Synthese (2009) 170:415–441
DOI 10.1007/s11229-007-9263-9

The good, the bad and the ugly

Philip Ebert · Stewart Shapiro

Accepted: 31 August 2007 / Published online: 16 October 2007

Abstract This paper discusses the neo-logicist approach to the foundations of mathematics by highlighting an issue that arises from looking at the Bad Company objection from an epistemological perspective. For the most part, our issue is independent of the details of any resolution of the Bad Company objection and, as we will show, it concerns other foundational approaches in the philosophy of mathematics. In the first two sections, we give a brief overview of the "Scottish" neo-logicist school, present a generic form of the Bad Company objection and introduce an epistemic issue connected to this general problem that will be the focus of the rest of the paper. In the third section, we present an alternative approach within philosophy of mathematics, a view that emerges from Hilbert's *Grundlagen der Geometrie* (1899, Leipzig: Teubner; *Foundations of geometry* (trans.: Townsend, E.). La Salle, Illinois: Open Court, 1959.). We will argue that Bad Company-style worries, and our concomitant epistemic issue, also affects this conception and other foundationalist approaches. In the following sections, we then offer various ways to address our epistemic concern, arguing, in the end, that none resolves the issue. The final section offers our own resolution which, however, runs against the foundationalist spirit of the Scottish neo-logicist program.

Keywords Abstraction principles · Neo-Logicism · Frege · Hilbert · Foundationalism · Implicit definitions · Basic knowledge

P. Ebert (✉)
Department of Philosophy, University of Stirling, Stirling FK9 4LA, Scotland, UK
e-mail: philipebert@mac.com

S. Shapiro
350 University Hall, 230 North Oval Mall, Columbus, OH, 43210, USA
e-mail: shapiro.4@osu.edu

1 Introduction

This paper discusses the neo-logicist approach to the foundations of mathematics by highlighting an issue that arises from looking at the *Bad Company* objection from an epistemological perspective. For the most part, our issue is independent of the details of any resolution of the Bad Company objection and it concerns other foundational approaches in the philosophy of mathematics.

After giving a very brief overview of the "Scottish" neo-logicist school, we present a generic form of the Bad Company objection and introduce an epistemic issue that goes with it. Then we present an alternative approach to the philosophy of mathematics, a view that emerges from Hilbert's *Grundlagen der Geometrie* (1899). This highlights the fact that Bad Company-style worries, and our concomitant epistemic issue, affect different conceptions within the philosophy of mathematics. We then offer various ways to address the epistemic concern, arguing that none resolves the issue. The final section offers our own resolution which, however, runs against the foundationalist spirit of the Scottish neo-logicist program.

2 The Scottish neo-logicist program

The Scottish neo-logicist program is to introduce mathematical theories via abstraction principles. The generic form of a second-order abstraction principle is:

(ABS) $\forall F \forall G(\Sigma(F) = \Sigma(G) \equiv E(F,G))$,

where Σ is a higher-order operator, denoting a function from properties (or concepts, or sets, or whatever is in the range of the monadic second-order variables) to individual objects, and E is an equivalence relation over properties (or concepts, or sets, or whatever).

The founding instance of the program is an abstraction now known as Hume's Principle (HP):

(HP) $\forall F \forall G(\#F = \#G \equiv (F \approx G))$,

where '$\#F$' is the cardinal number of F, and '$F \approx G$' is the statement that the F's are equinumerous with the G's. Hume's Principle and second-order logic suffices to derive the Peano axioms and so to account for arithmetic.

The program is to understand abstraction principles as a special form of implicit definition which, when successful, is true by stipulation. Abstraction principles, so conceived, are meant to account for a priori knowledge of arithmetic, analysis and, hopefully, set theory. Our focus is thus on the epistemological aspects of Scottish neo-logicism, the claims concerning the nature of mathematical knowledge. Other versions of logicism and neo-logicism have a more metaphysical focus, concerning the nature of mathematical truth. For example, some take the program to be that arithmetic is a body of conceptual truths. The present paper does not directly address those metaphysical concerns.

3 The Good, the Bad and the Ugly Company

The Bad Company objection is best seen as aimed at the epistemological aspects of neo-logicism. The original form of the objection is that abstraction principles cannot be a legitimate way to introduce mathematical theories, since some of them are inconsistent. The most notorious example, of course, is Gottlob Frege's own Basic Law V:

$$\forall F \forall G(\text{Ext}F = \text{Ext}G \equiv \forall x(Fx \equiv Gx)).$$

In like manner, an analogous abstraction on well-orderings (or relations generally) is inconsistent, falling to the Burali-Forti paradox. The argument is that if (HP) is good purely based on its form—if (HP) can ground knowledge of a branch of mathematics—then so are these other ones, since they have the same form (ABS). However, Basic Law V cannot be used to found a theory of sets or extensions, and an ordinal abstraction principle cannot be used to introduce a theory of ordinals (Weir (1998), and dialetheism, ala Priest (1987), notwithstanding). So (HP) cannot be used to introduce a theory of cardinal numbers, or so the argument goes.

Neil Tennant (1987, 236–237) raises an early formulation of the objection. He notes that the account of arithmetic developed by Crispin Wright (1983) may "with justice be called a naive theory of number". Tennant adds that this "is not to say that [Wright's theory] is inconsistent; it is only to stress the analogy with the naive theory of sets", based on Basic Law V.

Much of the contemporary discussion of the objection traces to articulations due to Michael Dummett (1991, 188–189) and George Boolos's (1997). Retrospectively, Dummett (1998, 375) observed the following:

> My complaint was the obvious one. In *Grundgesetze*, value-ranges are introduced in a manner precisely analogous to that in which Wright argued, in his book, that Frege ought to have introduced the cardinal numbers ...: and yet it was so far from being justified as to lead to actual contradiction. It therefore *could* not be maintained that this procedure is, in and of itself, legitimate.

The problem, according to Dummett, is that the neo-logicist procedure of stipulating abstraction principles cannot be regarded as reliable in generating knowledge-grounding statements.

Clearly, this is not a knock-down argument. Rather it raises a challenge, as Dummett (1998) continues: "Possibly some restriction, distinguishing the case of cardinal numbers from that of value-ranges could be framed. But Wright had put forward no such restriction; and hence his thesis, as it stood, could not be sound." And so the hunt was on to find appropriate constraints or restrictions to separate "good" cases—the knowledge-grounding abstraction principles—from the "bad" ones, and so resolve the Bad Company objection.

Consistency seems to be a necessary condition for a good abstraction (again, dialetheism notwithstanding). This suffices to separate Basic Law V and the ordinal abstraction from Hume's Principle. Yet, consistency alone is not enough to deal with other "bad" cases that were put forth by Richard Heck (1992), Boolos's (1997), and Wright (1997) himself. Heck, for example, proposes a scheme:

(HeΦ) $\forall F \forall G(\mathrm{E}F = \mathrm{E}G \equiv (\Phi \vee \forall x(Fx \equiv Gx)))$,

where Φ is a closed sentence in the second-order language, with no non-logical terminology. It is easy to see that the embedded formula $[\Phi \vee \forall x(Fx \equiv Gx)]$ defines an equivalence relation, and so (HeΦ) is in the form (ABS). Heck observed that (HeΦ) is satisfiable only if Φ is. Indeed (HeΦ) entails Φ. If Φ is the statement that there are no more than finitely many objects, then (HeΦ) is satisfiable only on finite domains. In contrast, (HP) is satisfiable only on (Dedekind) infinite domains, since it implies the existence and uniqueness of each natural number. So (HP) and (HeΦ) cannot both be true.

Similarly, Boolos's (1997) Parity Principle and Wright's (1997) Nuisance Principle are both satisfiable on any finite domain, but each of them is inconsistent with (HP) since neither is satisfiable on an infinite domain. In addition, Alan Weir (2003) has formulated a series of so-called "distraction principles". One of those is satisfiable only on limits in the series of inaccessible cardinals and another is satisfiable only on successors in the series of inaccessible cardinals. Again, only some of the distractions can be good, otherwise we end up with what Weir calls the "embarrassment of riches": there are too many presumably acceptable principles for them all to be acceptable. Clearly, then, unless further, well-motivated constraints are offered to isolate the "good" principles, the procedure of the neo-logicist has to be regarded as unreliable and so futile.

Over the last decade or so, this debate has seen various epicycles, producing constraints (conservativeness, modesty, stability/irenicity) on legitimate ("good") abstraction principles, and the logical relations among those constraints have been studied. It is perhaps fair for a neutral observer to conclude that the jury is still out on whether the Bad Company objection can be met—hence this special issue.

There is a closely related matter, which we contentiously label *Good Company*. Once the right criteria have been identified to rule out mutually inconsistent statements, why is it that only abstraction principles such as Hume's Principle are "good" and not also other postulations, such as the Dedekind-Peano axioms? In other words, why is the neo-logicist procedure—that of stipulating certain sentences to express truths—only applicable to abstraction principles, and not also to other statements of a different form?[1]

Stipulations of sentences that are not of the form of an abstraction principle will prove important for us, largely for purposes of illustration and to broaden the scope of Bad Company and its ultimate resolution. Yet, we are here not directly concerned with the exact criteria to separate the good postulations from the bad ones (i.e. Bad Company), nor with the scope of the good cases is (i.e., Good Company). Our focus instead is on the question of what—provided that the criteria for a stipulation to be good are understood—is the *epistemic status* of these criteria. That is, what does a theorist, say Wright's (1998) character "Hero", have to know or justifiably believe about the relevant criteria in order to acquire mathematical knowledge on the basis of abstraction principles or other stipulations? This issue turns out to be more problematic than one might think, and it is here that we think that the Bad Company objection turns *ugly*. Also, we believe that it is an issue that applies beyond the neo-logicist program, and

[1] See John MacFarlane's and Bob Hale and Crispin Wright's contributions to this special issue.

affects other epistemic foundationalist enterprises in the philosophy of mathematics. But, as we will see, it proves particularly ugly with respect to the neo-logicist approach.

