENCY

In Frege's 1884 view, abstraction to directions, shapes, and similar objects involves the idea that the two sides of an instance of (ABS) express the same content.

> The judgement "line a is parallel to line b", or, using symbols,
>
> $$a//b$$
>
> can be taken as an identity. If we do this, we obtain the concept of direction, and say: "the direction of a is identical with the direction of b". Thus we replace the symbol $//$ by the more generic symbol $=$, through removing what is specific to the content of the former and dividing it between a and b. We carve up the content in a way different from the original way, and this yields us a new concept.
>
> . . . We can obtain in a similar way from the parallelism of planes a concept corresponding to that of direction in the case of straight lines; I have seen the name "orientation" used for this. From geometrical similarity is derived the concept of shape, so that instead of "the two triangles are similar" we say "the two triangles are of identical shape" . . . [Frege, 1884, §64]

The 're-carving' that gets us from talk of parallel lines to talk of directions is represented here as merely an innocent repackaging. It introduces nothing new, but only restates the original content in a form whose equivalence with that original is part of what one grasps in understanding the meaning of direction terms. The innocence of the re-carving strategy underwrites the truth of the abstraction principle, and guarantees that the singular terms on the left-hand side have reference as long as the right-hand side is well-formed.

Russell's paradox does not show that these paradigm instances of abstraction to directions or to shapes are inconsistent; they are in this sense like the abstraction to numbers given in (HP). The question raised in the preceding section is that of why Frege didn't find it natural to react to the paradox by excluding inconsistent abstraction principles, while continuing to embrace the abstractionist strategy in the remaining unproblematic cases. Why, that is, didn't he replace his problematically general conception of abstraction with the more moderate (CON)?

(CON) If an abstraction principle is consistent, it's true.

(CON) has been discussed by those interested in the neo-logicist project, and has been found straightforwardly inadequate for the purposes of that project. And while (CON) is also problematic from the Fregean point of view, it is worth pausing to see that the reasons Frege would have for rejecting (CON) are quite different from those of the neo-logicist.

The immediate reason for the unsatisfactoriness of (CON) from the neo-logicist point of view is that there are pairs of abstraction principles each of which is consistent, but which are inconsistent with each other, and hence not both true.[3] For example, we have the pair:

(HP) $$\forall F \forall G(\mathrm{N}xFx = \mathrm{N}xGx \text{ iff } Fx \text{ equ } Gx),$$

(PP) $$\forall F \forall G(PxFx = PxGx \text{ iff } (F - G) \cup (G - F) \text{ is finite and even}).$$[4]

We then note with George Boolos that (HP) has infinite models but no finite ones, while PP has finite models but no infinite ones. So (CON) is false.

Things look somewhat different from a Fregean point of view. The first point to notice is that the inference alluded to above, one that we find so natural that it goes without saying, the inference from the existence of a model to the consistency of the principle so modeled, is not an inference that Frege sanctions.[5] In response to David Hilbert's use of models to demonstrate the consistency of sets of Euclidean axioms (or negations thereof), Frege says that:

> As far as ... lack of contradiction ... is concerned, Hilbert's investigation ... is vitiated by the fact that the sense of the axioms is by no means securely fixed. ... Hilbert was apparently deceived by the wording. If an axiom is worded in the same way, it is very easy to believe that it is the same axiom. But it depends on the sense; and this is different, depending on whether the words "point", "line", *etc.* are understood in the sense of Euclidean geometry or in a wider sense. [Frege, 1980, p. 91][6]

Similarly,

> A proof of non-contradictoriness ... cannot be given by saying that these rules have been proved as laws for the positive whole numbers and therefore must be without contradiction; for after all, they might conflict with the peculiar properties of the higher numbers, *e.g.*, that of yielding -1 when squared. ... [I]n virtue of the peculiar nature of the complex higher numbers there may arise a contradiction where so far as the positive whole numbers are concerned, no contradiction obtains. [Frege, 1885, p. 102]; English translation [Frege, 1984, p. 102]

In brief: from Frege's point of view, the existence of a model of a sentence shows that a particular thought is true: the thought expressed by that sentence

---

[3]See *e.g.*, [Boolos, 1990; Dummett, 1991; Hazen, 1985; Heck, 1992; Hodes, 1984].

[4]This is George Boolos's parity principle; see [Boolos, 1990]. In what follows, this relation between F and G will be abbreviated as 'F par G'.

[5]I have argued for this claim at length elsewhere; see my [1996; 2007; 2012].

