Neo-Logicism and A Priori Arithmetic

MPhil. Stud. Thesis

Tom Eckersley-Waites

Contents

Neo-Logicism and A Priori Arithmetic

Introduction

§1: The Development of Neo-Logicism

Frege's philosophy of arithmetic had at its core two central philosophical themes. The first is that arithmetical truths are truths about independently-existing objects. The second is that arithmetical truths are provable on the basis of logic and suitable definitions. Thus Frege subscribed both to a form of *platonism* and a form of *logicism* about parts of mathematics, of which arithmetic will be our concern here. In order to establish these philosophical results, Frege required that a technical task be completed first. Such a task comprised the development of a logically perfect language – his *Begriffsschrift* – in which one could interpret the truths of arithmetic. Thus Frege both invented a system of logic, and interpreted arithmetic in that logic.

On a modern understanding of logic as the investigation of what holds in any domain whatever, it becomes clear very quickly that these two core theses do not sit easily together. If truths of arithmetic are to be truths of logic and arithmetic is about particular objects, then logic can no longer be viewed in such a way. However, it is when the *epistemological* character of Frege's logicism is brought into focus that the apparent tension is resolved. We must hold apart the semantic, metaphysical and epistemological to put Frege's project in its proper place.

Firstly, there is the semantic thesis – what Wright calls "number-theoretic realism" – that holds that we should take "the Peano axioms... and their logical consequences, however that class is to be characterised, as being truths of some sort"[1] and that they are truths independently of whether or not

1 Wright (1983), p.xiv

they are ever formulated or in principle verifiable. Secondly there is the metaphysical thesis – Frege's platonism – that these mathematical statements are true or false in virtue of particular mathematical objects being some way or other. Thirdly there is Frege's epistemological logicism; that pure logic can provide a basis for our understanding and knowledge of mathematics. Wright stresses this by pointing out that "Frege's *question* is surely a good one: what [could be] the ultimate... source of our knowledge of number-theoretical statements?"[2] It is to what extent and why Frege and Wright's answers to this question are defective that will be examined in this thesis.

Frege most clearly expresses the centrality of this question in §62 of his *Grundlagen der Arithmetik* (1884) when he has finished giving his positive argument for platonism. Given his repudiation of the competing psychological theories of the metaphysics and epistemology of arithmetic, he asks "how, then, should a number be given to us, if we can have no idea or intuition of it?"[3]. His answer appeals to one of his guiding principles: his famous context 'Context Principle', the dictum that one should "never... ask for the meaning of a word in isolation, but only in the context of a proposition"[4]. This principle both legitimises and enforces the use of some kind of contextual definition of number if his logicism is to be vindicated. The first definition that Frege put up for evaluation was what became known as Hume's Principle:

(HP): $\forall F \forall G\ [\#F = \#G \leftrightarrow (F \approx G)]$ for $\approx$:= "can be put into 1-1 correspondence with"

However, he found it wanting not for mathematical but for philosophical reasons. He saw such definitions as insufficient to decide all identity statements of the form '#F = x' when x is not an expression of the form '#G' – for example, if x stood for Julius Caesar. He saw such "nonsensical examples" as a useful heuristic device to demonstrate the inadequacy of such a definition. He writes that "Naturally no-one is ever going to confuse [Julius Caesar] with [the number 3], but this is no

2 Ibid., p.xxi
3 Frege (1884), §62
4 Ibid., Introduction

thanks to our definition."[5] This lack of discriminatory power of the definition was for Frege an intractable difficulty.

Given his inability to overcome the Julius Caesar problem, Frege moved on to a definition based on extensions of concepts, the theory of which he saw as being a part of logic. As such, whilst his definition of number was to be explicit, he introduces extensions of concepts (in terms of which he would define number) using his preferred method of contextual definition. However, for an operator to be a part of logic would require logical laws to govern the use of that operator, which led Frege to introduce Basic Law V:

(BLV): $\forall F \forall G\ [\{x{:}Fx\} = \{x{:}Gx\} \leftrightarrow \forall x\ (Fx \leftrightarrow Gx)]$

Frege was a little wary of BLV, writing that:

> "A dispute can arise, so far as I can see, only with regard to my Basic Law concerning courses-of-values (V), which logicians have not yet expressly enunciated, and yet is what people have in mind, for example, where they speak of the extensions of concepts."[6]

And, later:

> "I have never concealed from myself its lack of self-evidence which the [other basic laws of logic] possess, and which must be properly demanded of a law of logic, and in fact I pointed out this weakness [in the quotation given above]."[7]

5 Ibid., §66
6 Frege (1893), Introduction
7 Frege (1903), Appendix

Frege nonetheless held that it was a logical law. Indeed, it seems obviously true – it states that two concepts have the same extensions just when the same objects fall under each concept. However, the adoption of BLV was disastrous. The principle entails a contradiction, giving rise to Russell's paradox. This can be most clearly seen by considering that BLV entails that every concept has an extension; as such, we can let F be the concept 'is a concept that is not a member of its own extension.' Thus the extension of F is made up of concepts. We then ask the question whether or not F is in its own extension: if it is, then it is not a member of its own extension, and so it is not; but if it is not, then it is a concept that is not a member of its own extension and hence is in the extension of F after all – a contradiction. Frege's own response to the paradox was to attempt to patch BLV in some way so as to block the derivation of the paradox. However, he had little faith in such a fix which indeed ultimately proved unsuccessful, as it still contained a contradiction. This failure was enough to make Frege abandon his logicist project.

Wright's *Frege's Conception of Numbers as Objects* (1983) is an attempt to demonstrate that Frege's judgement was premature. He develops the kernel of Frege's logicism in a slightly different way that does not fall foul of the same contradiction. The observation that motivates such a view is that we can break down Frege's derivation of the second-order Peano axioms (PA^2) from BLV into two stages. The first is to derive HP, and the second is from there to derive PA^2 – no further use of BLV is required. As such, Wright proposes to base the relevant definitions on HP rather than on BLV; he then argues doing so preserves many of the Fregean insights that motivated his philosophy of arithmetic.[8]

As with Frege's own programme, there were two parts to Wright's proposal; one mathematical, one philosophical. The technical result is what has become known as Frege's Theorem: that the second-order Peano axioms are equi-interpretable in a theory that appends HP to second-order logic. Call such a system second-order Frege Arithmetic (FA^2). The philosophical aspect is to assess the

8 Of course, this raises the substantial issue of why Wright found the Julius Caesar problem less intractable than did Frege. However, in this thesis I set this issue aside.

significance of such a result. Wright concedes that HP is not a truth of logic, and hence his programme is designed to establish a conclusion somewhat weaker than Frege's own conception. However, he claims that HP is analytic and that this is sufficient to demonstrate the analyticity of its logical consequences – in particular, the theorems of elementary number theory.

§2: Benacerraf's Challenge

In this thesis, I assess neo-logicism taken as a response to a specific challenge that can be seen most clearly in light of Benacerraf's classic paper 'Mathematical Truth' (1973). Benacerraf aims to give necessary conditions for a semantics and an epistemology of mathematics to be adequate before going on to show that these requirements are in tension. Both the semantic and epistemological requirements derive from the thought that there is nothing particularly special about mathematics, and as such our semantics (epistemology) for mathematical discourse should be the same as our semantics (epistemology) for any other area of discourse. What we need to do is "integrate" our mathematical knowledge with our corpus of knowledge from other areas of discourse.

These conditions are given a very intuitive and very minimal gloss. Benacerraf takes it that the only reasonable candidate for the more general semantics is a Tarskian theory of truth, whilst the only reasonable candidate for the more general epistemology is a causal theory. There are two important points here. Firstly, he acknowledges that there are difficulties in both theories, but he sees the devil as being in the detail. There is something compelling about referential semantics and a broadly empiricist conception of knowledge that gives us at least a prima facie reason to accept these requirements however they are to be cashed out. Secondly, if either of these minimal requirements are to be rejected as being appropriate for mathematics, we have two options; either we explain why mathematics is, despite appearances, 'special', or we give a new theory at the same level of generality that can make sense of mathematical and non-mathematical discourse and knowledge. This brings out the key point that Benacerraf's challenge cannot be brushed off by attacking his theoretical commitments; to do so would be to treat only the symptoms rather than the underlying disease.