4 The use of implicit definition in Hilbert's early program

It will prove instructive to consider another perspective on at least some branches of mathematics that also invokes stipulations of sentences, i.e. *implicit definitions*. These have roughly the same foundationalist outlook as the neo-logicist adopts with respect to legitimate abstraction principles. As we will see, these implicit definitions have their own Bad Company-style issues. Seeing how those are resolved in philosophical theory and in mathematical practice sheds light on the similar (perhaps identical) issue concerning neo-logicist abstractions.

David Hilbert's *Grundlagen der Geometrie* (1899) marked the culmination of a trend toward formalization in geometry. Hilbert was aware that, at some level, spatial intuition or observation remained the source of the axioms of geometry. In his *Grundlagen*, however, the role of intuition or observation is limited to motivation and heuristic. The book does not contain the phrase "implicit definition", but at least with hindsight, it is clear that it delivers, and is based upon, implicit definitions of geometric structures. The early pages contain phrases like "the axioms of this group define the idea expressed by the word 'between' ..." and "the axioms of this group define the notion of congruence or motion". The idea is summed up as follows:

> We think of ... points, straight lines, and planes as having certain mutual relations, which we indicate by means of such words as "are situated", "between", "parallel", "congruent", "continuous", etc. The complete and exact description of these relations follows as a consequence of the *axioms of geometry*.

Of course, Hilbert also says that the axioms express "certain related fundamental facts of our intuition", but in the mathematical development, all that remains of this "intuitive content" is the use of *words* like "point", "line", etc., and perhaps the diagrams that accompany some of the theorems.

One key idea is that the axiomatizations—stipulations—of the *Grundlagen* are free-standing, in the sense that *anything at all* can play the role of the undefined primitives of points, lines, planes, etc., so long as the axioms are satisfied. Otto Blumenthal reports that in a discussion in a Berlin train station in 1891, Hilbert said that in a proper axiomatization of geometry, "one must always be able to say, instead of 'points, straight lines, and planes', 'tables, chairs, and beer mugs'."[2] The connection to intuition or observation is broken for good. This approach bled into other areas of mathematics, such as arithmetic and analysis.

An implicit definition, in this sense, is the presentation of some axioms that contain one or more "primitive" terms. In the case of geometry, these might be "point", "line", and "plane". In the case of arithmetic, the primitives might be "number", "successor", "plus", and "times". The axiomatization characterizes the relations among the primitives, and only the relations among the primitives. Let us call this a "Hilbert-style

[2] "Lebensgeschichte" in Hilbert (1935, 388–429); the story is related on p. 403. See Stein (1988, 253), Coffa (1991, 135), and Hallett (1990, 201–202).

implicit definition". It is similar to a functional definition in the philosophy of mind (Shapiro 1997, Chap. 3, Sect. 6). In mathematics, if an implicit definition of this type is successful, then at least one structure is characterized by it, and the axioms are true of the structures, or structure, so characterized.

Frege himself objected to Hilbert's deployment of implicit definitions.[3] At least Hilbert's side of this dispute sheds some light on the perspective under study here. In a letter dated December 27, 1899, Frege articulated a traditional view concerning axioms and definitions. He claimed that in Hilbert (1899),

> ... the meanings of the words "point", "line", "between" are not given, but are assumed to be known in advance ... [I]t is also left unclear what you call a point. One first thinks of points in the sense of Euclidean geometry, a thought reinforced by the proposition that the axioms express fundamental facts of our intuition. But afterwards you think of a pair of numbers as a point ... Here the axioms are made to carry a burden that belongs to definitions ... [B]eside the old meaning of the word "axiom", ... there emerges another meaning but one which I cannot grasp.

For Frege, a definition should specify the meaning of a single word whose meaning has not yet been given, and it should only employ other words whose meanings are already known. In contrast, an axiom

> ... must not contain a word or sign whose sense and meaning, or whose contribution to the expression of a thought, was not already completely laid down, so that there is no doubt about the sense of the proposition and the thought it expresses. The only question can be whether this thought is true and what its truth rests on. Thus axioms and theorems can never try to lay down the meaning of a sign or word that occurs in them, but it must already be laid down.

In sum, Frege's point is that if the terms in the proposed axioms do not have meaning beforehand, then the statements cannot be true (or false), and thus they cannot be axioms. If, on the other hand, the primitives do have meaning beforehand, then the axioms cannot be definitions. Axioms should express *truths* and definitions should give the *meanings* and *fix the denotations* of certain terms. Frege argued that with a Hilbert-style implicit definition, neither job is accomplished. Hilbert rejected the dilemma. If the axiomatization is successful, then *both* jobs are accomplished at once.

Ironically, the neo-logicist agrees with Hilbert on this issue. He regards Hume's Principle as a meaning-constituting principle that provides a first grasp of the newly-introduced term "the (cardinal) number of". Moreover, the meaning of the newly introduced term is fixed by stipulation that the statement in which it occurs—Hume's Principle—is to express a truth. Hence for the neo-logicist, as for Hilbert, the two jobs are accomplished at once. If all goes well, the statement is an axiom that expresses a truth and a meaning is fixed.

[3] The much discussed correspondence between Frege and Hilbert is published in Frege (1976) and translated in Frege (1980).

In his reply, dated December 29, Hilbert addressed Frege's claim that the meanings of the words "point", "line", and "plane" are "not given, but are assumed to be known in advance":

> This is apparently where the cardinal point of the misunderstanding lies. I do not want to assume anything as known in advance. I regard my explanation . . . as the definition of the concepts point, line, plane . . . If one is looking for other definitions of a "point", e.g. through paraphrase in terms of extensionless, etc., then I must indeed oppose such attempts in the most decisive way; one is looking for something one can never find because there is nothing there; and everything gets lost and becomes vague and tangled and degenerates into a game of hide and seek.

Clearly though, not every proposed axiomatization is a successful implicit definition: some are inconsistent; some are not satisfiable. In these bad cases, no structure is characterized, and the axioms are not true by stipulation. So, similarly, we have to ask which are the good or successful cases of Hilbert-style stipulation, which are the bad cases and, importantly, what are the criteria that characterize the good ones and the bad ones respectively? This is the *Bad Company* objection. As with neo-logicism, we ask for criteria for determining when all has gone well with a stipulation.

For Hilbert the only requirement for the success of a stipulation is that an axiomatization be "consistent". It is not clear, however, whether he was thinking in terms of deductive consistency or satisfiability at that time, and it may be too much of an anachronism to insist on the distinction. To shore up various collections of axioms, Hilbert interprets them in various systems, often the real numbers. With hindsight, and a bit of anachronism, we can think of this as establishing either the satisfiability of the axiomatization, or else its relative deductive consistency: if real analysis is consistent, then so is the given axiomatization.

From Frege's perspective, in contrast, there is no need to concern oneself with the consistency of an axiomatization. The axioms are supposed to be (self-evident) *truths* about a given subject matter. He noted that from the truth of axioms, "it follows that they do not contradict one another". Hilbert emphatically rejected this more or less traditional perspective:

> . . . As long as I have been thinking, writing and lecturing on these things, I have been saying the exact reverse: if the arbitrarily given axioms do not contradict each other with all their consequences, then they are true and the things defined by them exist. This is for me the criterion of truth and existence.

Hilbert then repeated the role of what we are calling a Hilbert-style implicit definition, noting that it is impossible to give a definition of "point" in a few lines since "only the whole structure of axioms yields a complete definition". That is, this is the *only* way to define mathematical terms.

To be sure, there are important differences between Hilbert-style and neo-logicist stipulations. Each non-logical term in a (successful) Hilbert-style implicit definition stands for a piece—an object, property, function, or relation—of the structure or structures being defined. With the exception of the logical terminology (connectives, etc.), *no* term in a Hilbert-style implicit definition comes with a previously established meaning or extension. More importantly, the quantifiers in a Hilbert-style axiomatization range over (and only over) the places or concepts in the structure. Hilbert's axiomatization of geometry, for example, only has quantifiers that range over "points", "lines", and "planes"—the very items whose structure is being defined. Let us say that a Hilbert-style implicit definition is *self-contained*.

This is in stark contrast to typical abstractions. Recall the form of Hume's Principle:

(HP) $\forall F \forall G(\#F = \#G \equiv (F \approx G))$,

where '$F \approx G$' is the statement that the F's are equinumerous with the G's. Suppose we (try to) think of this as a Hilbert-style implicit definition. Notice, first, that (HP) has only one non-logical item, the '#' operator that is being defined, and the introduced ontology consists of the cardinal numbers. So, if (HP) is understood Hilbert-style, the first-order variables range over, *and only over*, cardinal numbers, and the second-order variables range over concepts or properties (or sets) of cardinal numbers. So construed, (HP) only allows us to count—assign cardinalities to—concepts of numbers.[4]

Clearly, (HP) is not to be understood in this way. The point of Scottish neo-logicism is that one can introduce talk of numbers as a "re-carving" of statements of equinumerosity concerning *any objects whatsoever*. Indeed, this is one of the motivations for logicism itself. The observation is that *any* concept (or at least any finite, sortal concept) has a cardinal number. Like logic, arithmetic applies to all subjects. As long as one is talking about objects at all, one can count them. So the opening quantifiers in (HP) are unrestricted.[5]

The general point here is related to what is sometimes called "Frege's constraint", the thesis that the proper foundation for a mathematical theory should flow from its typical applications (Hale 2000; Wright 2000). The logicist and neo-logicist alike argue that we must start with an account of how arithmetic applies to ordinary situations. This leads directly to Hume's Principle. To be sure, a theory introduced with a Hilbert-style implicit definition may find application, but an account of these applications must be tacked on separately.