[6]Frege to Liebmann, 29 July 1900.

when it is interpreted in the way given by the model. But the truth of that thought does not entail the consistency of the quite different thought in which we were originally interested: the thought expressed by the same sentence when given its original, or intended, interpretation. That one such thought can be true (and hence consistent) without guaranteeing the consistency of the other thought is due to the fact that consistency and inconsistency can turn on the contents of the non-logical terms appearing in the sentence. For an especially vivid example, consider the sentence (¬PA2), the negation of the conjunction of the second-order Peano axioms (expressed in a language with "0" and "successor" primitive); this sentence has a model, but the thought it expresses under its intended interpretation is from Frege's point of view the negation of a truth of logic, and hence presumably not consistent.[7]

Consider the following model[8] A for (HP):

A's universe is the natural numbers;
A assigns to the term-forming operator '$\mathrm{N}x \ldots x \ldots$' the function f_A:
For all $S \subseteq \mathbb{N}$:

$$f_A(S) = 0 \text{ if } \mathrm{card}(S) \text{ is infinite;}$$
$$f_A(S) = n + 1 \text{ if } \mathrm{card}(S) = n.$$

The thought assigned by A to the sentence (HP), which we shall henceforth call 'τ_A(HP)', is: $\forall F \subseteq \mathbb{N}\ \forall G \subseteq \mathbb{N}\ (f_A(F) = f_A(G) \text{ iff } F \text{ equ}_A\ G)$.[9]

The thought assigned to the sentence (HP) as that sentence is ordinarily understood, which we'll call 'τ(HP)', is: $\forall F \forall G(\mathrm{N}(F) = \mathrm{N}(G) \text{ iff } F \text{ equ } G)$, where '$F$' and '$G$' range over all first-level concepts.

Consider the model B for (PP):[10]

$\|B\| = \{0, 1, 2, 3\}$; 'P' is interpreted by the function f_B:
For all $S \subseteq \|B\|$:

$$f_B(S) = 0 \text{ if } \mathrm{card}(S) \text{ is even;}$$
$$f_B(S) = 1 \text{ otherwise.}$$

---

[7] Frege himself does not discuss models of arithmetic; his views about models and consistency arise in the setting of geometry. His treatment of the finite cardinals involves the proof, from principles that he takes to be purely logical, of principles whose conjunction is a straightforward equivalent of (PA2).

[8] [Boolos, 1990]; reprinted in [Boolos, 1998].

[9] Strictly speaking, a model in this sense is too coarse-grained to assign a determinate thought to a sentence, since the model fixes the reference but not the sense of each sentential component. Putting it more carefully, we should say that the model fixes for each sentence a class of thoughts with the same truth conditions. Those truth conditions are what matter for current purposes. 'F equ$_A$ G' expresses the existence of a function in A that maps F $1-1$ onto G.

[10] From [Boolos, 1990]. '$\|B\|$' stands in what follows for the domain of B.

Similarly, τ_B(PP) is: $\forall F \subseteq \|B\|\ \forall G \subseteq \|B\|\ (f_B(F) = f_B(G)$ iff F par G). And the thought expressed by (PP) when its quantifiers are unrestricted, τ(PP), is: $\forall F \forall G(P(F) = P(G)$ iff F par G).

For ease of reference, we list those four thoughts here:

$$\tau_A(\mathrm{HP}) : \forall F \subseteq \mathbb{N}\ \forall G \subseteq \mathbb{N}\ (f_A(F) = f_A(G) \text{ iff } F \text{ equ}_A\ G).$$

$$\tau(\mathrm{HP}) : \forall F \forall G(N(F) = N(G) \text{ iff } F \text{ equ } G),$$

$$\tau_B(\mathrm{PP}) : \forall F \subseteq \|B\|\ \forall G \subseteq \|B\|\ (f_B(F) = f_B(G) \text{ iff } F \text{ par } G).$$

$$\tau(\mathrm{PP}) : \forall F \forall G(P(F) = P(G) \text{ iff } F \text{ par } G).$$

That A is a model of (HP) means that τ_A(HP) is true. The theme expressed by Frege in his criticism of Hilbert is that this fact fails to entail the consistency of τ(HP). Similarly, that B is a model of (PP) is a matter of the truth of τ_B(PP). This, for Frege, does not show that τ(PP) itself is consistent.

τ(PP) is presumably false; it is so, certainly, if there are infinitely many objects. Further: if, as Frege thought, it's a truth of logic that there are infinitely many objects, then it's a truth of logic that τ(PP) is false, in which case it too is inconsistent.

On this point, the modern logician, and in particular the modern anti-neo-logicist, takes Frege to be simply confused, or to have a problematically eccentric view of what counts as 'logic'. For, the thought is, pure logic simply cannot guarantee the existence of objects. As Boolos puts it,

> We firmly believe that the existence of even two objects, let alone infinitely many, cannot be guaranteed by logic alone. After all, logical truth is just *truth no matter what things we may be talking about and no matter what our (nonlogical) words mean*. Since there might be fewer than two items that we happen to be talking about, we cannot take even $\exists x \exists y(x \neq y)$ to be valid. ([Boolos, 1987] in [Boolos, 1998, p. 199])

> A sentence is a logical truth only if it is true no matter what objects it speaks of and no matter to which of them its predicates or other non-logical words apply. . . . A sentence is not a logical truth if it is false when interpreted over a domain containing infinitely many things, and it is not a logical truth if, like Hume's Principle, it is false when only finitely many things belong to the domain. ([Boolos, 1990] in [Boolos, 1998, p. 214])

But it's worth pointing out that this line of thought involves a quick slide from the idea of logical truths as *purely general* to the idea of logical truths as *sentences true under every interpretation*. Frege himself is a firm believer in the generality of logic: as he understands it, there is no area of inquiry in which the principles of logic in any sense fail. But this doesn't mean that the pronouncements that logic makes about the whole universe must hold for small subsets of that universe. Indeed, it is difficult to see how, unless one were already committed to a model-centric understanding of logic, one would find that thought remotely plausible. Independent plausibility aside, the central point here is just that the idea of logical truths as sentences that are true under every interpretation simply begs the question against logicism, from Frege's point of view.