Thus we have two requirements that derive from a desire to treat mathematics on a par with other areas of discourse – one semantic, one epistemological. Benacerraf's main claim is that these requirements cannot be satisfied simultaneously. A commitment to Tarskian (referential) semantics for mathematical discourse involves a commitment to the existence of mathematical objects which seem to have no causal or empirical properties whatsoever. As such, we cannot explain the knowledge that we have in terms of our more general, empiricist, epistemology. Thus the challenge can only be met by weakening one or other requirement in some way; but such a strategy, of course, entails rejecting the very plausible and minimal conditions offered by Benacerraf.

This very brisk overview of Benacerraf's challenge roughly parallels Wright's presentation of the issue. According to Wright, we need go no further. He writes:

> "[T]he worry about the unintelligibility of abstract objects is a vaguely-based worry; and the empiricist challenge is in any case clear enough for our present purpose. For if we hold that we are capable of grasping abstract sortal concepts, then, at the very least, we attribute to ourselves the ability to identify and to distinguish among themselves objects which fall under those concepts, and to distinguish them from objects of other kinds. And now, how could such an ability possibly be acquired by a respectable empirical route when there is no such thing as an empirical confrontation with an abstract object – when abstract objects are constitutionally incapable of presenting themselves to us in experience? The just challenge posed to the platonist is that he explain how an understanding of any abstract sortal concept could be

> imparted to anyone whose concept-acquiring powers are subject to the constraints imposed by human sensory limitations."[9]

However, to see Benacerraf's challenge in such a fashion seems to me to be slightly too quick. Both Wright's way of setting up the problem and his way of solving it[10] seem to miss the mark. In chapter 1 I develop the challenge a little in order to help come to an assessment of whether or not Wright's proposed solution can be a good one.

Whilst most of the discussion of the problem will be deferred to chapter 1, it is very important (even at this early stage) to be clear on the precise nature of the epistemological challenge that is being put forward and hence the required nature of any appropriate neo-logicist resolution. The problem is not what we might term (following Burgess[11]) *hermeneutic*, where the aim is to make sense of actual mathematical epistemic practice, but nor is it straightforwardly *reconstructive* or *revolutionary*, where our aim is to say in what our mathematical epistemic practice could (or, on some interpretations, should) consist.

To deal with the former (somewhat uncontroversial) claim first; what is at issue is how our knowledge is possible, not the processes by which it becomes actual. The kind of sceptical challenge that is being put forward is that knowledge of even basic arithmetical facts is impossible if we take it that mathematical objects (such as numbers) are mind-independent, abstract objects. Thus an acceptable response is to demonstrate that a route to knowledge is possible – to demand that we discover the actual route to arithmetical knowledge is therefore misguided.

The latter claim requires more delicate handling. The problem with seeing the task as being simply reconstructive is to miss an important constraint of the challenge. This element is noted by various

9 Wright (1983), pp.5-6
10 In particular, see his (1983) §i and §xi respectively.
11 Introduced in Burgess (1983)

writers; as Dummett puts it, "in giving definitions [from which we can derive the laws of arithmetic], *we must be faithful to the received senses of arithmetical expressions*"[12]. Or Wright; "Frege's intention is not merely to introduce a language-game of natural number with some affinities to our pre-existent use of arithmetical concepts; rather, he means to give a philosophical account in depth of that pre-existent use."[13] This point is crucial, as it helps to bring out that we need to explain the knowledge that we in fact have, and not knowledge of some facts expressed by concepts that we have simply made up for our own purposes. We cannot redefine arithmetical and mathematical notions at will – to do that would give us a route to knowledge of a quite different kind to that which we actually have.

The point is most clearly seen by MacBride, who puts it as follows:

> "[T]he neo-Fregean claims [that] HP – properly understood – is nothing more than a *stipulation* that serves to introduce a novel operator into our language... And it is because HP is intended merely as a stipulation that the neo-Fregean feels able to legitimately claim that HP is a priori. Nevertheless, the neo-Fregean continues, HP provides a basis for grasping arithmetical truths a priori because (as Frege's Theorem demonstrates) the system that results from HP and second-order logic allows for a reconstruction of ordinary arithmetical practice in the following sense. It – Frege arithmetic – suffices for the interpretation of the laws of ordinary arithmetic and the proof of their interpretations. It is in virtue of the interpretative powers of the system HP engenders that the neo-Fregean takes himself to be retrospectively entitled to characterise HP as an arithmetical principle."[14]

The key point here is that if HP or a similar principle is to ground our knowledge of *arithmetic*, then

12 Dummett (1991), p.179 (emphasis added)
13 Wright (1983), p.106
14 MacBride (2002), p.129

there is an important constraint on its acceptability that is determined by 'ordinary arithmetical practice.' The stipulation of HP is both acceptable and relevant only if it can be demonstrated that it characterises the concept of cardinal number that is used to frame arithmetical truths. We can introduce a new operator at will, governed by principles that will be analytic and a priori by stipulation, but that will be of assistance only if the new operator turns out to correspond exactly to an old operator. Thus an appropriate response to the sceptic is to provide an *explanation* of how we could come to have the knowledge that we in fact have; to give a response in terms of how we could have had knowledge had we used different, reconstructed, concepts is inadequate.

§3: The Neo-Logicist Solution

Wright's response to Benacerraf, then, is that we can explain our arithmetical knowledge in terms of mastery of arithmetical discourse. Such mastery is attained by understanding HP, which is accorded the status of an analytic truth. As such, it is true just in virtue of the meanings of the terms contained within the principle and so we can know it just by understanding those meanings. But if any logical consequence of an analytic truth is itself an analytic truth, then the truths of arithmetic (by Frege's Theorem) are likewise analytic, and hence arithmetical statements can be known simply by understanding the meanings of the terms contained within. Thus there is an a priori route to knowledge of arithmetic. We understand the meaning of the constituents of HP, thereby making us able to know the truth of HP itself. This understanding (and understanding of second-order logic) is sufficient to derive any truth of arithmetic that is entailed by the Peano axioms. As such, if analyticity transmits across logical consequence, we can know any truth of arithmetic entailed by the Peano axioms.[15]

15 A certain amount of care is required here. The claim is limited to truths derivable from the Peano axioms rather than all truths of arithmetic because the argument given here will not help us ascertain the truth of the Gödel sentence for FA^2, for example. However, the neo-logicist can argue that to require that we need use his approach to explain *all* our arithmetical knowledge is too demanding – we only need to be able to explain enough of such knowledge to get

It is worth noting that whether or not this kind of proposal is correct, it does at least have the requisite structure; it seeks to explain how arithmetical knowledge is possible. It is *arithmetical* knowledge because HP is supposed to adequately characterise our concept of cardinal number, and it is *knowledge* at all because HP and its logical consequences are analytic. Wright is well aware of this when he assesses the significance of his conclusion as follows:

> "Frege's idea is one of immense importance if it can be sustained. Abstract objects are sometimes thought of as constituting a 'third realm', a sphere of being truly additional to and independent from the concrete world of causal space-time. It is this conception of the abstract which generates the well-known epistemological problems to which nominalism and various forms of reductionism and structuralism attempt to respond. But if anything like – or to the extent that – the Fregean conception can be sustained, there can be no epistemological difficulties posed by thought about and knowledge of any particular species of abstract objects which are not already present on the right-hand side of the abstraction which initially introduces the concept of that species. That is not to say that all such problems disappear. But it is a tremendous gain in manageability. And the blanket idea, that there has to be a general problem about thought and knowledge of abstracta, just in virtue of their being abstract, is quite undermined."[16]

Thus Wright's proposal, if it can be sustained, is one that tackles Benacerraf's challenge head-on. He takes it that there is a genuine explanatory challenge here, but he believes that there is a satisfactory response based on Frege's insights.

us started, so to speak. This point is rather more delicate than this brief discussion might indicate, but owing to spatial constraints I cannot give it the attention it deserves.

16 Wright (1997) pp.278-9

§4: Neo-Logicism and the Solution Space.