There is a further crucial difference between Hilbert-style implicit definitions and neo-logicist abstractions like (HP). On this matter, the neo-logicist is closer to Frege. For Hilbert, the satisfiability (or relative consistency) of the set of axioms is sufficient for their truth, whereas for the neo-logicist, a crucial issue is also the uniqueness of the objects referred to by the relevant terms involved. At most, Hilbert-style definitions

[4] Hume's Principle, understood as a Hilbert-style implicit definition, would be of interest to a Pythagorean, who believes that All is Number. More seriously, and for what it is worth, (HP), so construed, is satisfiable in a domain of size κ if and only if there are exactly κ-many cardinal numbers (in the sense of ZFC) less than or equal to κ. That is, the implicit definition defines a structure on $\aleph_0$ and on any fixed point in the series of alephs (i.e., any cardinal κ such that $\kappa = \aleph_\kappa$). If we allow interpretations whose size is not a set, and if global choice holds, then (HP), so construed, is satisfiable on V (assuming that ZF correctly describes the pure sets). Notice, incidentally, that the derivation of the Dedekind-Peano axioms from (HP), what is now called "Frege's theorem", still works when (HP) is understood in this Hilbert-style way. Frege's theorem makes no assumptions, one way or the other, concerning whether there are non-numbers in the range of the quantifiers. Thanks to Roy Cook here.

[5] Some philosophers and logicians demur from the possibility of unrestricted quantification (see, for example, Rayo and Uzquiano (2006)). We need not broach those issues here; one can think of (HP) as systematically ambiguous. The point here is that its quantifiers range over arbitrary objects, not just cardinal numbers.

require only that there be sufficient objects to satisfy the system (whether they be real numbers, beer mugs, or tables and chairs). The neo-logicist demands that the resulting expressions function as singular terms and that they refer to independently existing objects. As we shall soon see, these differences between the programs do not prevent both from having Bad Company issues.

5 Bad Company objection and the epistemic status of the criteria

As noted, it is clear that not every attempt to give a Hilbert-style implicit definition succeeds. In particular, a proposed axiomatization may not define anything—there may be nothing that answers to the axioms. One obvious way that an axiomatization can fail is if it is inconsistent. The Dedekind-Peano axioms, together with a statement that there is a largest prime number, would be a failed axiomatization. So was the original presentation of the λ-calculus.

Deductive consistency is thus necessary for acceptability of a Hilbert-style implicit definition. Arguably, it is not sufficient. Let PA be the conjunction of the second-order Dedekind-Peano axioms for arithmetic, and let CON_{PA} be a typical arithmetization of the statement that PA is consistent. Kurt Gödel's second incompleteness theorem entails that PA together with $\neg CON_{PA}$ is consistent. But this "theory" has no models: PA is categorical, in that all of its models are isomorphic to the natural numbers and CON_{PA} is true on the natural numbers (presumably) and thus in all models of PA. So PA together with $\neg CON_{PA}$ is not satisfiable. So any attempt to stipulate $PA+\neg CON_{PA}$ as a Hilbert-style implicit definition would fail just about as badly as an attempt to stipulate PA and the statement that there is a largest prime number, or, for that matter, an attempt to stipulate $PA+\neg PA$. One can, of course, do deductions from these various premises, but there is no reason to think that any consequence of these stipulations is true of the natural numbers, or of anything else for that matter.[6]

Of course, it won't help much to lay down a principle that a Hilbert-style stipulation is acceptable only if the axiomatization is satisfiable, in some intuitive sense. This is just to say that a stipulation is acceptable only if there is a structure it characterizes. We were wondering just which stipulations characterize structures.

Shapiro (1997) introduces a new primitive, called "coherence", and states that an implicit definition defines a structure if (and only if) it is coherent. It is not possible

[6] This conclusion depends on interpreting the second-order quantifiers in Hilbert-style implicit definitions via standard semantics. We assume, for example, the monadic property variables range over every collection of objects in the domain. If the theories are first-order, or second-order with Henkin semantics, then deductive consistency does suffice for satisfiability, in light of Gödel's and Henkin's completeness theorems. Neo-logicism also invokes second-order languages. The opening quantifiers in Hume's Principle are second-order, and the embedded definition of equinumerosity is second-order, since it invokes a quantifier over relations. The aforementioned point that certain abstraction principles are satisfiable only on finite domains also relies on standard semantics. This matter is a staple in the literature on neo-logicism. On the other hand, neo-logicism is perhaps more concerned with deduction than with model-theoretic satisfiability. Even so, its derivations rely on impredicative instances of the comprehension scheme (Linnebo 2004). If we think of the comprehension scheme as expanding with the linguistic resources, similar points apply. A sentence analogous to $\neg CON_{PA}$ is inconsistent with the theory consisting of Hume's Principle plus a Tarskian truth-predicate, and with a sufficiently rich theory of natural numbers and sets of natural numbers. Moreover, much of the model theory can be recaptured from within. See Shapiro (1991). Nevertheless, some of the points below do not turn on whether the semantics is standard.

to give a non-trivial mathematical definition of coherence, but perhaps we can say enough about it to convey the idea.[7] Or perhaps not. We need not engage that issue here. Clearly however, there are some conditions which an acceptable Hilbert-style implicit definition must meet. Call these conditions $\mathbf{C_H}$. The idea is that an axiomatization is successful—defines at least one structure—only if it meets $\mathbf{C_H}$. We need not speculate on the details of these conditions here, except to insist that consistency is among them.

We extend the same courtesy to the neo-logicist. It is clear that any acceptable abstraction principle must be consistent, and, as above, it is also clear that consistency is not sufficient for acceptability. As noted, there are consistent abstractions that are satisfiable only on finite universes (see note 6). Such an abstraction is incompatible with Hume's Principle. There is an abstraction that is satisfiable only on limits in the series of inaccessibles and there is an abstraction that is satisfiable only on successors in the series of inaccessibles (Weir 2003). Each of these is consistent and, indeed, satisfiable if there are enough inaccessible cardinals, and yet the two abstractions cannot both be acceptable—they are not jointly satisfiable. Over the years, neo-logicists have developed a number of criteria on successful abstractions: consistency, conservativeness, etc. The details are not under scrutiny here. We just note that there are conditions $\mathbf{C_F}$ on acceptable abstractions. The idea is that an abstraction is successful—introduces a type of abstract and is true of said abstracts—only if it meets $\mathbf{C_F}$. We only insist that consistency is among these conditions. The conditions $\mathbf{C_F}$ and $\mathbf{C_H}$ may be the same or they may be different. We do not care.

6 Epistemic status of the $\mathbf{C_H}$ and $\mathbf{C_F}$

Suppose that a mathematician, call her *Emma*, attempts a Hilbert-style implicit definition. And suppose that Wright's character *Hero* attempts a neo-logicist abstraction. Again, Emma's stipulation succeeds only if $\mathbf{C_H}$ is met and Hero's stipulation succeeds only if $\mathbf{C_F}$ is met. Our central question here concerns the epistemic status of the conditions $\mathbf{C_H}$ and $\mathbf{C_F}$. What does Emma or Hero have to know or at least be justified in believing about $\mathbf{C_H}$ or $\mathbf{C_F}$, respectively, before they can be credited with knowledge of the stipulated statement? We explore several options below, but find none satisfactory, at least when it comes to the epistemological goals of neo-logicism.

[7] Coherence is loosely modeled by the mathematical notion of satisfiability. Recall that as the notion is defined, a set of sentences is satisfiable if and only if there is a *set* that satisfies it. There are two potential problems. First, there might be axiomatizations that are true if interpreted on a proper class, like V, but not on any set. In this case, the mathematical notion of satisfiability may not coincide with the one relevant to Hilbert-style implicit definitions. This won't be problematic if a certain reflection principle holds, yielding the existence of so-called small large cardinals (Shapiro (1987), (1991, Chapter 6, Sect. 6.3)). More importantly, the mathematical notion of satisfiability presupposes a theory of sets (or classes), together with an account of how a given axiomatization is satisfied in a given set (or class). Do we obtain *that* theory with a Hilbert-style implicit definition? Does the latter have to be satisfiable as well? In what? Regress threatens—depending on how one construes the foundational issues. We return to this issue in the final section.

6.1 Ya really gotta know

The first and perhaps most natural option is a requirement that Emma (or Hero) must *be in a position to show* that the conditions $\mathbf{C_H}$ (or $\mathbf{C_F}$) are met in order to be credited with knowledge of the implicit definition in question.

This requirement on knowledge is a broadly internalist one. Within epistemology, access internalism is the view that for a subject to know a proposition *p*, she has to be in a position to access the knowledge-conferring justification for *p* by a priori reasoning or self-knowledge.[8] To be sure, in this paper we do not attempt a philosophical analysis of the concept of knowledge, nor even the compound concept of mathematical knowledge. Rather, we aim to investigate how certain requirements for knowledge, if construed in a broadly internalist way, may be incompatible with certain foundationalist approaches to mathematical knowledge.

It does seem intuitive, at least at first, to think that the holding of the conditions $\mathbf{C_H}$ (or $\mathbf{C_F}$) is at least part of the knowledge-conferring justification that a stipulation supposedly confers. Arguably, in mathematics the standard for knowledge is proof, especially since the conditions in question here ($\mathbf{C_H}$, $\mathbf{C_F}$) are not always obvious or self-evident. This suggests that Emma and Hero have to be in a position to prove that the conditions for the implicit definitions in question are met in order to know it. Thus our first option.