Let's return to a consideration of (CON), according to which consistency is a guarantee of truth for abstraction principles.

For Frege, as discussed above, the existence of a model of a sentence doesn't guarantee the consistency of the thought ordinarily expressed by that sentence. Some important examples of this include (¬PA2) (above), and (¬HP), the negation of (HP). Each of these has a model, but this relatively mundane fact is no argument against Frege's logicism, according to which each is in conflict with the deliverances of pure logic.

Similar difficulties attach, from Frege's point of view, to the use of models to prove the consistency of concepts defined by partially interpreted sentences. Taking the term-forming operator 'N$x \ldots x \ldots$' to appear schematically in (HP), that sentence defines a condition met by a 1-place second-level function η if and only if η is a function from all first-level concepts to objects, and it assigns the same object to two concepts iff those concepts are equinumerous with each other. The consistency of this condition is a matter of the consistency of the thought that there is such a function, which from Frege's point of view is equivalent to the consistency of the thought that there exist the relevant objects, those that will form the range of the function.[11] It is a matter, as one might put it, of the consistency of the existence of numbers. The model of (HP) that we've seen provides a function, and objects, that meet a restricted version of this condition, a version restricted in the collection of concepts over which the function is defined, and hence in the collection of concepts to which the number-like objects must be assigned. On Frege's view, the consistency of this restricted condition is immediately implied by the truth of the thought τ_A(HP), since the truth of this thought entails the satisfaction of that condition. But, as above, the satisfaction of this restricted condition is not sufficient to show the consistency of its more general counterpart, that defined by the original (HP).

The difficulty of demonstrating consistency via the presentation of models therefore provides a roadblock, from Frege's point of view, in the way of the standard modern argument against (CON). Nevertheless, (CON) itself is of no help, even from Frege's point of view, as a means of justifying particular abstraction principles. For an appeal to (CON) is essentially an appeal to the idea that if the condition defined by an abstraction principle is consistent, then that condition is instantiated. And there are two substantial problems, from Frege's point of view, with such a line of thought. The first is that unless we have a non-question-begging way of proving the consistency of a condition, *i.e.*, one that does not turn on a prior demonstration that the condition is instantiated, then the principle (CON) can never be applied. And by Frege's lights, such

---

[11] Given such a collection of objects, the existence of a function that maps concepts to them in the way required is immediate, assuming Frege's rich conception of functional existence.

a non-circular method of proving consistency has not yet been exhibited. The second difficulty is that the crucial inference involved, from the consistency of a condition to its instantiation, is in Frege's view simply fallacious. As he puts it in *Grundlagen* [1884]:

> §94 A concept is still admissible even though its defining characteristics do contain a contradiction; all that we are forbidden to do is to presuppose that something falls under it. But even if a concept contains no contradiction, we still cannot infer that for that reason something falls under it ... [I]t does not follow that because we see no contradiction there is none there, nor does a clear and full definition afford any guarantee against it.
>
> §95 Strictly, of course, we can only establish that a concept is free from contradiction by first producing something that falls under it. The converse inference is a fallacy. . . .

And in a letter to Hilbert:

> Suppose we knew that the propositions
>
> 1. X is an intelligent being,
> 2. X is omnipresent,
> 3. X is omnipotent,
>
> together with all their consequences did not contradict one another; could we infer from this that there was an omnipotent, omnipresent, intelligent being? This is not evident to me. (Frege to Hilbert, 6 January 1900 [Frege, 1979, p. 47])

That we would not contradict ourselves in supposing the existence of a function or a collection of objects of a certain kind is no guarantee, as Frege sees it, that such a function or collection actually exists. Hence in order to justify (CON), from Frege's perspective, one would have to give a special reason for taking consistency to imply existence in the particular case of abstraction principles. That is to say, one would need a *proof* of (CON).

3. STRICTER CONDITIONS

Neo-logicists have reacted to the modern argument against (CON) by tightening up the conditions in its antecedent: the abstraction principles taken to be reliable are not all of the consistent ones, but the ones meeting a stronger condition, *e.g.*, that of 'strong stability'.[12] The fundamental idea is that, for some such condition, which we'll henceforth simply call the *safety condition*, any

[12] An abstraction principle A is *strongly stable* if there is some cardinal γ such that all of A's models are of cardinality at least γ, and for every cardinal $\kappa \geq \gamma$, A has a model

abstraction principle that satisfies it is thereby eligible to serve as a definition of its contained term-forming operator and of the terms so formed. Understood as definitions, the principles are, according to this strategy, sufficiently analytic to stand as fundamental truths in a logicist reduction. In the case of arithmetic, the important point is that (HP) satisfies the safety condition, and that such problematic principles as Basic Law V and (PP) do not.

In comparing Frege's understanding of the fundamental principles of arithmetic with the neo-logicist approach, we will sidestep the vexed question of how to understand the sense in which abstraction principles meeting the safety condition are taken to be 'analytic', and focus on the weaker condition of the guarantee of their truth, and the consequent guarantee of the existence of objects referred to by their singular terms. That is, we'll focus on the following two claims, taking (HP) as our central example:

(S) If an abstraction principle meets the safety condition, then it is true and its introduced singular terms refer;

and

(SHP) (HP) meets the safety condition.