It is worth pausing at this point to note that taking Benacerraf's challenge as a starting point provides a good way to reveal the commitments of each of the major 'isms' of 20th Century philosophy of mathematics.[17] For example, Hilbertian formalists deny that (at least non-finitary) mathematical statements are truths or falsehoods about mathematical objects (and hence reject the semantic requirement), maintaining instead that what we take to be truths are theorems of a particular formal system. Fictionalists endorse one part of Benacerraf's semantic requirement by maintaining that we have the same kinds of truth *conditions* for mathematical statements as for natural language, but deny that there are any such objects that could make such statements true[18]. Much of the work, then, involves giving both an account of knowledge and an account of what makes some mathematical statements more acceptable than others. Intuitionists, on the other hand, can be agnostic on whether or not Benacerraf's semantic requirement of *homogeneity* is a good one[19]. The claim is that we need non-classical semantics (and logic) for mathematics, and this does not give rise to the problems of reference and knowledge brought in by Tarskian semantics. Thus they either reject Benacerraf's problem outright or claim that it stems from a commitment to classical logic and semantics in natural language.

Nearer to an acceptance of the challenge is Quinean holism. The holist takes the key question to be not how we can explain our knowledge of facts about mathematical objects, but whether the existence of and our knowledge of such objects is required by our best scientific theories. If so, then whilst we do not have an explanation of our knowledge per se, we do at least have a justification of

17 Incidentally, this brings out that neo-logicism is the only 'ism' that accepts that there is a genuine explanatory gap highlighted by Benacerraf's paper and seeks to fill it.
18 Of course fictionalists do not wish to deny that all mathematical or arithmetical statements are literally true. 'There is no greatest prime number' would be an instance of a true claim on a fictionalist account, albeit one justified in a rather different way than on a classical, platonist account.
19 Although, of course, an intuitionist need not be; Dummett, for instance, recommends that the use of non-classical logic and semantics should be more widespread than just mathematics.

our beliefs – that they are entailed by some more fundamental theory. As such, the holist sees Benacerraf as going too far in demanding an explanation of our knowledge over and above a pragmatic view of its indispensability. Again, the challenge is rejected – Benacerraf has not provided necessary conditions on the acceptability of a philosophy of mathematics.

Of the prominent philosophical views of mathematics on offer today, neo-logicism is the only one that tackles the issues raised by the challenge without in some way seeking to weaken or reject it. Therefore an investigation of neo-logicism is philosophically worthwhile provided, of course, that the challenge is a good one. As such, in the first chapter I motivate and defend the importance of the challenge applied to platonist theories of mathematics. I then explore the first part of the apparently promising neo-logicist line of response based on the linguistic turn as underwritten by the Context Principle.

However, even if this linguistic response to the challenge is in principle legitimate, for all that will have been said there is no guarantee that this strategy will work. If there is no acceptable way to put flesh on the bones, then neo-logicism is no more than a pipe dream. As such, in the remainder of the thesis I examine and assess possible developments of the core neo-logicist theses. In chapter 2 I scrutinise Wright's theory, concluding that whilst his approach is inadequate, it fails for interesting reasons. As such, in chapter 3 I give a reconstruction of Wright's argument and alter some key premises such that a marginally weaker and significantly more plausible conclusion is reached. I close with a brief discussion of the outstanding issues.

Chapter 1

§1: Benacerraf's Challenge

In his seminal (1973), Benacerraf outlines a dilemma for any philosophy of mathematics. He offers two constraints on any acceptable theory that are mutually exclusive. The first is that we must have "a homogeneous semantic theory in which semantics for the propositions of mathematics parallel the semantics for the rest of the language."[20] The second is that the same must hold for our epistemology of mathematics – there should be a similar parallel between how we know purely mathematical statements and how we know other kinds of statement. Another way of putting this latter requirement is that we must be able to explain our mathematical knowledge in terms of a more general theory of knowledge. Benacerraf argues that these constraints are necessary conditions for the adequacy of any account and that moreover that one can only be met at the expense of the other.

He first of all gives a sketch of a minimal version of what he calls "the standard view".[21] He introduces such a view by way of an illustration. He gives the following two sentences:

A. There are at least three large cities older than New York
B. There are at least three perfect numbers greater than 17

And asks whether they both share the following logico-grammatical form:

C. There are at least three FG's that bear R to a.[22]

20 Benacerraf (1973) p.403
21 Ibid., p.410
22 Ibid., p.405

Benacerraf claims that (A) ought to be analysed as being of the form of (C), with 'a' naming some particular object ('New York'). The two key components of such an analysis are that singular terms are used to name particular objects, and that the resulting proposition is truth-evaluable. Benacerraf further notes that many philosophers have been less willing to analyse (B) in a parallel fashion. In order to endorse an alternative analysis, one must reject one of the components given above. As such, either surface grammar is no indication of an appropriate semantics, or mathematical statements are not truth-evaluable. Benacerraf claims that all accounts that lack one or other component are defective, as they fail to satisfy his semantic requirement.

The semantic requirement has two components that Benacerraf does not fully distinguish. The first is that the semantics for mathematical statements should be the same as for natural language semantics – if there are differences, these should "emerge at the level of the analysis of the reference of the singular terms and predicates."[23] The second aspect is that we must explain why our mathematical theorems are *true*, rather than merely being theorems of some formal system. Benacerraf takes it that Tarskian semantics is the only one that can provide for this first (semantic) requirement. This has as a natural bedfellow some form of platonist ontology of mathematics, as the point of adopting such a (more general) semantics is to allow us to treat Benacerraf's (B) as we would (A) – namely, on the model of (C). In order for us to do this (and have (B) come out true) there must be some objects to which the singular terms in (B) refer.[24] These objects must be either mind-dependent or mind-independent, and if the latter then either physical or abstract. Mathematical objects are generally taken to be abstract, for good reason; whilst the alternatives may avoid Benacerraf's problem, such accounts are unsatisfactory for a variety of reasons, a fuller discussion of which is beyond the scope of this thesis. As such, I will take it as a premise that if there are mathematical objects, then they are abstract; this is an assumption of what Benacerraf calls

23 Ibid., p.406

24 It is worth noting that it is here where the two aspects of Benacerraf's semantic requirement come apart – one is free to adopt Tarskian semantics without believing in the literal truth of mathematical statements, as fictionalists do. In this chapter I will focus on the consequences of treating the semantic requirement as being linked to the truth of mathematical theorems, as Benacerraf does.

the standard, platonist view.[25]

Benacerraf's bias towards a Tarskian semantics springs, by his own admission, from a lack of viable alternatives. However, Benacerraf takes this to be because such semantics are the only viable ones for natural language as a whole – he writes that his bias "stems simply from the fact that he has given us the only viable *systematic general account* we have of truth."[26] There are plenty of alternative formulations of *mathematical* semantics, but each is inadequate without an explanation of how truth in referential languages links with truth in the new mathematical language. The point is that any new semantics must have the right level of generality, but on that level there are no good alternatives to Tarski.

So much for the first constraint; we account for the truth of mathematical theorems by appeal to the existence of abstract objects and their possession of certain properties. What of the second? The claim is that "we have mathematical knowledge, and it is none the less knowledge for being mathematical."[27] Moreover, "it must be possible to link up what it is for p to be true with my belief that p."[28] This condition is a feature of any epistemology – the difference between any two accounts comes in at the level of *how* truth and belief are linked. Benacerraf takes it that the requisite link must be of the form of a causal explanation – it must be possible to establish some appropriate causal explanation between the truth of p and my belief that p in order for me to count as knowing that p. This, for Benacerraf, is the more general epistemology into which our account of mathematical truth must mesh.

Thus we have two constraints – that we must be able to explain our mathematical knowledge, and that mathematical statements must be truth-evaluable in the same way as natural language statements. Can both be met simultaneously? Benacerraf's claim is that they cannot. The standard

25 Loc. cit.
26 Ibid., p.411 (emphasis added)
27 Ibid., p.409
28 Loc. cit.

view's platonist metaphysics consists of non-spatial, non-temporal, non-causal objects. There cannot be any causal chain involving such objects, so *a fortiori* there can be no causal explanation of how the truth of some statement p that refers to one of these objects is linked to one's belief that p. As such, mathematical knowledge is impossible without violating Benacerraf's second constraint.