Something like this may underlie Hilbert's *Grundlagen* (1899). As noted above, he typically supports an axiomatization by showing how to interpret the primitive terms in such a way that the axioms are true. The interpretation is usually in the real numbers. So perhaps we can read Hilbert as taking on the burden of *justifying* his implicit definitions. If so, then perhaps there is such a burden. The present option is to set the burden high: a theorist must be able to prove (or maybe cite a proof) that $\mathbf{C_H}$ or $\mathbf{C_F}$ are met as part what it takes for a successful implicit definition.

However, at least some of the items in the conditions $\mathbf{C_H}$ and $\mathbf{C_F}$ are themselves mathematical. The mathematical notion of satisfiability presupposes a background theory of sets or interpretations, but perhaps satisfiability, in this form, is not part of $\mathbf{C_H}$ or $\mathbf{C_F}$. Still, we saw, or at least assumed, that consistency is among the conditions. Hilbert himself came to understand that deductive consistency is itself a mathematical matter. It invokes a theory of syntax: sentences, definitions, rules, and deductions. How is this meta-theory to be established? Is the ontology of sentences, deductions, and so forth to be introduced via a neo-logicist or Hilbert-style implicit definition? If so, how is *this* implicit definition to be established? Does Emma or Hero have to show that the meta-theory, too, meets $\mathbf{C_H}$ or $\mathbf{C_F}$, respectively? How is it possible for either of them to do that? What is the meta-meta-theory?

In addition to this potential regress, we encounter a vicious circularity with the more basic mathematical theories. Formulas in a rigorous deductive system are, in

[8] For a helpful survey of the debate and different characterisations of internalism and externalism, see Pappas (2005) or Pryor (2001). Our characterization is a simplification of Pappas's characterization of access internalism. The use of other, more sophisticated versions of internalism may complicate the present point, but we must leave that discussion for another occasion.

a sense, structurally similar to natural numbers, and deductive consistency corresponds to an arithmetic property of collections of natural numbers. More formally, arithmetic and the theory of syntax are definitionally equivalent (see, for example, Corcoran et al. (1974)). The structuralist concludes that the theory of syntax just *is* arithmetic, or all but arithmetic, but we need not insist on that much here. Whatever we say about the Hilbertian, the neo-logicist surely is not a structuralist. The point here is that the tight structural connection between numbers and strings indicates that, at least prima facie, the two theories are, in the relevant sense, equally problematic (or unproblematic). Neither theory has an epistemic foundational advantage over the other.

Again, the neo-logicist holds that the natural numbers are, or at least can be, introduced via the stipulation of (HP). The Hilbertian holds that the natural numbers can be introduced by the Dedekind-Peano axioms. If we insist on the present internalist approach—that a subject has to be in a position to show that the conditions $\mathbf{C_H}$ or $\mathbf{C_F}$ are met for the relevant axioms in question—then before we can hold that either of these is a successful implicit definition of the natural numbers, the stipulater, Emma or Hero, has to be in a position to show that the abstraction or axiomatization is consistent. But he or she cannot do that without having on board a theory that is equivalent to—and equi-consistent with—arithmetic.

The later Hilbert program suggested a resolution to this circularity and regress, or it would have if Hilbert were still thinking in terms of implicit definitions in the 1920's and Gödel had not intervened. The idea is that questions of deductive consistency are to be settled in finitary arithmetic, and, crucially, finitary arithmetic is *not* to be understood as introduced via an implicit definition. Rather, it is founded on something like Kantian intuition, or perhaps better, finitary arithmetic simply needs no foundation. It is not that one cannot raise epistemological issues about this branch of knowledge, but there simply is no more secure standpoint from which to resolve such questions. If one somehow comes to doubt finitary arithmetic, it is not clear that she can go on to think about anything at all (Tait 1981). We thus take finitary arithmetic for granted, and use that theory to discharge the burden of justifying other, more powerful and more contentious mathematical theories—full Dedekind-Peano arithmetic, real analysis, set theory, etc.

But, of course, this was not to be. The structural connections between the arithmetic of the natural numbers and syntactical properties of well-formed-formulas, such as consistency, was tightened and exploited in Gödel's celebrated incompleteness theorems, and subsequent work in meta-mathematics. It is not possible to prove the consistency of, say, real analysis in real analysis, much less in a weaker and safer theory, one that does not need a foundation. A fortiori, since consistency is among the conditions on both Hilbert-style and neo-logicist implicit definitions, and provided that the target theory T is sufficiently rich, it is not possible to discharge the burden and establish $\mathbf{C_H}$ or $\mathbf{C_F}$ in T, much less in finitary arithmetic, or some other theory that needs no justification.[9]

[9] As is often noted, we cannot be definitive here, since there is no consensus on what finitary arithmetic is supposed to be. See Detlefsen (1986). We also have not ruled out the possibility that the stipulater can acquire the relevant mathematical knowledge (that the condition $\mathbf{C_F}$ or $\mathbf{C_H}$ is met) by means short of proof.

In sum, it is surely too much to demand that Hero has to be able to *prove* that HP is consistent in order to be able to put it forward as a knowledge-grounding implicit definition, and it is too much to demand that Emma be able to prove that Dedekind-Peano arithmetic is consistent so that she can stipulate its axioms successfully. We take it that, thanks to Gödel, this (admittedly simple-minded) internalist option is a non-starter. We somehow have to lower the bar. The trick is to do so while leave some semblance of a foundationalist program in place.

6.2 If it's good, then it's good

Our second option is the opposite of the first. The thesis to be explored now is that the stipulating theorist has *no* substantial epistemic burden. If a Hilbert-style implicit definition meets $\mathbf{C_H}$, then any stipulation of it is successful. According to this option, if $\mathbf{C_H}$ is in fact met, then the axiomatization defines a structure (or structures), and the stipulater, Emma, can be credited with knowing that the axiomatization is true of this structure (or these structures). Similarly, suppose that Hero puts forward an abstraction principle, like HP. On our second option, the only requirement is that the abstraction principle in fact meets $\mathbf{C_F}$. If it does, then the requisite abstracts have been successfully introduced, and Hero knows that the abstraction principle is true of them.

Our second option thus has the theorist confront Bad Company-type worries and stare them down, at least so far as the epistemological burdens go. On this option, if the relevant conditions $\mathbf{C_H}$ or $\mathbf{C_F}$ are met then the implicit definitions are true and knowledge-grounding. The bad, or failed axiomatizations are rejected just because they do not satisfy the requisite conditions, $\mathbf{C_H}$ or $\mathbf{C_F}$ respectively.

This is not to say that there is no epistemic role for relative consistency proofs and the like. It is not as if Hilbert wasted his time in providing them. In showing that Euclidean geometry can be interpreted in real analysis, the theorist learns that if real analysis is acceptable then so is Euclidean geometry. So she does not have to worry about the latter, unless she is worried about the former. If a question is actually raised about the acceptability of the axiomatization, the theorist has something to say. On the present option, however, the theorist needs no such assurance to push on. She can take the risk, and live dangerously. We have known since Gödel that we have to fly without a safety net. If all is in fact well anyway—if $\mathbf{C_H}$ is met—then the stipulating theorist knows that the stipulated sentences (axioms or abstraction) are true of the defined subject matter. The same goes for the neo-logicist.

Our first option, just rejected, invoked an analogy with internalist epistemology. Here we make a comparison with externalism. According to some accounts, if a belief is produced by a reliable mechanism, such as ordinary vision, then it is justified. The subject need not have an accessible warrant that the mechanism in question is reliable. It is not that there is no point in verifying this reliability (to the extent that this is

Footnote 9 continued
Unless $\mathbf{C_F}$ or $\mathbf{C_H}$ happens to be self-evident—whatever that means—it is not clear how this knowledge is to be acquired.

possible). For example, the subject can consult an optometrist to deal with questions about the reliability of his eyes. The externalist thesis is that the subject need not be able to verify the reliability in order to be justified. Analogously, on our second option here, Emma need not be able to verify that $\mathbf{C_H}$ holds, and Hero need not verify that $\mathbf{C_F}$ holds, in order for each to be successful in their stipulations. By hypothesis, the stipulation in question is successful if and only if $\mathbf{C_H}$ or $\mathbf{C_F}$ holds.

Again, this paper is not an exercise in epistemology, and we do not attempt detailed views on the nature of mathematical knowledge. Rather, we aim to investigate how certain requirements for knowledge, now if they are broadly externally construed, may be incompatible with certain foundational approaches to mathematical knowledge. For simplicity, we use this rather simple version of externalism, and do not supply many details, in order to raise the general concern. As commonly understood, the externalist insists that a subject need not have an internally available warrant that the conditions for knowledge are met in order to be credited with knowledge. Analogously, here we do not insist that our stipulater be in a position to show that the relevant condition, $\mathbf{C_H}$ or $\mathbf{C_F}$, is met, in order to be credited with knowledge of a stipulated implicit definition.

It is often argued, or claimed, that internalist epistemology leads to skepticism: if the bar is set that high, then there simply is no knowledge, or at least no non-trivial knowledge of ordinary things. In the previous sub-section, we have argued that something like this is indeed the case on an internalist construal of the conditions on successful implicit definition. We argue here that the second option sets too low of a standard. It makes mathematical knowledge come too easy.