An immediate difficulty with (S) and (SHP), from the Fregean point of view, is that they suffer from exactly the problems noted above with (CON). Each of the safety conditions under discussion in the neo-logicist project is model-theoretic; each is the requirement that the abstraction principle has models of a certain kind (and, in some cases, that it has models *only* of a certain kind). For just the reasons given above, the existence of models (even of interesting classes of them) of an abstraction principle does not suffice, from Frege's point of view, even for the minimal condition of the consistency of the thought expressed by that principle under its intended interpretation.

A second and more important difficulty is that the safety conditions of the neo-logicist project are just like consistency in that their satisfaction by a proposed abstraction principle provides no guarantee of the existence of the proposed objects. It is important to note that the argument that an abstraction principle has models of a certain kind (*i.e.*, that it satisfies a given safety constraint) is not itself intended as a proof of the existence of objects named by the abstraction principle's singular terms. For the idea is not that the background set theory (used to provide proofs of model existence) itself generates the abstract objects in question. It is rather that the background set theory, and the models it is used to construct, are intended to demonstrate that the abstraction principle has some general structural features, which structural features themselves are taken to suffice for the existence of the objects named by its

of cardinality κ. See especially [Cook, 2012]; also [Burgess, 2005; Fine, 2002; Linnebo and Uzquiano, 2009; Weir, 2003; Wright, 1997]. For a nice overview, see [Linnebo, 2009].

singular terms. Recall that Frege's objection to an inference from consistency to existence is that such an inference is ungrounded unless we have a prior guarantee that all consistent conditions are instantiated. Just the same point is applicable here: the inference from the safety of an abstraction principle to the existence of abstracta meeting its defining condition is ungrounded unless we have a prior guarantee that all conditions defined by such principles are instantiated. Without a proof, then, that safe abstraction principles are in fact instantiated, we are in no position to assume (S).

In order to understand Frege's conception of what's required for demonstrations of existence, we need to understand how he thought of his own introduction of mathematical objects, as value ranges governed by Basic Law V. A crucial point to consider is Frege's understanding of how his appeal to value ranges differed from the more 'creative' methods used by some of his contemporaries to introduce mathematical objects, methods that he found wanting. Regarding Dedekind's assumption of the existence of irrationals corresponding to cuts, *i.e.*, Dedekind's claim that, as quoted by Frege, 'Whenever ... we are presented with a cut (A_1, A_2) that is not generated by a rational number, we create for ourselves a new, irrational number a, which we regard as completely defined by this cut ...' [1903, §139], Frege comments:

> This creating is the heart of the matter. ... Here the question is whether creating is possible at all; whether, if it is possible, it is so without constraint; or whether certain laws have to be obeyed while creating. In the last case, before one could carry out an act of creation, one would first have to prove that, according to these laws, the creation is justified. These examinations are entirely missing here, and thus the main point is missing; what is missing is that on which the cogency of proofs conducted using irrational numbers depends. [1903, §139]

Regarding Hankel's treatment of complex numbers in terms of 'units', Frege remarks that Hankel's proofs are 'no more than a perplexing sleight of hand; for nowhere is it proven that there are such units, nowhere is it proven that it is legitimate to create them' [1903, §141].

Two sections later, Frege criticizes the creative strategy of Stolz in now familiar terms:

> To be sure, this creative power is constrained [in Stolz's view] by the addition that these properties must not contradict each other ... . How is it to be recognized that properties do not contradict each other? There seems to be no other criterion than to find the properties in question in one and the same object. In that case, however, the creative power that many mathematicians award themselves is as good as worthless. For before performing the act of creation, they now have to prove that the properties that they want to attribute to the object to be created ... do not contradict each other; and this they can do, it seems, only by proving that there is an object which has all these properties. If they

> can do that, there is no need to create such an object in the first place. [1903, §143]
>
> ... Incidentally, how is it known that avoiding contradiction is the only constraint that creation has to obey? [1903, §144]

In short, as Frege puts it here, that 'presumed creative power of the mathematicians' that consists in inferring existence from apparent consistency 'has to be regarded as worthless' [1903, §144].

A pressing question for our purposes is how Frege takes his own invocation of value ranges to escape the difficulties just outlined. Here is Frege's answer:

> It has thus become plausible that creating proper is not available to the mathematician; or at least, that it is tied to conditions that make it worthless. Against this, it could be pointed out that in the first volume (§3, §9, §10) we ourselves created new objects, namely value ranges. What did we in fact do there? Or to begin with: what did we not do? We did not list properties and then say: we create a thing that has these properties. Rather, we said: if one function (of first-level with one argument) and a second function are so constituted that both always have the same value for the same argument, then one may say instead: the value range of the first function is the same as the value range of the second. We then recognize something in common to both functions and this we call the value range both of the first function and of the second function. That we have the right so to acknowledge what is common and that, accordingly, we can convert the generality of an equality into an equality (identity) must be regarded as a basic law of logic. [1903, §146]

And finally, commenting on the habits of previous mathematicians,

> It is certainly clear that the mentioned possibility of conversion has, in fact, always been made use of; it is just that the coinciding is predicated of the functions themselves, rather than of the value ranges. If a first function always has the same value for the same argument as a second, it is customary to say: 'The first function is the same as the second' or 'Both functions coincide'. Even though this expression is different from our own, still, the generality of an equality is here converted into an equality (identity). [*Here appears a footnote about the impermissibility of e.g.*, '$f = g$'.]
>
> When logicians have long spoken of the extension of a concept and mathematicians have spoken of sets, classes, and manifolds, then such a conversion forms the basis of this too; for, one may well take it that what mathematicians call a set, *etc.*, is really nothing but the extension of a concept, even if they are not always clearly aware of this.