§2. Benacerraf Generalised

Under Benacerraf's own formulation of the problem, it might appear open to reject the either Tarskian semantics or a causal theory of knowledge. However, to think that this is adequate would be to miss both the force and the generality of the challenge. The theoretical commitments to Tarskian semantics and a causal theory of knowledge come from their being representative of a more general species of theory. The semantic issue is that our choice of semantics is constrained by the kind of analysis required of (A), (B) and (C); Benacerraf does not claim that Tarskian semantics is *necessarily* the only option, but he claims (almost in passing) that it is both the only viable option and that it is only the referential aspect of such semantics that matter.

Similarly, the causal theory of knowledge is being used to fill a theoretical lacuna. The point is that our mathematical knowledge is in need of *explanation*, but that this explanation must have the form of a more general epistemology that can incorporate or “integrate”[29] our knowledge of mathematics. It is therefore an open question whether such an explanation need posit causal links. But to merely say that the causal theory of knowledge is inappropriate for explaining mathematical knowledge is to leave an explanatory gap. If the causal theory is rejected, some alternative theory is required to fill its place – otherwise we cannot explain our mathematical knowledge.

29 Peacocke (1999), ch.1

The real puzzle is to explain how we, flesh-and-blood creatures, can come to know anything about the world. Any explanation must be able to account for the difference between knowing facts about objects and not having such knowledge; Benacerraf takes it that the best (but not the only) candidate theory is a causal one. However it seems impossible that in the case of abstract objects there could be anything (causal or otherwise) to ground an explanation of such a difference. As Field notes, "the problem is that the claims that the platonist makes about mathematical objects appear to rule out any reasonable strategy for explaining the systematic correlation in question."[30] There is a tension between the ways in which we take ourselves to know about the world and the ways in which we could in principle come to know about mathematical objects. Thus any reasonable positive accounts of the semantics and epistemology of mathematics are incompatible.

Another way in which we might further generalise Benacerraf's challenge is to note that the difficulty that he identifies is largely independent of philosophy of mathematics. This is noted by Benacerraf when he says that "although it will often be convenient to present my discussion in terms of theories of mathematical truth, we should always bear in mind that what is really at issue is our over-all philosophical view."[31] Thus the version of the challenge under scrutiny here is equally applicable to any area of discourse that resolves semantic issues by appeal to abstract objects. An example is modal discourse, in which one might appeal to possible worlds in order to give a semantics for statements about necessary or contingent truths. If these worlds are taken to be abstract in any way (such as sets of true sentences, or some kind of abstract version of the actual world), then the same problem would then arise: how can we explain any knowledge of these possible worlds? The challenge to supporters of abstract objects is that difficulties arise just in virtue of their being abstract.

A further issue is that whilst Benacerraf emphasises the epistemological nature of the problem for those who adopt a platonist metaphysical picture, there is a parallel problem for reference and

30 Field (1984), p.231
31 Benacerraf (1973), p.405

thought. The idea here is that Field's point above – that any reasonable strategy for explaining how we might know about abstract objects is ruled out by their being abstract – applies equally well to strategies for explaining how we are able to even think about these objects. The problem is to account for some feature of our thinking about mathematical objects and the objects themselves; the claim is that this connection is inexplicable. Again, the problem is independent of any particular theory of reference – what matters is that on *any* way of filling out how we can refer to objects, abstract objects will be referentially inaccessible.

The referential challenge is to ask that given the possibility of referential failure, what justifies our thinking that we are referentially successful? It is important to note here that the relevant question is not what *constitutes* referential success – this question is one appropriately answered by giving a theory of reference. Instead, we must ask how (even if it is conceded that we are referentially successful when talking about mathematical objects) we can explain how we could possibly be referentially successful. This in turn brings out two important features of the challenge. Firstly, it is independent of any particular theory of reference. Secondly, the issue is not predominately ontological – the point is that whilst there may or may not be such things as abstract objects, there is a difficulty in explaining how we, with our limitations, could think about or know facts about these utterly undetectable objects.

With these issues in mind, we should consider Ebert's[32] thought-provoking reconstruction of Benacerraf's challenge as being based on the following premises:

1. *Homogeneous Semantic Theory*: the demand that we adopt a general and systematic theory of truth, which – for Benacerraf – should be a Tarskian theory of truth.
2. *Surface Grammar*: the demand to respect the surface grammar of mathematical discourse.
3. *Reference and Object-Directed Thought*: the demand to explain how the objects posited by

32 Ebert (2006)

the semantic theory can, in principle, be in the range of directed thought and talk of the subjects.

4. *Knowledge*: the demand to reconcile the truths of the subject matter with what can be known by ordinary human thinkers. Crucial here is to provide an explanation of how a subject can have mathematical knowledge and on what basis the subject can claim such knowledge.[33]

It seems to me that this kind of reconstruction is a good one, even if Ebert's own attempt slightly misses the mark. For instance, it is not clear that Ebert is committed (as Benacerraf is) to taking certain mathematical statements to be *true*, even if the semantics for mathematical discourse is to mirror that of natural language as demanded by *Homogeneous Semantic Theory*. But despite this, the thrust of his first premise captures the generality requirement put forward by Benacerraf.

Ebert's *Surface Grammar* is a demand for mathematical syntax and semantics to be treated as part of a more general theory. This dictates that syntax should be as good a guide to semantics in mathematical discourse as in natural language. It may be objected that in natural language, syntax is a very poor guide to semantics. Whilst this is true, it does not get to the heart of the issue. The requirement is not that a syntactically well-formed sentence should be equally well-formed semantically, but only that syntax informs semantics with respect to reference. Mathematical singular terms that 'look referential' should refer.

Of course, there are plenty of instances in natural language where what might be taken to be a singular term is clearly non-referential – take Wright's cases of "I did it for John's <u>sake</u>", or "the <u>whereabouts</u> of the Prime Minister is unknown"[34]. However, not every noun phrase need be taken to be a singular term. This observation prompts two questions: when is a noun a singular term? And in the mathematical cases of interest, are the noun phrases singular terms? A discussion of the former question involves investigating the ways of refining Frege and Dummett's criteria for singular

33 Ibid., p.6
34 Wright (1983), p.26

termhood and lies beyond the scope of this thesis[35]. However the second question is of interest, and to answer in the affirmative is a good way to capture the key point of the *Surface Grammar* requirement. It is a demand that the syntax and semantics of mathematical discourse does not *systematically* depart from the normal cases in which syntax is a good guide to semantics. Once we have this, we can then restrict ourselves to considering sentences such as Benacerraf's (B) above – those containing mathematical singular terms that, by *Surface Grammar*, refer.

Reference and *Knowledge* seem to bring out what is demanded by the referential and epistemological challenge respectively. The difficulty is to *explain* how ("in principle") we could possibly have acquired the knowledge of facts about abstract mathematical objects and the ability to think about and make reference to such objects. Ebert's parsing of the debate in such a way is done so that he can introduce what he labels the *Fundamental Assumption* of various responses that he discusses: that "if there is a priori mathematical knowledge and the mathematical discourse is construed at face value, then there has to be some form of acquaintance with the objects involved that underwrites this knowledge."[36] Whilst this seems too strong – for instance, the requirement that a subject be *acquainted* with an object far outstrips the requirement of the existence of a causal link between subject and object given in Benacerraf's paper – Ebert is right to attempt to pin down the commitments of the genus of theories of knowledge or reference of which the causal theory is a species.

A more successful attempt to identify this kind of commitment is made by Hale and Wright when they criticise the idea that

> "knowledge of truths about objects of any kind must involve some form of prior interaction or engagement with those objects. That notion is naturally taken to call

35 Hale attempts to build on earlier work by Dummett in order to give such an account – see Hale (1994), (1996) and Dummett (1973), especially pp.54-8, 174-9.