We cannot be completely definitive here without speculating on the details of $\mathbf{C_H}$ or $\mathbf{C_F}$. Assume, for the time being, that if an axiomatization is satisfiable, then it meets $\mathbf{C_H}$, and assume that if an abstraction principle is satisfiable, then it meets $\mathbf{C_F}$. Alternately, we can assume any axiomatization that is logically equivalent to a good one is itself good.[10] We take it for granted, at least for the sake of this argument, that the Dedekind-Peano axiomatization PA meets $\mathbf{C_H}$, and that Hume's Principle HP meets $\mathbf{C_F}$. Otherwise, there is no point to the indicated programs.

Let Q be any true statement in the language of arithmetic. To make things definite, let Q be the statement of Fermat's last theorem, and suppose that we go back to the period before this result was known. Suppose that Emma, tired of searching for a proof, simply stipulates PA&Q as a Hilbert-style implicit definition. Under the prevailing assumptions, PA&Q meets $\mathbf{C_H}$. Indeed, PA&Q is satisfiable: second-order PA has only one model, up to isomorphism, and by assumption Q is true in that model.[11] Moreover, PA&Q is logically equivalent to PA—it has the same models. So, under the

[10] These assumptions are implicit in much of the literature on the Bad Company objection. The proposed conditions on good abstraction principles—stability, irenicity, and the like—are often stated in model-theoretic terms. If the conditions are understood this way, then they do not distinguish between two abstraction principles that have the same models. Thanks to Roy Cook for pointing this out.

[11] We assume here that the axiomatizations and stipulations are to be understood as second-order with standard semantics. One can get the same results if we think of the comprehension scheme as extending to include formulas in other languages, as in note 6 above. If the axiomatizations or neo-logicist stipulations are interpreted with Henkin semantics (or if they are first-order), then one can make similar points if the example Q is a deductive consequence of PA, although the arguments will be a bit different.

current, simple-minded externalist option, Emma's stipulation is successful: it defines a structure, and she knows that PA&Q is true of that structure.

It is worse. Since she knows the relevant logic, Emma knows that all models of PA&Q are isomorphic to the natural numbers. And she knows that isomorphic structures are equivalent. By using the comprehension scheme and the induction axioms in PA and PA&Q, something like this can be established internally (i.e., without explicit reference to model theory), in ordinary second-order deductive systems. Under the present assumptions, Emma knows that there is a structure defined by PA&Q, and, of course, Q holds in every such structure. So Emma knows that Q holds of the natural numbers, the intended model of PA. What did we need the Wiles' proof for?

Bertrand Russell (1993, 71) once wrote that the "method of 'postulating' what we want has many advantages; they are the same as the advantages of theft over honest toil." Surely, that is the proper conclusion to draw for Hilbert-style implicit definitions, on our second, loose option. There must be *some* epistemic burden on anyone who attempts to put forward a Hilbert-style implicit definition. Demanding that the stipulater provide a proof of consistency is perhaps too much, but not demanding anything of him—beyond the mere fact that the conditions are satisfied—is surely too little.

Essentially the same goes for neo-logicism, on our second, externalist option for Bad Company. To be sure, the neo-logicist need not envisage someone trying to stipulate HP&Q, as an implicit definition, since this formula is not an abstraction principle. Hume's Principle has the proper form (ABS), but HP&Q does not. But we can modify a technique from Richard Heck (1992) to do the trick.

Without loss of generality, assume that the only non-logical terminology in the language of arithmetic are the symbols for zero and the successor function. Let Φ be a sentence in this language that does not contain any occurrences of the variables x, f, and X. Let $\Phi[X, f, x]$ be the result of replacing each occurrence of the symbol for zero with x, replacing each occurrence of the successor symbol with f, and then restricting the quantifiers to X. So, for example, $(\neg\forall y(sy = 0))[X, f, x]$ is $(\neg\forall y(Xy \rightarrow fy = x))$. As above, let Q be any sentence in the language of arithmetic that is true but unknown at the given time. Let Q^* be the sentence:

$$\forall X\forall f\forall x(\text{PA}[X, f, x] \rightarrow Q[X, f, x]).$$

In effect, Q^* says that Q holds in all structures that satisfy PA, but without explicit mention of models and the like. Given the categoricity of PA, this amounts to a statement that Q holds in all structures isomorphic to the natural numbers. Notice that Q^* has no non-logical terminology. Moreover, Q^* is logically true if and only if Q is true of the natural numbers (Shapiro 1991, Chap. 4, Corollary 4.9).

Now suppose that Hero attempts to stipulate the following abstraction principle:

$$(\text{HP}+Q): \quad \forall F\forall G(\Pi(F) = \Pi(G) \equiv [F \approx G \;\&\; (\neg Q^* \rightarrow \forall x(Fx \equiv Gx))]),$$

where '$F \approx G$' is the usual statement that the F's are equinumerous with the G's. Notice that like (HP) and Basic Law V, (HP+Q) is a logical abstraction, in that it has no non-logical terminology on its right-hand side. It says that two concepts have the same abstract—the same Π—just in case they are equinumerous and, if $\neg Q^*$, they are co-extensive.

By hypothesis, Q is true of the natural numbers. It follows that (HP+Q) has the same models as (HP). Indeed, suppose that (HP) is true in a given structure (or in the universe). Then, by Frege's theorem, there is a sub-structure that satisfies PA, namely the structure of the defined cardinal numbers. By hypothesis, Q is true in that structure, and so $Q*$ is also true. It follows that (HP+Q) is true if we interpret $\Pi(F)$ as the cardinal number of F. Conversely, suppose that (HP+Q) is true in a given structure (or in the universe). It follows, by an argument similar to that leading to Russell's paradox, that $\neg Q*$ is false in the structure (or in the universe). So $\forall F \forall G(\Pi(F) = \Pi(G) \equiv F \approx G)$ is true in the structure (or in the universe). It thus easily follows that (HP) is satisfiable in that same structure (or in the universe): just interpret the cardinal number operator as Π. In other words, (HP) and (HP+Q) have the same models.

Under the prevailing assumptions, then, (HP+Q) satisfies $\mathbf{C_F}$—assuming, of course, that (HP) does. So according to our present externalist option, Hero *successfully* stipulates (HP+Q), and thereby comes to know that (HP+Q) is true of the introduced abstracts—the Π's. But Hero can then also deduce, via an argument much like Frege's theorem (combined with techniques associated with Dedekind) that the finite Π's are isomorphic to the finite cardinal numbers introduced by (HP), and he knows that isomorphic structures are equivalent. Since Hero also knows $Q*$, he thereby knows that Q holds of the natural numbers. Thus, as with Hilbert-style implicit definitions, the neo-logicist bypasses all the hard and ingenious work that Wiles did in establishing Fermat's last theorem. Here, too, we have theft over honest toil.

Our second, externalist option did not start life as the thesis that every arithmetic truth is known, or can easily become known, nor as the thesis that every true belief is knowledge. The present option is just a loose thesis about the requirements on successful stipulation, either Hilbert-style or neo-logicist. But the option leads to an absurd view much like this. Mathematical knowledge comes too easy. In particular, the key role of proofs as extending knowledge is undermined since we can always conjoin a true belief to the axioms and take this new statement to be known.[12]

This worry could be met if one could give an account of which propositions are eligible to be axioms and which should be (non-trivial) theorems—assuming that these do not substantially overlap. Our challenge is that on the Hilbert account no such distinction can be drawn—at least not on the basis of the *form* of the implicit definition. Also, due to Heck's technique, no such distinction can be drawn in the case of abstraction principles and least not on the basis of the form of an abstraction principle.

[12] To continue the analogy with externalism, an advocate of this loose option might try to make the view more palatable by insisting on a distinction between Hero or Emma knowing an implicit definition p and Hero or Emma being able to *responsibly claim* knowledge of p. Or perhaps our advocate will concede that, say, Emma does know PA+Q and that Hero knows HP+Q, but that neither knows that he or she knows the statement in question. It seems to us that the conclusions are still bad enough; the view is still plenty unpalatable. But the underlying issues here would take us too far afield.

6.3 Presumption of innocence

One way to back off, slightly, from the pure externalist position would be to declare that a proposed Hilbert style implicit definition or neo-logicist abstraction is acceptable, and known to be true, provided only that (1) the relevant conditions, $\mathbf{C_H}$ or $\mathbf{C_F}$, hold and that (2) the theorist sees that he or she has no good reason to think otherwise. Thus, for example, Hero knows that Hume's Principle is true and Emma knows that the axiomatization of arithmetic is true because there is no particular reason to think that the conditions are not met. The only burden on them is to make sure that the usual arguments to paradox are blocked, and that nothing else fishy is going on. The idea here is that the *default* position is that an implicit definition is good.[13] This sketch of the third way will suffice to highlight our concerns with it.

Once again, let PA be the conjunction of the second-order Dedekind-Peano axioms, and let Q be a true but presently unknown statement in the language of arithmetic. Suppose that Emma attempts to stipulate PA&Q as a Hilbert-style implicit definition. On the present option, she should be skeptical. She can see that PA&Q is satisfiable if and only if Q is true of the natural numbers, and, by hypothesis, she has no idea whether that is so or not. So she sees that something is fishy, and thus has to retract the stipulation. It is not good after all, at least not yet.