> We are thus not really doing anything new by means of this conversion; but we do it in full awareness and by appealing to a basic law of logic. And what we do in this way is completely different from the arbitrary, lawless creation of numbers by many mathematicians. [1903, §147]

Frege holds that the introduction of mathematical objects is something about which one must be especially cautious, and about whose justifying principles one must be especially forthright. In a project whose purpose is to bring to light the principles on which the science rests, the invocation of objects must be accompanied by an explicit statement of the principles underwriting their existence. The problematic procedure that Frege finds in the work of his contemporaries, involving the unjustified assumption of consistency and the unjustified inference from this to existence, is therefore made triply deficient by its lack of a clear statement of the principle(s) on which those existence claims are based.[13]

That Frege's own procedure does not suffer from precisely the flaws of his contemporaries' methods is clear: he does not infer existence from the apparent consistency or other general features of defining characteristics, and he provides explicit statements of his fundamental principles. Nevertheless, as was presumably evident to Frege himself, the defense he offers of the fundamental law in question, *i.e.*, of Law V, is not up to his own standards of rigor. His central point in the passage just quoted is that some version of the law has 'always been made use of' in logic and mathematics; this thought is supplemented as the passage continues with a quite un-Fregean appeal to something like the claim that without such a law, we couldn't get what we want, namely a logical foundation for arithmetic:

> If there are logical objects at all — and the objects of arithmetic are such — then there must also be a means to grasp them, to recognise them. The basic law of logic which permits the transformation of the generality of an equality into an equality serves for this purpose. Without such a means, a scientific foundation of arithmetic would be impossible. For us it serves the purposes that other mathematicians intend to achieve by the

[13] For a helpful discussion of Frege's introduction of logical objects, see [Ruffino, 2003]. Ruffino argues that what qualifies extensions as 'logical' in Frege's treatment is their essential occurrence in all discourse about functions, as representatives of those functions. On this view, the fact that numbers, unless reduced to extensions, do not play such a role undermines any attempt to provide a parallel justification for the claim that objects introduced directly by (HP) are logical. For a contrary view, see [Schirn, 2006]. Though the account proposed here does not engage the question whether talk about concepts is, for Frege, talk of their extensions, it converges with both Ruffino and Schirn on the critical claim that extensions (value ranges) are *logical* objects for Frege in virtue of their essential role in logic. It differs from Ruffino in holding, with Schirn, that the question of the logical status of a collection of objects is secondary to the question of the logical status of the principle that underwrites their existence.

> creation of new numbers. Our hope is thus that from the eight functions whose names are listed in I, §31, we can develop, as from one seed, the whole wealth of objects and functions that mathematics deals with. Can our procedure be called a creation? The discussion of this question can easily degenerate into a quarrel about words. In any case, our creation, if one wishes so to call it, is not unconstrained and arbitrary, but rather the way of proceeding, and its permissibility, is settled once and for all. And with this, all the difficulties and concerns that otherwise put into question the logical possibility of creation vanish; and by means of our value-ranges we may hope to achieve everything that these other approaches fall short of. [Frege, 1903, §147]

The weakness of his own justification aside, the interesting point for our purposes is the foil against which Frege presents his own approach, that of the 'lawless' or 'unconstrained and arbitrary' practice of his fellow mathematicians. With his own approach, says Frege, 'all the difficulties and concerns that otherwise put into question the logical possibility of creation fall by the wayside'. Those difficulties and concerns, to recapitulate, are: (i) the unjustified inference from apparent consistency to consistency; (ii) the unjustified inference from consistency to existence; and (iii) the failure to lay down explicitly the principles employed.

In the setting of *Grundgesetze*'s strict demands on existence proofs, a project relying tacitly on a nontrivial but ungrounded principle like (S) will count as a retrograde move, a step back into precisely the kinds of 'difficulties and concerns' that Frege takes to plague the work of his contemporaries. The failure to include (S) explicitly in the basic principles of a science that relies on it is precisely difficulty (iii), and the existential inference it licenses, unless independently justified, instantiates versions of (i) and (ii).

4. RE-CARVING AGAIN

Frege seems to hold in *Grundlagen* that instances of (ABS) are self-evidently true, for reasons similar to those invoked by the neo-logicist in appropriate cases: there is something like an analytic equivalence between left-hand and right-hand sides. The strong requirement, found in *Grundgesetze*, that existential principles be explicitly laid out in advance is not yet present here. One might ask, then, whether despite its conflict with the strong proof-theoretic demands of *Grundgesetze*, the neo-logicist's strategy is in accord with the looser approach to existence found in *Grundlagen*.