36 Ibid., p.16

> for some kind of physical connection... and so is obviously inimical to the abstract. But it is not obviously a notion that must be accepted. To be sure: if we forget the 'prior' and the notion is given a sufficiently broad construal, so that possession of any sort of identifying knowledge of an object suffices for 'engagement', the idea reduces, near enough, to a truism – one can hardly be credited with knowledge of truths about objects unless one knows which objects are in question. But so construed, it need raise no hurdle for platonism. The crucial thought – we should say: 'mistake' – is the additional idea that such engagement is presupposed by and must be already in place before any knowledge of truths about objects can be had."[37]

The point is that there is no principle that *must* be accepted that can be used to tell against the possibility of knowledge of facts about abstract objects. However, it is not fair to say that the lack of such a plausible principle draws the sting from Benacerraf's challenge; as has been emphasised, the challenge should be seen as a call for a *positive explanation* rather than being a straightforward objection to platonism. The attraction to the kinds of "prior interaction or engagement" principles comes from the fact that such interaction or engagement with objects can help explain knowledge of facts about those objects; thus the rejection of such a principle leaves unfilled the explanatory gap highlighted by Benacerraf.

Instead, then, let us try the following reconstruction along the same lines as Ebert:

1. The correct semantics for mathematical discourse is the same as that for natural language.
2. The syntax of mathematical statements is a good guide to the semantics – it is not systematically misleading. For instance, mathematical singular terms refer.
3. The correct theory of reference for mathematics is the same as that for natural language, a theory which must explain how subjects can think about and refer to the objects posited by

37 Hale & Wright (2002), p.114

the semantic theory.

4. The correct theory of knowledge for mathematics is the same as that for natural language, a theory which must explain how subjects could come to know facts about the objects posited by the semantic theory (if such facts are known).

This seems to capture what is right about Ebert's premises, and fits well with a parallel 'fundamental assumption': that the explanation demanded by (3) and (4) must involve the notion of “prior interaction or engagement”. This final assumption is somewhat loosely formulated, but Hale and Wright's point is that of the ways in which it could be fleshed out, each is unobjectionable if true or false if hostile to the platonist.

At first blush it seems obvious that accepting something like this assumption is necessary, and that as such it may not be rejected lightly. It seems to unite the genus of theories of reference and knowledge to which I alluded earlier, and does seem to have a certain explanatory power – for how could we know facts about objects without such prior engagement? Therefore to reject the assumption leaves a lacuna. In the remainder of this chapter I will discuss Hale and Wright's own (neo-Fregean) proposal – that we can know (at least) certain arithmetical truths by inference from a principle governing numerical identity, a principle that in turn has a particular explanatory epistemic status.

§3: The Linguistic Turn

Frege's linguistic turn is briefly summarised by Dummett as follows:

> “[Frege's] solution was to invoke the Context Principle: only in the context of a sentence does a word have meaning. On the strength of this, Frege converts the

> problem into an enquiry how the sense of sentences containing terms for numbers are to be fixed. *There* is the linguistic turn. The Context Principle is stated as an explicitly linguistic one, a principle concerning the meanings of words and their occurrence in sentences; and so an epistemological problem, with ontological overtones, is by its means converted into one about the meanings of sentences."[38]

The crucial idea is that the 'normal' order of explanatory priority is reversed. Rather than explaining how we know the fact expressed by given sentences in terms of how we know facts about the constituents of that sentence, we can know facts about the objects to which each constituent of the sentence refers by understanding the contribution it makes to the truth-conditions of that sentence.

Benacerraf's problem, when suitably generalised, is a demand for a positive explanation of a connection that, it is charged, cannot be adequate. This is because certain kinds of objects cannot fit into our account of semantics and truth whilst being epistemically accessible. The response is that we can gain knowledge of facts about such objects by understanding what it takes for sentences about such objects to be true. There are two thoughts here; firstly, that we do not need some *prior* understanding of the meaning of expressions that purport to refer to abstract objects in order to attain such mastery of the relevant discourse. Secondly, we do not need to explain how we can know facts about certain objects in terms of their possession of properties that make them epistemologically or referentially accessible – as indicated by the earlier quotation from Field, such an approach is unpromising. Instead an alternative explanatory route is available.

The neo-Fregean strategy applies the linguistic turn to abstraction principles in order to fix the sense of terms that purport to refer to numbers. Fregean Abstractions have the following form:

(FA): $\forall F \forall G\, [\Phi F = \Phi G \leftrightarrow (F \approx G)]$

38 Dummett (1991), p.111

Where F and G can be objects or concepts of any level, Φ is a term-forming functional operator on F and G, and ≈ is an equivalence relation on F and G. A rough-and-ready characterisation of the main idea is to say that one can introduce singular terms on the LHS of the equivalence to give a criterion of identity to the objects to which such terms purport to refer. As the equivalence is a necessary one, we can say that both the LHS and RHS of the equivalence have the same truth-conditions. We then point out that the RHS seems to be epistemologically and referentially unproblematic; as such, we conclude that we can come to know facts about the objects that are the referents of the terms of the LHS and be warranted in taking our use of the terms of the LHS to be instances of referential success by citing their connection to the facts given by the truth-conditions of that expressed by the RHS.

This characterisation brings out that this application of the linguistic turn is best understood as a conjunction of two theses. This can be illustrated by considering the example used by Frege, the Directional Equivalence:

(DE): The direction of line a is the same as the direction as line b iff a is parallel to b.

Prima facie, the LHS of the biconditional makes explicit reference to abstract objects – namely, directions – whilst the RHS does not. Thus there is a substantive question as to which objects a speaker is committed if he is to endorse these equivalences. The first claim, then, is that each side of the abstraction makes reference to distinct kinds of objects and that they are as such *ontologically plenitudinous*: we should take it that in the case of DE, one ought to believe that both the concrete objects 'lines' and the abstract objects 'directions' exist. The second thesis is that we can explain the knowledge that we have about these directions as being derivative knowledge from knowledge of facts about parallel lines. Thus the RHS has explanatory or *epistemic priority*[39] over the LHS. This

39 A brief terminological note: Hale – cf. his (1995), pp.205-7 – calls these claims the 'epistemic priority' thesis and 'ontological priority' thesis. I dislike the latter as it seems to suggest a (somewhat implausible) reading of abstractions that there is no commitment to the referents of the RHS at all. The point is that in the illustrative case of DE, we are committed to the existence of directions *as well as* (not *instead of*) lines.

claim is that we can explain what facts we know about the objects referred to by the LHS by what we know about claims that fit the model of the RHS.

Both the epistemic priority thesis (EPT) and the ontological plenitude thesis (OPT) cannot be taken to be utterly general theses, but they need not be for them to serve the neo-logicist's purposes. For instance, the EPT should not be taken to apply to an abstraction principle governing criteria of identity for shapes – that 'shape (A) = shape (B) iff A is similar to B' – as in such a case, it is conceivably more plausible to explain similarity in terms of sameness of shape. However, each thesis need only apply to particular classes of abstraction principles which in turn may or may not have a neat characterisation. What is needed are general considerations in favour of each thesis that may or may not have exceptional cases; all that needs to be established is that a given abstraction fits the mould that is being pushed here.

With the OPT and the EPT made explicit, we can bring the characterisation of the linguistic turn into sharper focus and see exactly how it is supposed to solve the reference problem. The LHS contains terms that purport to refer; given the possibility of referential failure, what gives us warrant to deny that this is such a case? The response is that given the OPT, we cannot endorse the apparently obvious abstraction without thereby believing in the abstract objects to which the LHS appears to make reference. By the EPT we can come to know the facts expressed by the abstraction under their description in terms of the vocabulary of the RHS (say, in terms of talk about parallel lines) without requiring any understanding in terms of the vocabulary of the LHS (in terms of directions). As such, we can justify our belief in abstract objects by our belief in the abstraction and our understanding of the facts that make the RHS of a given abstraction true.

My claim here is not that there *is* any such abstraction principle that could justify our belief in mathematical abstracta – whether or not there is such a principle is a matter I investigate in chapters 2 and 3. The claim at this point is only that this strategy could provide the bare bones of an acceptable starting point – in other words, that there could be such a principle remains an epistemic possibility. To dispute this claim is to dispute the legitimacy of this kind of performance of the linguistic turn. It seems to me that there are three main ways one could do this. One could reject the EPT, reject the OPT, or reject the idea that the alleged necessary equivalence of abstractions are necessary in a way that underwrites the EPT and OPT. To reject the EPT is in effect to claim that it is insufficient for knowledge, or to warrant referential success, to understand the relevant abstractions and their RHS. To reject the OPT or the coherence of any suitable necessity of the abstraction is to claim that abstractions cannot in principle be adequate to bear this sort of burden.