Similarly, just above, we had Hero propose the following abstraction principle:

$$(\text{HP+}Q): \quad \forall F \forall G(\Pi(F) = \Pi(G) \equiv [F \approx G \,\&\, (\neg Q^* \rightarrow \forall x(Fx \equiv Gx))]),$$

Being a competent logician, who is familiar with Russell's paradox, Hero should be able to see that (HP+Q) is satisfiable if and only if Q is true of the natural numbers. And, again, he has no idea whether that is so. So something is fishy here, too, and the default position is defeated.[14]

On both of these counts, however, the conclusion that something is fishy and thus that the stipulation is defeated turns on an artifact of the way the stipulations were set up in the last sub-section. Suppose Emma stumbles onto a more complex axiomatization E that, as it happens, is (deductively) equivalent to PA+Q, but she has no idea of this equivalence,[15] and sees no reason to think that the axiomatization might fail to meet the condition $\mathbf{C_H}$. So the default is maintained, and the axiomatization E is good. Emma knows that it is true. Suppose that the mathematical community at large accepts this. Similarly, Hero can stumble onto a complex abstraction H that, as it happens, is equivalent to HP+Q, but does not have the straightforward

---

13 Agustín Rayo (2003) explores and criticizes this option for other reasons than we do here. He presents a holistic resolution which is, in some respects, similar to the account developed in the closing section of this paper.

14 Wright's (1997) second conservativeness requirement on good abstractions is that if a conclusion Φ is derived via disjunctive syllogism from a known paradox, then Φ has to be shown to be true independently of that demonstration. That is exactly the problem here: Q is not known independently of the derivation from the disjunctive abstraction principle.

15 Of course, if Q is itself a deductive consequence of PA, then PA itself is such an axiomatization. We have in mind here an axiomatization that is significantly different from PA and from PA&Q, so that the deduction of Q from the axiomatization strikes one as theft over honest toil. Admittedly, we have no way of making this precise.

disjunctive form, and there is no reason to think that Q or Basic Law V plays any role in it. Then, according to the proposal of this section, the stipulation of H is good, by default. Again, suppose the mathematical community at large accepts it, and starts working in this field, publishing articles about it in major journals, and the like.

Suppose that some time later, a brilliant mathematician discovers how to model the natural numbers in the ontology of E or in the abstracts generated by H. A bit later, someone else derives Q from E and/or H. They are then in position to conclude that Q is true of the natural numbers. By the plan of this section—presumption of innocence—Q is now known to be a truth of arithmetic. We take this scenario as a *reductio ad absurdum* of the intermediate proposal at hand. It is not as if the community has done no honest toil, as happened with pure externalism, but the toil is not of the right kind to establish a truth of arithmetic.

It is possible, we suppose, that the community would regard, and perhaps should regard, these results as belated discoveries that something fishy is going on, after all. That is, the discovery that the new theory yields a "proof" that Q holds in the natural numbers amounts to a defeater for the stipulation.[16] Perhaps, but the situation could also be taken as a proof of Q. It might depend on how confident they are in the new theory, and on the role it takes in the community.

Part of the problem here is that the notion of "nothing fishy is going on" is too obscure to characterize a serious position, at least as it stands. Surely, it is the burden of the proponent of this general idea to clarify the conditions for when the subject can assume that "nothing fishy is going on" and so assume the default success of an implicit definition. We turn to one such proposal.

6.4 Entitlements: a recent proposal

In 2004, Wright is mainly concerned with external-world skepticism, and he uses a specific type of warrant, so-called "entitlements", to respond to such skepticism. He notes that the same idea can also be used for the project of acquiring mathematical or logical knowledge, and thus it may help with the present problem.

Wright's program begins with the observation that in pursuing any type of cognitive project—perceiving the world, learning about the past, acquiring mathematical knowledge, etc.—a subject needs to make specific presuppositions. For example, when someone undertakes a project of finding a phone number in the directory, he needs to take the reliability of his senses for granted. In this case, the presupposition can be challenged, and our subject can get his eyes tested in response to such a challenge. However, this further investigation would in turn also raise its own presuppositions which are of a similar type. The optometrist can get her equipment checked, and have her own eyes checked, but these would involve further presuppositions. In practice the regress ends. It has to. Wright's (2004, 189) point is that "whenever any cognitive achievement takes place, it does so in a context of specific presuppositions, which are not themselves an expression of any cognitive achievement to date".

[16] A referee suggested this outcome.

As a result, there is an element of adventure in any cognitive achievement. It is not just mathematics that has to fly without a safety net. In any cognitive project, there will be certain presuppositions that have not been subject to a cognitive investigation and have to be assumed on trust, i.e. without evidential justification. Wright introduces the notion of entitlement to capture the idea that at least in many cases, it is not irrational or unwarranted to invest trust in a presupposition. In typical cases, the trust is rational. He writes:

> Let me try to harness these ideas to a definite proposal about entitlement. First (to tidy up a bit) a definition: let us say that P is a presupposition of a particular cognitive project if to doubt P (in advance) would rationally commit one to doubting the significance or competence of the project.
>
> Then the relevant kind of entitlement—an entitlement of cognitive project—may be proposed to be any presupposition of a cognitive project meeting the following additional two conditions:
>
> i. We have no sufficient reason to believe that P is untrue and
> ii. The attempt to justify P would involve further presuppositions in turn of no more secure a prior standing, ..., and so on without limit; so that someone pursuing the relevant enquiry who accepted that there is nevertheless an onus to justify P would implicitly undertake a commitment to an infinite regress of justificatory projects, each concerned to vindicate the presupposition of its predecessor (Wright 2004, 190).

So, the idea is that whenever we need to carry out a specific cognitive project whose presuppositions fulfil the first and second requirement stated, then "we should—are rationally entitled to—just go ahead and trust that the [presuppositions] are met" Wright (2004, 190–191). This idea resonates, in parts, the underlying thought of the presumption of innocence, that one can take the success of a definition as a default provided that there seems nothing wrong with it, i.e., that it fulfills condition (i). What is added here is that any attempt to justify the presupposition in question would lead to a hopeless regress.

Let us now consider if and how this proposal could apply to the cognitive project of accumulating mathematical knowledge by means of Hilbert-style or neo-logicist stipulations. Clearly, it is a presupposition of this cognitive project that the conditions $\mathbf{C_H}$ or $\mathbf{C_F}$ are met. As noted before, consistency is among both conditions. Indeed, to doubt the consistency of Hume's Principle, for example, would rationally commit someone to doubting the significance of the entire neo-logicist cognitive project (*pace* dialetheism), or at least this instance of it. And clearly, to doubt the consistency of the Dedekind-Peano axioms would call into question the project of stipulating it (again, dialetheism notwithstanding).

Does this particular presupposition—the consistency of Hume's Principle and the Dedekind-Peano axioms—fulfil Wright's two conditions to qualify as an entitlement? Consider the second condition first, and let us focus on Hume's Principle. The issue is whether an attempt to justify the claim that this abstraction is consistent would involve further presuppositions which are of no more secure a prior standing.

In light of the incompleteness theorem, it seems that the only foreseeable means to show that Hume's Principle is consistent is to provide a relative consistency proof—relative to another mathematical system, or relative to the soundness of the system used in the proof. This system again incurs the same presupposition, namely that it is consistent. As a result, it is reasonable to suppose that an enquiry to justify this presupposition would itself "involve further presuppositions in turn of no more secure a prior standing, . . . , and so on without limit". And thereby one would "implicitly undertake a commitment to an infinite regress of justificatory projects, each concerned to vindicate the presupposition of its predecessor." So, it seems, Wright's second condition can be regarded as fulfilled.

The first condition is trickier. It aims to capture in more sophisticated philosophical terminology the idea, noted earlier, that "nothing fishy is going on". Yet, it seems to us the first conditions is still unclear in important respects. First, "having no (sufficient) reason" should not—at least on our understanding—be understood "externally" in the sense of "there being no (sufficient) reason". To explain, if Hume's Principle is in fact consistent, there would be no (sufficient) reason to doubt it—in the "realm of reason" as it were. However, this should not automatically mean that a subject *has* no (sufficient) reason to doubt it. Otherwise we are back on the externalist-type lines of two sub-sections back, and condition (i) does not do much work in relation to the thinking subject.

So we should look for an understanding of "having no sufficient reason" along more internalist lines. Once again, demanding a proof that all is well is asking for too much. Even so, the thinker, Hero or Emma, has a genuine epistemic responsibility, and the condition is relative to the information-state of the subject. Take as a proposal the following:

> Once all available evidence (in a very broad sense) which might show that a presupposition P is untrue has been surveyed, and it is established that it does not undermine P, then the subject has no sufficient reason to believe P is untrue.

In the spirit of Wright's proposal, the key idea here is that the notion of "available evidence" is confined to what can be reasonably expected from an epistemic subject.

Let us return to the consistency of Hume's Principle as a presupposition and see how it complies to condition (i) for entitlements, as articulated thus far. Do we have sufficient reason to believe this presupposition to be untrue? Well, there is the Bad Company objection. In particular, there are inconsistent principles looking rather similar to Hume's Principle. So, in order to nullify any sufficient reason to believe that Hume's Principle is inconsistent, we need to look at all the available evidence which would defeat the presupposition. This means we need to look at whether known types of paradoxes can be developed from Hume's Principle. Once we have surveyed all the "usual suspects", Russell's and Burali-Forti's constructions for example, and all other evidence we can think of, which might show that the presupposition of consistency is not met, and we see that Hume's Principle is not subject to any of them, then we can consider the first condition to be fulfilled. In these circumstances there is "no sufficient reason to believe Hume's Principle is consistent to be untrue". Although this is a mere sketch, it does promise to be an alternative to the pure and strict internalist and the pure and loose externalist responses rejected above.

Nevertheless, we still run into variants of the problems highlighted in the previous sub-section. To continue our main example, suppose that Hero stumbles across a complex abstraction principle A that is deductively equivalent to HP+Q but he has no idea of this equivalence and he sees no reason to believe that any of the known paradoxes might apply to the principles. Any attempt to generate a Russell or Burali-Forti paradox on the defined abstracts fails. So, on the present proposal, Hero can be considered to have no sufficient reason to believe the principle to be untrue and thereby (assuming condition (ii) is fulfilled) acquire an entitlement for the statement in question. This yields cheap knowledge that Q is true of the natural numbers.