The account of abstraction in *Grundlagen*, as applied to the case of directions, is one on which

(dir) $$\text{the direction of } a = \text{the direction of } b$$

is just a re-phrasing of what's expressed by the corresponding

(par) $$a \text{ is parallel to } b.$$

The idea would seem to be that the implication between (dir) and (par) is immediate and self-evident: anyone who was unsure whether the first followed from the second would thereby count as confused about the meaning of one or the other sentence. Similarly for the immediately-implied

(ex) $$\exists x(x = \text{the direction of } b);$$

understanding the sentences in question suffices for knowledge that (ex) follows from (dir), and that (dir) follows from (par).

One way to understand the central point is in terms of the sufficiency of a straightforward semantic intention. The intention to use the term-forming operator 'the direction of' in such a way that instances of (dir) re-state the contents of the corresponding instances of (par) is all that's required in order to give that term-forming operator this role. Hence that intention is all that's required, on this picture, to guarantee that the existence of directions, as expressed by (ex), follows immediately from the truth of an instance of (dir). In general: given an equivalence relation Φ, the intention to use an associated term-forming operator *the* f *of* ... in such a way that instances of (ABS) are analytically true is self-fulfilling, and this fact is self-evident.

The governing idea of the sufficiency of such an intention, though, would seem to be just what is undermined by the paradox. The intention to use a term-forming operator in a way that satisfies an instance of (ABS) can be shown to be unfulfillable in virtue of the non-existence of the required objects. Hence it shows that the existence of the objects is not a trivial consequence of the truth (or the falsehood) of the relevant instance of (EQU). That one intends to use abstract terms in accordance with an instance of (ABS) does not suffice to make the left-hand and the right-hand sides of that instance equivalent in content.[14]

When the background idea of the self-fulfilling nature of the abstractionist semantic intention is replaced by the more modest (S), we no longer have obvious counterexamples. But we have also strayed rather far from what appears to be the Fregean idea, that the equivalence between the two sides of (relevant) abstraction principles is self-evident, something that can be doubted only by those who are confused about the semantics of the singular terms in question. For it is not a sign of confusion about the meaning of the term-forming operator to be unsure about *e.g.*, the truth of (S), or of the claim that a particular abstraction principle satisfies the safety condition. Hence it is not a sign of confusion to be unsure about whether a given instance of (ID) is appropriately semantically similar to its corresponding instance of (EQU).

The replacement, then, of Frege's background assumption of the triviality of re-carving with a background assumption of something like (S) would seem to amount to the replacement of self-evident (hence immediately well-grounded) abstraction principles with abstraction principles whose justification rests on the non-trivial and as yet unjustified (S).

---

[14] This, I take it, is part of the point emphasized by Dummett in *e.g.*, [1991, Chap. 15].

5. MODERN RESULTS

It is worth pausing briefly over the independent question not of the justification for (HP), but of the sense in which (HP) can serve as a ground of arithmetic. Frege's thought was that the connection between (HP) and the truths of arithmetic was that of proof: every truth of arithmetic was to have been rigorously provable, via a finite sequence of logical inferences, from (HP). That this was so was to have been demonstrated by the provability from (HP) of Frege's own version of the Dedekind-Peano axioms, which he presumed would suffice for the proof of all of arithmetic. We now know that this natural hope was misguided: we cannot prove all of arithmetic from (HP), or from any recursive collection of fundamental principles.[15] The sense in which (HP) 'suffices' for arithmetic is that it (in virtue of implying the second-order axioms for arithmetic, which we know to be categorical) has as models only structures isomorphic to the natural numbers in their natural order. In short, every model of (HP) is a model of each sentence of true arithmetic.

One might wonder how closely this relationship between (HP) and arithmetic approximates the relationship that Frege sought to demonstrate. As we have seen above, the strategy of invoking models of sentences to demonstrate logical properties of the thoughts ordinarily expressed by those sentences is not altogether smooth from the Fregean point of view. The deductive relationship that Frege sought to establish between (HP) and arithmetic has a straightforward epistemological payoff: the existence of a finite proof of an arithmetical truth τ from (HP), using just logic, fairly immediately shows that one can, in principle, come to know τ just by means of knowledge of (HP) and possession of logical reasoning skills. The fact that every model of the sentence expressing (HP) is a model of the canonical sentence expressing τ has no such payoff. Here there are two difficulties. The first is that it is far from clear that truth preservation across models for such a pair of sentences gives evidence of an important justificatory link between the thoughts in question. The second is that the procedure that's required in order to determine that each model of the HP-sentence is a model of the τ-sentence involves non-trivial claims about set-theoretic structures, reasoning which is, especially in light of the paradox, very clearly not what Frege (or we) should count as 'purely logical'.