It is important to note here that we are assuming at this point that abstraction principles are an acceptable form of (something like) implicit definition. This assumption is roundly criticised by Dummett, amongst others, but a discussion of this would take us too far afield[40]. The core of the neo-Fregean strategy is to use abstraction principles to generate results of particular interest, and then claim that (by the linguistic turn) these kinds of results will be sufficient. Of course, if abstraction principles are utterly without merit, then the strategy will fail; but this need not yet be of concern. The relevant issue here is whether abstraction principles can in principle generate the desired results, something that will not be the case if any of the objections raised in the previous paragraph have any substance.

In assessing the substance of these objections to the legitimacy of the linguistic turn, it will become clear that Frege's Context Principle is indispensable to one who wishes to defend the neo-Fregean corner. Hence before continuing with an examination of the EPT and OPT, it seems worthwhile taking the time to make clear exactly which tools are available to the neo-Fregean. As such, in the next section I discuss the Context Principle before going on to see how it can be used to justify the EPT and OPT.

40 See Dummett (1991), pp.226-9

§4: The Context Principle

The Context Principle was first stated by Frege in the introduction to *Grundlagen*, writing that one should "never... ask for the meaning of a word in isolation, but only in the context of a proposition."[41] Frege took it as a methodological principle of some sort; but as Dummett[42] points out, the Context Principle is first and foremost a principle governing meaning. However, it was endorsed by Frege before he first drew the distinction between the sense and the reference of expressions, and as such it is not clear at first whether it is right to interpret the Context Principle as governing sense, reference or both.[43]

To treat it as a principle governing sense is to give it a very plausible reading, as it amounts to "the conceptual priority of thoughts over their constituents: the constituents *can* be grasped only as potential constituents of complete thoughts"[44]; it is moreover a principle that "governed Frege's thinking from start to finish."[45] According to Hale and Wright[46], it is this line of thought that is developed by Wittgenstein in his critique of the Augustinian conception of language. The idea here is that we do not learn a (first) language in a piecemeal way by grasping the individual constituents of thoughts and combining them to create complete thoughts; instead, we grasp those constituents by understanding the contribution that they make to the thought. This is obviously fairly compressed, but the main idea seems compelling; moreover it is one that has the assent of the main protagonists in the debate.

But what of treating the Context Principle as governing reference? According to Dummett, it is

41 Frege (1884), Introduction
42 Dummett (1991), p.183
43 Whether or not it was *Frege's* intention to endorse the principle in any such way is not my main concern. I will only be interested in assessing the plausibility of the principle as so construed.
44 Ibid., p.184
45 Loc. cit.
46 See Hale and Wright (2002), especially pp.113-9

coming to an assessment on this matter that is the primary task of much of Wright's (1983). Wright interprets the Context Principle as "the thesis of the priority of syntactic over ontological categories."[47] This is explicated as the view that "the category of objects in particular, is to be explained as comprising everything which might be referred to by a singular term, where it is understood that possession of reference is imposed on a singular term by its occurrence in true statements of an appropriate type."[48] Thus it is not that we assess whether or not the singular terms in a given sentence have reference before going on to assign a truth value to that sentence; instead we make that assignment of a truth value and understand the reference of terms by the contribution that they make to that assignment.

Views of this kind would have two further requirements; that there is a plausible means by which we can determine the truth-conditions of sentences, and that there is a plausible characterisation of singular termhood. However, it is important to get the dialectic right in each case. With respect to the latter, syntactic, requirement, Dummett proposes that we "understand [Wright's term 'syntactic role'] somewhat loosely; we may perhaps leave problems of syntactic classification *to be dealt with as they arise*, contenting ourselves with the reflection that we can in practice judge reasonably well whether or not an expression would count as being a 'proper name', even if we cannot precisely formulate the principles underlying our judgements."[49] The point here is that there need not be a demand for a detailed account of what counts as a singular term – good judgement is all that is required.

How do things stand with respect to our other, truth-conditional, requirement? At first it is not clear exactly to what this requirement amounts; if we are working within a broadly referential semantics, sentences containing apparently referring terms are true or false only if the relevant terms actually refer. But perhaps to see the requirement in this way reveals that we are free to choose our method

47 Wright (1983), p.51
48 Ibid., p.52
49 Dummett (1991), pp.185-6 (emphasis mine)

of determining truth-conditions apart from and prior to our worrying about the reference of the relevant singular terms. Again, the point seems to be that (at least for the time being) we are entitled to use our judgement.

How might this bear on our problem of justifying our belief in the existence of certain kinds of abstract objects and our mathematical knowledge? To answer this it is helpful to consider Fine's[50] enlightening gloss on the Context Principle. He points out that the problem arises because we have no way of distinguishing actually referring terms from merely potentially referring terms with respect to certain kinds of abstract objects. We take it that abstractions are a suitable class of sentences that contain the relevant singular terms, and that we are laying down the abstraction principles "from a standpoint in which the existence of the objects that are to be assigned to the terms is not presupposed."[51] However, if such a stipulation is successful then we have guaranteed referents for the singular terms of the LHS. He puts it in terms of an adoption of the Context Principle as, in effect, a way of bridging the gap between potential and actual reference. He writes

> "The general idea behind the principle...is that linguistic practice may be partly constitutive of reference. The fact that certain terms are used in a certain way may guarantee, in conjunction with the appropriate non-linguistic facts, that those terms refer and that they refer to what they do... The view is that the apparently referential behaviour of terms may help secure their reference: terms that behave *as if* they refer *will* refer, given that the appropriate non-linguistic facts are in their favour."[52]

Thus Fine's interpretation of the Context Principle seems to provide, in one sense, a very direct solution to the reference problem. The world may or may not be such that for any given potentially referring term, that term actually refers; the point is that working out whether or not we succeed is not a matter of giving an account of an apparently mysterious connection between ourselves and the

50 Fine (2002)
51 Ibid., p.56
52 Ibid., p.57

relevant objects, but is instead a matter of certain non-linguistic facts obtaining and of linguistic practices. The alleged problem for platonism is that even if there are abstract objects, there is no way that we can account for our ability to refer to them. One reading of Fine's response is that the Context Principle can do just that – if there are such abstract objects, then linguistic practice will help guarantee referential success. However, whilst this response is along the right lines it is somewhat oversimplified; I will return to it in discussion of the OPT.

A similarly provisional and underdeveloped response to the knowledge problem (to which I will return in chapter 3) can be seen to come from these considerations. This point is addressed by Hale and Wright who (in the context of a discussion of DE taken as an illustrative abstraction) say that:

> "So long as we can ascertain that certain lines are parallel, there need be no *further* problem about our knowledge of certain basic kinds of facts about directions, for all their abstractness. For provided that the concept *direction* can be implicitly defined by Fregean abstraction, we can know statements of direction-identity to be true just by knowing the truth of the appropriate statements of parallelism among lines. We can do so for the unremarkable reason that the truth-conditions of the former are fixed by stipulation to coincide with those of the latter."[53]

Part of what makes this kind of response merely provisional is that it appeals to the truth-conditions of statements about abstract objects being *stipulated* to coincide with truth-conditions of statements about concrete objects. Similarly, there is the proviso that the line of response requires one to take Fregean abstractions as implicit definitions. As I go on to discuss in later chapters, it is not clear to me either that they should be treated as such or that they need to be; however, let us grant this for now to help bring out the force of the response.

53 Hale & Wright (2002), p.119. I have removed illustrative references to the discussion of numerical identity.

How we know facts about abstract objects neatly builds on how we know that our reference to abstract objects is successful. As Hale and Wright point out, we need only know the truth of the relevant equivalence captured by an abstraction principle and the truth of certain statements about concrete objects to know that certain statements about abstract objects are true. But we can know the truth of the relevant abstractions by understanding the meaning of the terms contained within them – if they are analytically true by stipulation – and hence there is "no *further* problem" about knowing the truth of statements about abstract objects. Thus if we can know that abstractions succeed in referring to the abstract objects that they do in fact refer (as is suggested by our reading of Fine's discussion of the Context Principle), we have a route to knowledge of facts about those abstract objects.