This type of ignorance about the content of foundational principles is ubiquitous within mathematical practice. Connections between axiomatizations are discovered all the time. If the entitlement approach is to apply to these cases, then it is open to the problem that far too many statements that intuitively seem ineligible to be an entitlement, say because they are complex or not sufficiently foundational, turn out to be so.

There is an additional worry that with the entitlement approach, the more "ignorant" a subject is, the easier it will be to satisfy the above condition. That is, an uneducated stipulater will not know of many defeaters and so it will be easier for this subject to establish that there is no sufficient reason to believe P is untrue. The fewer "usual subjects" the subject—or her community—is aware of, the less work she has to do to maintain the entitlement.

These are some of the many issues that the entitlement theorist has to resolve to make his view viable. Here, we mainly note that many problems seem to affect such a position that attempts to locate itself somewhere between a pure externalist and a pure internalist approach. This is not to say that it fails rather that more work needs to be done to consider it a genuine alternative.[17]

7 Holism: an unFregean move

There is another general problem in the vicinity of the entitlement, or presumption of innocence proposals, no matter how they are articulated further. In effect, Wright's proposal adopts provisos of some kind concerning potential defeaters. If conditions (i) and (ii) are met concerning certain proposed foundational stipulations, then the stipulater can trust that the stipulations are successful and thus true. This invites our running question concerning the epistemic status of the provisos, a sort of ugliness issue one level up. Does the stipulating subject have to *know* that conditions (i) and (ii) are met, or does he have to have some reason to think that they are met, or does it just have to be the case that conditions are met? The latter option, leads us back into the externalist direction broached above, and fails to be suitable for the foundationalist approach, as above. The former options raise important questions about what exactly the subject has to know about the provisos, and so a further epicycle is

[17] For further critical discussion of the entitlement approach applied to mathematics, see Ebert (ms), Pedersen (2005) and Pedersen (ms).

about to be started. Is the subject *entitled* to the thesis that there are no defeaters, and that any attempt to investigate the consistency, say, of the stipulation would lead to a regress?

Consider the second condition, that the "attempt to justify" the presuppositions "would involve further presuppositions in turn of no more secure a prior standing, ..., and so on without limit". We noted above this condition is plausibly met in the cases of Hume's Principle and PA. Presumably, the subject has to *know*, or have good reason to believe, *that* any attempt to certify the consistency of Hume's Principle, for example, would involve a presupposition no more secure. It seems to us that the only way to show this would be to give a relative consistency proof.

In the original articulation of Scottish neo-logicism, Wright (1983) conjectured that HP (which was then called $N^{=}$) is consistent. Boolos (1987) showed that HP is equi-consistent with second order Peano arithmetic, the system we have been calling PA (Parsons 1964; Hodes 1984). Since no one (or hardly anyone) seriously doubts the consistency of second-order Peano arithmetic, no one (or hardly anyone) seriously doubts the consistency of HP. Boolos's result provided some sort of legitimization for Scottish neo-logicism, and it made the program more respectable, or so it seems.

However, we are puzzled as the point, or value, of the equi-consistency result from the perspective of the neo-logicist himself. How does the equi-consistency bear on what Hero knows, or what he can take himself to know? Recall that the thesis of neo-logicism is that Hero acquires knowledge of arithmetic by stipulating HP. From this epistemological perspective, one would think that there should be no epistemological payoff in providing relative consistency proofs. It is not just that such proofs cannot offer any kind of guarantee that HP is in good order. That much is not on the agenda anyway—thanks to Gödel. Our question here is how the relative consistency proofs can even help—how they can assure Hero that Wright's condition (ii) is met. From the foundational perspective, an appeal to another mathematical theory is simply not available. And it would seem disingenuous to appeal to another axiomatization of *arithmetic*.

In the previous section, we claimed that Wright's condition (ii) is met for HP and PA. The argument appealed to the second incompleteness theorem, which is itself a piece of mathematics. In effect, the incompleteness theorems are proved in formal systems at least as strong as arithmetic. It is hard to see how Hero can appeal to them, until arithmetic is established.

In the end, this problem may or may not be resolved to the satisfaction of the neo-logicist.[18] To close this article, we briefly propose another reaction to ugliness issues like this. It is a perspective on implicit definitions and stipulations that is at odds with the foundationalist aspects of neo-logicism.

According to Michael Hallett (1990), Hilbert's view, at the time of the publication of his *Grundlagen* and the correspondence with Frege, was that if an axiomatization accords with mathematics *as developed thus far*, then one can conclude that the axiomatization is true, and its "objects" exist—in the only sense of existence relevant

[18] Hale and Wright (2001, Introduction) provide some discussion of the role of meta-theory in neo-logicism.

to mathematics. Of course, the conclusion is defeasible, but that is good enough. Translated into the present context, the idea is that all the Hilbertian requires is that the "new" theory introduced by the axiomatization should fit smoothly into existing mathematical practice.

According to Hallett, Hilbert proposed "consistency" as a mathematically serviceable gloss on this notion of "smooth fit", but he never took consistency to be a wholesale replacement of the intuitive criterion for acceptability, namely that the axiomatization fit into existing practice.

Proof-theoretic consistency is mathematically tractable, and, of course, a most productive research program emerged. A better gloss on the relevant notion of acceptability would be *satisfiability*. We shore up an axiomatization by interpreting it in an established mathematical theory. This, in fact, is the practice of Hilbert's *Grundlagen*. As noted, after presenting an axiomatization, he shows how to interpret it, usually in the real numbers. Presumably, there is no serious doubt about the real numbers.

On occasion, existence issues come up in mathematical practice itself. Since the nineteenth century, at least, these questions are usually resolved in the way described here, by interpreting the questioned theory in an established one. Non-Euclidean geometries were shorn up by showing how to interpret them in Euclidean space.

Mark Wilson (1993, Sect. III) illustrates the historical development and acceptance of a space-time with an "affine" structure on the temporal slices:

> . . . the acceptance of . . . non-traditional structures poses a delicate problem for philosophy of mathematics, viz., how can the novel structures be brought under the umbrella of *safe* mathematics? Certainly, we rightly feel, after sufficient doodles have been deposited on coffee shop napkins, that we understand the intended structure . . . But it is hard to find a fully satisfactory way that permits a smooth integration of non-standard structures into mathematics . . . We would hope that "any coherent structure we can dream up is worthy of mathematical study . . ." The rub comes when we try to determine whether a proposed structure is "coherent" or not. Raw "intuition" cannot always be trusted; even the great Riemann accepted structures as coherent that later turned out to be impossible. *Existence principles* beyond "it seems okay to me" are needed to decide whether a proposed novel structure is genuinely coherent . . . [L]ate nineteenth century mathematicians recognized that . . . existence principles . . . need to piggyback eventually upon some accepted range of more traditional mathematical structure, such as the ontological frames of arithmetic or Euclidean geometry. In . . . our century, set theory has become the canonical backdrop to which questions of structural existence are referred. (pp. 208–209)

Within the community of professional mathematicians, if not philosophers, a set-theoretic proof of satisfiability resolves any legitimate questions of existence. In effect, the iterative hierarchy of set theory has become the canonical proving ground for existence questions.

To sum up, and return to our theme, the holistic proposal is that the only epistemic burden on Emma is to provide a satisfiability proof of a proposed axiomatization, along the lines of those in Hilbert's *Grundlagen*. Of course, this approach to the Hilbertian version of Bad Company is not consistent with the foundational goals of the later

Hilbert program. The procedure of modeling proposed axiomatizations in the iterative hierarchy cannot assure us that we won't be "driven from Cantor's paradise", to paraphrase Hilbert (1925), should set theory be inconsistent or otherwise unsuitable as a mathematical theory. In a sense, we are *relying* on Cantor's paradise, or at least Zermelo's paradise, to justify our new axiomatizations. But, once again, the mathematical community flies without a safety net—certainly without a net safer than arithmetic, analysis, and set theory.

There are different ways to extend this holistic proposal to the Scottish neo-logicist project. One is to adopt exactly the same criterion of existence: an abstraction principle is good once Hero shows it to be satisfiable, and he is free to use any established mathematical theory in interpreting an abstraction principle. So Hero is right to be content with Boolos's demonstration that (HP) is satisfiable on any countably infinite domain.

This first maneuver is to treat abstraction principles more along the lines of Hilbert-style implicit definitions, as in Sect. 2 above, and it does not resolve the standard Bad Company issues for neo-logicism. The parity principle and the nuisance principle are both satisfiable, on any finite domain, but both are inconsistent with (HP). Moreover, this plan misses one of the main points of (HP), namely to serve as a theory of cardinal number in all of its applications. To fulfil this role, the quantifiers have to be unrestricted.

A better way to incorporate the holistic theme into neo-logicism is to require that Hero show—in set theory presumably—that his proposed abstraction principle is satisfiable on any sufficiently large domain. This has become known as a requirement of *stability*.[19] Then we know that whatever the universe may be, the abstraction principle can be made true on it. The parity principle and the nuisance principle fail this test, of course, since they are not satisfiable on any infinite domain. If we assume the axiom of choice, then Hume's Principle passes the test, since Boolos's technique shows that (HP) can be satisfied on any infinite, well-ordered domain. The abstraction principles introduced for real analysis in Hale (2000) and Shapiro (2000) pass the test as well (mostly because their quantifiers are restricted).