In short, the Fregean goal of providing axioms from which we might deduce each truth of arithmetic is significantly different from the goal pursued by his fellow logicist Dedekind, that of providing axioms that give a categorical description of the number-sequence. An axiomatization of the latter kind tells us which general features of structures already given will suffice to characterize an ω-sequence, but cannot tell us which fundamental truths suffice for the finite demonstration of each arithmetical truth. Here again, as with the distinction between Frege and the neo-logicist, the core of the difficulty in the non-Fregean

---

[15] We presuppose here the kind of Fregean realism that would guarantee the expressibility of the truths of arithmetic as a complete set of sentences of a language of arithmetic.

program is the absence of an appropriate grounding for the existence claims of arithmetic. And as the incompleteness of second-order logic shows us, there is no clear path from the kind of foundation provided by Dedekind's project to the kind sought by Frege: no way to infer the epistemically significant grounding of a body of truths from sentences known merely to characterize the isomorphism-class of their home structure. Hence the sense in which (HP) 'suffices' for arithmetic is not in any immediate way a sense that would have sufficed for Frege's purposes.

6. ASSESSMENT

It is agreed by modern neo-logicists and their anti-neo-logicist interlocutors that (HP) is reliable: it is provably consistent (in a post-Fregean sense); so we know at least that as a foundational principle it won't be disastrous in the way that Law V was. The point on which the two camps are divided is the question whether (HP) is sufficiently content-free, sufficiently 'analytic', that its use as a foundation for arithmetic would underwrite the claim that arithmetic is analytic.

Frege's views are arguably not compatible with those of either camp. As outlined above, the use of a model to demonstrate the 'consistency of (HP)' demonstrates the truth of the thought assigned to the HP-sentence by that model. It also demonstrates the satisfiability of the general structural condition implicitly defined by that sentence. But neither of these suffices to answer what Frege would have considered the important consistency question, that of the thought expressed by the HP-sentence when its quantifiers range over all first-level concepts, and its numerical term-forming operator refers to the real numberer. Frege thinks, no doubt, that (HP) *is* consistent in this important sense, and continues to think this after the paradox: he does not think that the coherence of our practice of assigning numbers to concepts is in any way thrown into doubt by Russell's letter. But to know or to firmly believe that (HP) is consistent is not to have a proof of its consistency. And, critically, neither its known nor its proven consistency offers from Frege's point of view a means of demonstrating its truth. Neither, as we have seen, does its satisfaction of strong stability or other safety conditions of the kind central to the neo-logicist project.

Frege's strong requirement that mathematical proof cannot involve any tacit existential presuppositions means that the effect of Russell's paradox, in undermining both the general abstractionist strategy of *Grundlagen* and its refinement as Law V of *Grundgesetze*, leaves a large gap in the foundation. Frege's own final thoughts on the matter were that no other principle of logic would be able to fill the gap: the 'logical source of knowledge', he suggests, is insufficient to provide objects, with the result that we are likely to need to turn to geometry, a science whose existential claims are grounded in intuition, for the needed proofs of existence.[16] The upshot of the preceding considerations is

---

[16]Frege's unpublished 'A new attempt at a foundation for arithmetic', from 1924–25, says of the 'logical source of knowledge' that: 'it seems that this on its own cannot yield

that the idea of reformulating the foundations of the logicist project so as to count (HP) as fundamental would have amounted, from Frege's point of view, as the proposal simply to ignore the gap.

This does not yet provide reason to prefer Frege's own strategy, with an explicit proof of existence (perhaps via geometry) to the neo-logicist strategy. The question of which such strategy is to be preferred, if either, will turn in part on whether Frege's strict requirements for existence proofs are reasonable. One way to put the question is to ask to what extent we ought to follow Frege in taking mathematical objects to be just like other objects, in the sense that the question of their existence is always a further question after we have settled general facts about the consistency, fruitfulness, categoricity, and so on of their defining conditions. One might hold on the contrary that mathematical objects, or objects reached via some non-trivial process of abstraction, have the special characteristic that their existence follows, *contra* Frege, from some general feature of their defining conditions. One thing that the preceding discussion is meant to clarify is that such a conception of existence will involve a significant departure from Frege's own, not unreasonable, view. And while the paradox obviously undermines Frege's pre-1902 conception of the innocence of re-carving and of the status of value ranges, it would not by itself seem to offer any reason to reject his strong views about the grounding of mathematical existence statements.

REFERENCES

Blanchette, Patricia [1996]: 'Frege and Hilbert on consistency', *The Journal of Philosophy* **43**, 317–336.

——— [2007]: 'Frege on consistency and conceptual analysis', *Philosophia Mathematica* (3) **15**, 321–346.

——— [2012]: *Frege's Conception of Logic*. Oxford University Press.

Boolos, George [1987]: 'The consistency of Frege's *Foundations of Arithmetic*', in J. Thompson, ed, *On Being and Saying: Essays in Honor of Richard Cartwright*, pp. 3–20. Cambridge, Mass.: MIT Press. Reprinted in [Boolos, 1998], pp. 183–201. Page references are to this reprint.

——— [1990]: 'The standard of equality of numbers', in George Boolos, ed., *Meaning and Method: Essays in Honor of Hilary Putnam*, pp. 261–277. Harvard University Press. Reprinted in [Boolos, 1998], pp. 301–314. Page references are to this reprint.

——— [1998]: *Logic, Logic and Logic*. Richard Jeffrey, ed. Harvard University Press.

Boolos, George, and Richard Heck [1998]: '*Die Grundlagen der Arithmetik* §§82–83', in [Boolos, 1998], pp. 315–338.