§5: Objections to the Linguistic Turn

Recall that we have an argument for our belief in abstract objects being justified that makes essential use of abstraction principles. The idea is that we have impredicative necessary equivalences that demonstrate that the same fact can be expressed with or without (explicit) reference to certain classes of objects. From this, we give reasons to think that our acceptance of such abstractions justifies our use of the singular terms contained in them as genuinely referring terms, and moreover justifies our belief in the existence of their purported referents. There are therefore three theses open to objection: the thought that abstractions express an important kind of necessary equivalence that underwrites the EPT and the OPT, the OPT itself and the EPT itself. I will assess each of these in turn.

§5.1: Abstractions and Equivalence

The attempt to establish the EPT and OPT by appeal to features of abstraction principles will only get off the ground once we are clear about the kind of equivalence that abstractions are supposed to express. We can say that the LHS and RHS 'state the same fact' or something similar, but this is both unnecessarily unclear and leaves unanswered two questions about any further interpretation. The first is whether it is plausible to say that abstractions are principles governing 'sameness of fact' (however that is cashed out), and the second is whether this 'sameness of fact' will be adequate to underwrite the EPT and OPT.

Whilst Dummett[54] and Hale[55] argue that treating abstractions as stating that the LHS and RHS *have equivalent senses* or *express the same thought* is not consistent with the EPT and Frege's understanding of the sense of sentences as being composed of the senses of their constituent parts, it seems to me extremely plausible to interpret abstractions as making a claim about the *truth-conditions* of the LHS and RHS. The point is supposed to be that for the LHS and RHS to express the same thought is to say that there is no epistemic priority between one or the other – to grasp the content of the LHS *just is* to grasp the content of the RHS and vice versa. However, to say that abstractions express an equivalence of truth-conditions seems more fitting in that it allows for one side to be epistemically prior to the other without prejudging which side has that priority. There is of course the outstanding question of whether the EPT and the OPT have any traction in the cases relevant to our discussion, but to press this is just to move on to objecting to the EPT or OPT on independent grounds. Thus provided the proponent of the linguistic turn is cautious at this point, there is very limited mileage in objecting to this aspect of abstraction principles.

54 Dummett (1991), pp.168-70
55 Hale (1995), pp.192-7

§5.2: The Ontological Plenitude Thesis

The most obvious line of objection to the OPT is to say that we should take the RHS of abstractions to be a better guide to the constitution of the class of objects in which we ought to believe if we endorse the abstraction. The idea is that even if the EPT is true, it does not help us to establish the OPT. The proper conclusion to draw is not that the LHS has inherited the RHS's status as being epistemologically and referentially unproblematic despite it being about abstract objects, but that the LHS is epistemologically and referentially unproblematic *only because* it is *not* really about abstract objects at all. All parties can agree that both sides of the abstraction are about the same thing, but the claim is that we have reason to think that the surface grammar of the RHS is a better guide to our ontological commitments than that of LHS. We have to decide whether this is a case in which the surface grammar (of either side of the abstraction) is likely to be misleading, and the heart of the objection is that there is an allegedly coherent and convincing explanation of how the neo-logicist gets things wrong. As noted in §2, there are plenty of examples in natural language of terms that 'look referential' but are not; the objection here is that terms purporting to refer to abstract objects can seen as an example of such pseudo-reference, as in the case of 'sakes' or 'whereabouts'. Treating them as genuinely referential is motivated only by a desire to make the identity claim of the LHS a genuine identity. This does not warrant inflating our ontology.

It seems to me that there is a promising line of response to this kind of objection that draws on the Context Principle as a principle governing reference. Recall the slightly flat-footed and oversimplified view presaged at the end of §4; we have terms that behave referentially, and by the Context Principle such terms can be guaranteed reference by the contribution that they make to true sentences. As such, if the appropriate sentences (the abstractions) are in fact true – and the potentially referring terms actually refer – then we will be justified in believing in the existence of the relevant abstract objects in virtue of our justification in believing such relevant sentences.

The difficulty with this line of response is that using the Context Principle only justifies our using the relevant singular terms that stand for abstract objects in the way that we do. However we are not yet warranted in drawing the further conclusion that we are justified in believing in abstract objects; if some coherent alternative is available, we have not yet done enough. It is a position of this sort that Dummett endorses, taking an 'intermediate' reading of abstraction and calling it 'tolerant reductionism' (in contrast to the 'austere' reading of the 'intolerant reductionist'). The intolerant reductionist cannot endorse the Context Principle as a principle governing reference owing to considerations discussed in the foregoing paragraphs, but the tolerant reductionist is not so constrained – he can make such an endorsement but need not be committed to the OPT as a result.

Dummett's claim is that to cast the debate as being between the robust and austere readings is to fail to keep separate two distinct aspects of Frege's account of reference: reference as a relation between a name and its bearer, and reference as semantic role – as Wright puts it, “the contribution it makes to determining the references (truth-values) of complex expressions (sentences) in which it features.”[56] On Dummett's view, to fully determine the reference of a singular term is to say both how that term relates to a particular non-linguistic object and how that term relates to other (more complex) linguistic expressions. Thus the Context Principle (as a principle governing reference) informs our reading of abstractions only inasmuch as it informs our understanding of the purely linguistic aspect of the reference of the singular terms of the LHS.

Dummett does not want to make the implausible claim that these two aspects of reference are independent of one another – indeed, he makes the case that for concrete objects, the two aspects are coincident. When reference to concrete objects is in question, there can be no way for a singular term to make a contribution to the truth conditions of more complex expressions except by there being some external object that has the requisite properties and behaves in the requisite fashion. However, Dummett argues that there is a relevant disanalogy with abstract singular terms, and that

56 Wright (1983), p.66

as such we cannot justify taking such terms to possess realistic reference.

I do not intend to discuss the details of Dummett's argument to withhold attribution of realistic reference to abstract singular terms for two reasons. The first is that even if it is successful, it is left open that there could be some other way to justify taking thin and realistic reference of abstract singular terms to be coincident (and, of course, that such terms would indeed refer to abstract objects). Secondly, this more nuanced line of objection to the OPT is clearly reliant on there being a coherent and stable middle ground between the austere and robust readings of abstractions. It seems to me that one can follow Hale[57] in putting pressure on the plausibility of this necessary condition for Dummett's argument to get off the ground.

Hale parses Dummett's objection in terms of 'thin' and 'realistic' reference, where a term has realistic reference only if it has an object in the world as its referent and thin reference only if its semantic role is to contribute to the truth-conditions of more complex expressions. He goes on to say that "it is an objection to platonism that directions can be objects only of thin reference *only if* it is contended that objects of realistic reference may be held to exist in some sense in which objects of thin reference may not – and for that contention Dummett offers no argument whatsoever."[58] The point is clearly supposed to be that we need to have a reason for thinking that a term having thin reference could fail to adequately justify a belief that it also has realistic reference, and no such reason has been provided.

This is undoubtedly the key issue, but first it is helpful to get clear on the precise dialectical situation. There is already a standing objection to platonism based on Benacerraf's challenge, and Dummett's position is supposed to demonstrate a shortcoming in a particular line of response. His distinction between two aspects of reference – one purely semantic, the other a relation between a linguistic expression and an external object – is key to this. The difference in these two aspects of

57 Hale (1995)
58 Ibid., p.205

reference can and should be granted at the conceptual level; the burden of argument is on the platonist to show that even if there is a worthwhile conceptual distinction to be drawn between thin and realistic reference, the distinction would disappear at the ontological level. In other words, the platonist must show that thin reference of terms for abstract objects is going to suffice for the existential guarantees that the platonist requires. Such an argument would demonstrate that Dummett's distinction need not be an obstacle to justifying our beliefs in abstract objects, and hence that he would not have succeeded in exposing that the current line of response to Benacerraf lacks sufficient motivation.