A third way to incorporate the holistic theme is just to require that Hero show that his abstraction principle can be interpreted in the iterative hierarchy, or perhaps the iterative hierarchy with urelements. The idea is that quantifiers in the abstraction principle should be interpreted as ranging over all of V, or all of the set-theoretic universe. It is widely held that any legitimate mathematical theory can be interpreted on the iterative hierarchy. If this is so, then the present test would give abstraction principles the required generality. If we assume a global choice principle, then (HP) passes this test as well, along Boolos's lines.

All of these holistic proposals have Hero make use of other mathematical theories, usually set theory, in order to shore up abstraction principles. In Shapiro (2004, 2005), this is called an "external" perspective. Clearly, the program is incompatible with the epistemological goals of Frege's logicism. The point of that, recall, was to show that

[19] An abstraction principle is said to be *irenic* if it is consistent with every conservative abstraction principle. Weir (2003) shows that if we interpret the conditions model-theoretically, in terms of satisfiability, an abstraction principle is stable if and only if it is irenic.

the basic principles of arithmetic and real analysis are analytic: they can be derived from basic principles of logic, plus definitions. By relying on an external set theory, we compromise that goal: the logicist would have to show that set theory itself is free from empirical elements, Kantian intuition, etc. How could we do that?

Neo-logicism has a similar foundational aim. One stated purpose is to show that branches of mathematics can be known a priori, without relying on Kantian intuition, empirical considerations (like indispensability in science), a direct grasp of mathematical objects, and the like. Wright (1997, 210–211) wrote:

> Frege's theorem will . . . ensure . . . that the fundamental laws of arithmetic can be derived within a system of second-order logic augmented by a principle whose role is to *explain*, if not exactly to define, the general notion of identity of cardinal number, and that this explanation proceeds in terms of a notion which can be defined in terms of second-order logic. If such an explanatory principle . . . can be regarded as *analytic*, then that should suffice . . . to demonstrate the analyticity of arithmetic. Even if that term is found troubling, . . . it will remain that Hume's Principle—like any principle serving implicitly to define a certain concept—will be available without significant epistemological presupposition . . . So one clear a priori route into a recognition of the truth of . . . the fundamental laws of arithmetic . . . will have been made out . . . Such an epistemological route . . . would be an outcome still worth describing as logicism . . .

Clearly, this foundational goal is undermined if we adopt the present holistic response to our issue. Hero would have to rely on a derivation in set theory, or some other established mathematical theory, before he can know that his stipulation is good—before he knows that it is true. If the background theory, typically set theory, itself has "significant epistemological presupposition", then so does the neo-logicist's arithmetic.

Acknowledgements We are indebted to discussion and written comments from many friends and colleagues. The list includes Roy Cook, Øystein Linnebo, Paul McCallion, Nikolaj Pedersen, Marcus Rossberg, William Stirton, Gabriel Uzquiano, Crispin Wright, Elia Zardini, and two anonymous referees for this special issue.

References

Boolos, G. (1987). The consistency of Frege's *Foundations of arithmetic*. In J. J. Thompson (Ed.), *On being and saying: Essays for Richard Cartwright* (pp. 3–20). Cambridge, Massachusetts: The MIT Press, reprinted in Boolos (1998), 183–201, and Demopoulos (1995), 211–233.

Boolos, G. (1997). Is Hume's principle analytic? In R. Heck, Jr. (Ed.), *Language, thought, and logic* (pp. 245–261). Oxford: Oxford University Press, reprinted in Boolos (1998), 301–314.

Boolos, G. (1998). *Logic, logic, and logic*. Cambridge: Harvard University Press.

Coffa, A. (1991). *The semantic tradition from Kant to Carnap*. Cambridge: Cambridge University Press.

Corcoran, J., Frank, W., & Maloney, M. (1974). String theory. *Journal of Symbolic Logic, 39*, 625–637.

Demopoulos, W. (1995). *Frege's philosophy of mathematics*. Cambridge: Harvard University Press.

Detlefsen, M. (1986). *Hilbert's program*. Dordrecht: D. Reidel Publishing Company.

Dummett, M. (1991). *Frege: Philosophy of mathematics*. Cambridge: Harvard University Press.

Dummett, M. (1998). Neo-Fregeans: In Bad Company? in Schirn (1998), 369–387.

Ebert, P. (ms.), What foundational knowledge could not be.

Frege, G. (1976). David Hilbert. In G. Gabriel, H. Hermes, F. Kambartel, & C. Thiel (Eds.), *Wissenschaftlicher Briefwechsel* (Chapter XV). Hamburg: Felix Meiner.
Frege, G. (1980). *Philosophical and mathematical correspondence*. Oxford: Basil Blackwell.
Hale, B. (2000). Reals by abstraction. *Philosophia Mathematica, 3*(8), 100–123; reprinted in Hale and Wright (2001), 399–420.
Hale, B., & Wright, C. (2001). *The reason's proper study*. Oxford: Oxford University Press.
Hallett, M. (1990). Physicalism, reductionism and Hilbert. In A. D. Irvine (Ed.), *Physicalism in mathematics* (pp. 183–257). Dordrecht: Kluwer Academic Publishers.
Heck, R. (1992). On the consistency of second-order contextual definitions. *Nous, 26*, 491–494.
Hilbert, D. (1899). *Grundlagen der Geometrie*, Leipzig: Teubner. English edition: (1959). *Foundations of geometry* (Trans.: Townsend, E.). La Salle, Illinois: Open Court.
Hilbert, D. (1925). Über das Unendliche. *Mathematische Annalen, 95*, 161–190; translated as On the infinite. van Heijenoort (1967), 369–392.
Hilbert, D. (1935). *Gesammelte Abhandlungen, Dritter Band*. Berlin: Julius Springer.
Hodes, H. (1984). Logicism and the ontological commitments of arithmetic. *Journal of Philosophy, 81*, 123–149.
Linnebo, Ø. (2004). Predicative fragments of Frege arithmetic. *Bulletin of Symbolic Logic, 10*, 153–174.
Pappas, G. (2005). Internalist vs. externalist conceptions of epistemic justification. In E. N. Zalta (Ed.), *The Stanford Internet encyclopedia of philosophy*. http://plato.stanford.edu/archives/spr2005/entries/justep-intext/
Parsons, C. (1964). Frege's theory of number. In M. Black (Ed.), *Philosophy in America* (pp. 180–203). Ithaca, New York: Cornell University Press; also in Parsons, C. (1983). *Mathematics in philosophy* (pp. 150–175). Ithaca, New York: Cornell University Press.
Pedersen, N. (2005). *Entitlement in mathematics*. Ph.D. Thesis, University of St. Andrews.
Pedersen, N. (ms), Mathematical skepticism and entitlement.
Priest, G. (1987). *In contradiction: A study of the transconsistent*. Dordrecht: Martinus Nijhoff Publishers.
Pryor, J. (2001). Highlights of Recent Epistemology. *British Journal for the Philosophy of Science, 52*, 95–124.
Rayo, A. (2003). Success by default. *Philosophia Mathematica, 3*(11), 305–322.
Rayo, A., & Uzquiano, G. (Eds.) (2006). *Absolute generality*. Oxford: Oxford University Press.
Russell, B. (1993). *Introduction to mathematical philosophy*. New York: Dover (first published in 1919).
Schirn, M. (Ed.) (1998). *The philosophy of mathematics today*. Oxford: Oxford University Press.
Shapiro, S. (1987). Principles of reflection and second-order logic. *Journal of Philosophical Logic, 16*, 309–333.
Shapiro, S. (1991). *Foundations without foundationalism: A case for second-order logic*. Oxford: Oxford University Press.
Shapiro, S. (1997). *Philosophy of mathematics: Structure and ontology*. New York: Oxford University Press.
Shapiro, S. (2004). The nature and limits of abstraction. *Philosophical Quarterly, 54*, 166–174.
Shapiro, S. (2005). Sets and abstracts. *Philosophical Studies, 122*, 315–332.
Stein, H. (1988). *Logos*, logic, and *Logistiké*: Some philosophical remarks on the nineteenth century transformation of mathematics. In W. Aspray & P. Kitcher (Eds.), *History and philosophy of modern mathematics* (Vol. XI, pp. 238–259). Minneapolis, Minnesota Studies in the Philosophy of Science: University of Minnesota Press.
Tait, W. (1981). Finitism. *The Journal of Philosophy, 78*, 524–546.
Tennant, N. (1987). *Anti-realism and logic*, Oxford: Oxford University Press.
Van Heijenoort, J. (1967). *From Frege to Gödel*. Cambridge: Harvard University Press.
Weir, A. (1998). Naïve set theory is innocent. *Mind, 107*, 763–798.
Weir, A. (2003). Neo-Fregeanism: an embarrassment of riches. *Notre Dame Journal of Formal Logic, 44*, 13–48.
Wilson, M. (1993). There's a hole and a bucket, dear Leibniz. *Midwest Studies in Philosophy, 18*, 202–241.
Wright, C. (1983). *Frege's conception of numbers as objects*. Aberdeen: Aberdeen University Press.
Wright, C. (1997). On the philosophical significance of Frege's theorem. In R. Heck, Jr. (Ed.), *Language, thought, and logic* (pp. 201–244). Oxford: Oxford University Press; reprinted in Hale and Wright (2001), 272–306.
Wright, C. (1998). On the harmless impredicativity of $N^{=}$ (Hume's Principle). In Schirn (1998), 339–368; reprinted in Hale and Wright (2001), 229–255.

Wright, C. (2000). Neo-Fregean foundations for real analysis: Some reflections on Frege's constraint. *Notre Dame Journal of Formal Logic, 41*, 317–334.

Wright, C. (2004). Warrant for nothing (and foundations for free)? *Proceedings of the Aristotelian Society Supplementary, 78*, 167–212.