Burgess, John P. [2005]: *Fixing Frege*. Princeton University Press.

Cook, Roy [2012]: 'Conservativeness, stability, and abstraction', *British Journal for Philosophy of Science* **63**, 673–696.

us any objects', and 'Since probably on its own the logical source of knowledge cannot yield numbers . . ., we will appeal to the geometrical source of knowledge.' ([Frege, 1983, pp. 298–302]; English translation in [Frege, 1979, 278–281])

Demopoulos, William, ed. [1994]: *Frege's Philosophy of Mathematics*. Harvard University Press.

Dummett, Michael [1991]: *Frege: Philosophy of Mathematics*. Harvard University Press.

Fine, Kit [2002]: *The Limits of Abstraction*. Oxford University Press.

Frege, Gottlob [1884]: *Die Grundlagen der Arithmetik, Eine logisch mathematische Untersuchung über den Begriff der Zahl*. Breslau: Wilhelm Koebner. English translation [Frege, 1953]; quotations are from this translation.

——— [1885]: 'Über formale Theorien der Arithmetik', *Sitzungsberichte der Jenaischen Gesellschaft für Medizin und Naturwissenschaft, Jenaische Zeitschrift für Naturwissenschaft* **19**, 94–104. Reprinted in [Frege, 1990], pp. 103–111. English translation in [Frege, 1984], pp. 112–121.

——— [1893]: *Grundgesetze der Arithmetik, Band I*. Jena: Hermann Pohle. English translation [Frege, 2013]; quotations are from this translation.

——— [1903]: *Grundgesetze der Arithmetik, Band II*. Jena: Hermann Pohle. English translation [Frege, 2013]; quotations are from this translation.

——— [1953]: *The Foundations of Arithmetic*. English translation of [Frege, 1884], J.L. Austin, trans. Oxford: Blackwell. Quotations are from this translation.

——— [1979]: *Posthumous Writings*. Chicago University Press. H. Hermes, F. Kambartel, and F. Kaulbach, eds. P. Long and R. White, trans. This is an English translation of most of [Frege, 1983].

——— [1980]: *Philosophical and Mathematical Correspondence*. G. Gabriel, H. Hermes, F. Kambartel, C. Thiel, and A. Veraart, eds. Abridged from the German edition by B. McGuiness. H. Kaal, trans. University of Chicago Press.

——— [1983]: *Nachgelassene Schriften*. 2nd rev. ed. H. Hermes, F. Kambartel, and F. Kaulbach, eds. Hamburg: Felix Meiner Verlag.

——— [1984]: *Collected Papers on Mathematics, Logic, and Philosophy*. Oxford: Blackwell. B. McGuinness, ed. Translated by M. Black, V.H. Dudman, P. Geach, H. Kaal, E.-H.W. Kluge, B. McGuinness, and R.H. Stoothoff. This is a translation of most of [Frege, 1990].

——— [1990]: *Kleine Schriften*. I. Angelelli, ed. 2nd ed. Hildesheim: Georg Olms.

——— [2013]: *The Basic Laws of Arithmetic*. Oxford University Press. [Frege, 1893] and [Frege, 1903] translated and edited by P. Ebert and M. Rossberg with C. Wright. Quotations are from this translation.

Hazen, A. [1985]: 'Review of [Wright, 1983]', *Australasian Journal of Philosophy* **63**, 250–254.

Heck, Richard [1992]: 'On the consistency of 2nd-order contextual definitions', *Noûs* **26**, 491–495.

——— [1993]: 'The development of arithmetic in Frege's *Grundgesetze der Arithmetik*', *The Journal of Symbolic Logic* **58**, 579–601.

Hodes, Harold [1984]: 'Logicism and the ontological commitments of arithmetic', *The Journal of Philosophy* **81**, 123–149.

Linnebo, Øystein [2009]: 'Introduction [to special issue on the Bad Company Problem]', *Synthese* **170**, 371–391.

Linnebo, Øystein, and Gabriel Uzquiano [2009]: 'Which abstraction principles are acceptable? Some limitative results', *British Journal for Philosophy of Science* **60**, 239–252.

Parsons, Charles [1965]: 'Frege's theory of number', in Max Black, ed, *Philosophy in America*, pp. 180–203. Ithaca, N.Y.: Cornell University Press. Reprinted with postscript in [Parsons, 1983]. Reprinted (with same postscript) in [Demopoulos, 1994].

——— [1983]: *Mathematics in Philosophy: Selected Essays*. Ithaca, N.Y.: Cornell University Press, 1983.

Ruffino, Marco [2003]: 'Why Frege would not be a neo-fregean', *Mind* **112**, 51–78.

Schirn, Matthias [2006]: 'Concepts, extensions, and Frege's logicist project', *Mind* **115**, 983–1005.

Weir, Alan [2003]: 'Neo-fregeanism: An embarrassment of riches', *Notre Dame Journal of Formal Logic* **44**, 13–48.

Wright, Crispin [1983]: *Frege's Conception of Numbers as Objects*. Aberdeen University Press.

——— [1997]: 'The philosophical significance of Frege's theorem', in Richard Heck, ed, *Language, Thought and Logic: Essays in Honour of Michael Dummett*. Oxford: Clarendon Press.