The question, then, is whether Hale has provided such an argument. Hale (as adumbrated by Wright[59]) suggests that Dummett's intermediate position can be upheld only if there can be different senses of 'exist' – one to correspond to the (realistic) existence of referents of terms with realistic reference, the other to correspond to the (thin) existence of referents of terms with merely thin reference. The point is that without such a conceptual distinction, the distinction between the two notions of reference will indeed disappear at the ontological level – whilst we will have succeeded in pointing out two distinct features of what happens when a term refers, we will not have succeeded in showing that they can be held apart. But this is just what is required for the intermediate reading of abstractions to stand alone and not collapse into either the austere or robust reading. Hale notes that Dummett is (for good and familiar reasons) unwilling to endorse a distinction between realistic and thin existence, but without such a distinction it is very hard to see how the neo-Fregean conclusion – the OPT – can be resisted in this Dummettian way.

59 Wright (1983), p.83

§5.3: The Epistemic Priority Thesis

Finally, we have the EPT. Recall that this thesis is a claim that it is sufficient to understand abstractions and the description of the facts expressed in RHS-vocabulary in order to be able recognise the facts as expressed in LHS-vocabulary. Our putative objector is therefore claiming that on the linguistic turn, the order of explanation is wrong. In order to understand what makes statements about abstract objects true, we must understand something about those abstract objects. Thus an explanation of how we know anything about abstract objects in terms of senses of statements about those objects is invalidated. However, whilst a position of this sort is very natural, it is rarely argued for. Moreover, as Hale and Wright note, it is this idea that is

> "what is embodied in the Augustinian conception of language that is put up for rebuttal at the very outset of the *Philosophical Investigations*; and the prime spur towards the 'naturalist' tendency which finds abstract objects per se problematical is the idea, at the heart of the Augustinian conception, that some... causal relationship must lie at the roots of all intelligible thought of, and hence reference to objects of a particular kind. While this is not the place to enlarge on the relevant points in detail, it is by no means an uncommon reading of *Philosophical Investigations* to believe that the book as a whole accomplishes a compelling critique of this idea."[60]

The thought here is that the use of knowledge of senses of sentences to explain the sense of expressions is not so outlandish, whilst its contrary is not as compelling as might first be thought. If language more generally can be seen to work this way, it should certainly work for each area of discourse in which an integration problem arises.

60 Hale and Wright (2002), p.116

However another (somewhat easier) way to defend the EPT is to argue that it follows from some of the less controversial aspects of the discussion of the Context Principle in §4. The EPT is the result of an application of the Context Principle to abstractions. If (contrary to the EPT) the LHS were to have epistemic priority, then we would have to have some grasp of the objects referred to by the singular terms of the LHS *prior* to our giving the relevant identity conditions embodied by the RHS of an abstraction. If this is to be possible, however, we would have to have grasped the sense of such terms independent of propositions that are true of them – precisely what is ruled out by an endorsement of the Context Principle as a principle governing sense. Thus endorsing the Context Principle will give us sufficient grounds for endorsing the EPT.

If this line of argument is cogent, is the legitimacy of the linguistic turn in principle now beyond question? There is one further (albeit somewhat desperate) ploy available. The thought is that one should accept this neo-Fregean line of argument up to a point, but see the conclusion as a kind of *reductio ad absurdum*. The conclusion, recall, is that one can use abstractions to transfer the status of claims about the RHS as epistemically and referentially unproblematic to claims about the referents of the terms of the LHS. The proposed response would be to say that we already (from Benacerraf's argument) have good reason to think that the objects of the LHS are insurmountably epistemically and referentially problematic; given that the heart of the EPT is to say that the LHS and RHS have the same epistemic status, we should take it that the RHS is just as problematic as the LHS – and that is our reductio. There are no problems referring to or understanding facts about particular parallel lines, for instance. As such, we must jettison one of our premises – the OPT, the EPT, or that abstractions capture some sort of interesting necessary equivalence.

This kind of response is simply question-begging and misses the point of the linguistic turn entirely. It is clear that the impossibility of resolving Benacerraf's dilemma is a further premise in the foregoing objection but it is a premise that is taken to be sacrosanct for no good reason. Indeed, the point of the linguistic turn is to take Benacerraf's challenge seriously but demonstrate that it is *not* impossible to meet. There is nothing contained in the preceding line of argument that gives us independent reason to take the challenge as impossible to meet, and as such it can be rejected.

§6: Conclusion

I have so far defended the view that to take the linguistic turn to resolve Benacerraf's challenge is an open epistemic possibility. The fairly minimal (although nonetheless substantial) claim is that no prior characterisation of our connection to abstract objects is necessary for knowledge of facts about them or reference to them; instead, we can attain knowledge by understanding the senses of appropriate statements in which abstract object terms appear. However, it is not yet clear that there are any such appropriate statements. Whilst the neo-Fregean will claim that there is an abstraction that will do the job, this raises the issue of which such abstractions are suitable candidates. In the following chapter I investigate whether what is taken (by Wright and others) to be the best candidate for solving Benacerraf's problem with respect to number, Hume's Principle, is fit for purpose.

Chapter 2

The *locus classicus* for neo-Fregean logicism is Wright (1983), in which he attempts to establish a view that consists of two components; one mathematical, one philosophical. The former, technical, result has become known as Frege's Theorem – that second-order Dedekind-Peano arithmetic[61] is equi-interpretable in Fregean second-order logic with what has become known as Hume's Principle

(HP): $\forall F \forall G\, [\#F = \#G \leftrightarrow (F \approx G)]$ for $\approx$:= "$\exists$ 1-1 correspondence"

appended as an axiom. The domain of HP is taken to be the class of all Fregean concepts (with suitable qualification to be explicated later), whilst the # operator is a function that is taken by Wright to be short for 'the cardinal number of'. The latter, philosophical, project then consists in assessing the significance of Frege's Theorem. The main thought here is that HP and second-order logic each have a certain epistemological status, and that 'recognisable logical consequence' can preserve such a status.

The idea, as noted, is to answer the Benacerrafian challenge by using a strategy based on adopting an abstraction principle as an axiom. The abstraction, HP, is designed therefore to have two epistemological roles that correspond to two parts of Wright's argument. The first is to guarantee that we can have knowledge of facts about the referents of the terms of the LHS by understanding what it takes for the RHS to be true. It is this part of his argument that I assess in Chapter 1. The second is to provide an epistemologically unproblematic foundation for any theory that takes such an abstraction as an axiom. Frege's Theorem guarantees that if epistemic status may be transmitted across recognisable logical consequence, then arithmetic will have the same epistemic status as that of HP.

61 Henceforth I will drop the qualifier 'second-order Dedekind-Peano'.

It is worth pausing at this point to make a brief digression. As the foregoing remarks indicate, Wright's neo-logicism is designed to take on Benacerraf's challenge entirely at the level of addressing the epistemological issues that it raises. Another form of neo-logicism, championed by (for instance) Tennant[62] and Rumfitt[63] seeks to attack on both fronts by changing the semantic requirements and solving a consequently different set of epistemological problems. The idea is to work in a free logic rather than standard second-order logic; the key difference is that on a free logic, singular terms do not need to refer. The difficulty for a theorist working in such a logic, then, is not to give an account of how we can justify thinking that our terms refer but how we justify a further auxiliary principle that tells us when a singular term refers and when it is an empty name.

There are some interesting features of any neo-logicist system governed by a free logic, of which I will note three. The first is that for Wright, whether or not this kind of approach is correct, it is certainly overly tentative; in his view there is no need to allow for the possibility (or even the coherence) of numerical terms failing to refer. Secondly, the line of argument that I will advance in §2 that makes essential use of counterexamples to HP would be relatively ineffectual; such cases could only (at most) be purported counterexamples to the conjunction of HP and the auxiliary principle governing reference of singular terms, and as such HP can be maintained in its full generality (provided that one accommodates the potential problem case by making a change to the auxiliary principle). Thirdly, giving an assessment of the epistemic status of the non-logical principles of a neo-logicist programme would no longer be limited to an assessment of particular abstractions – instead, we would have to assess the status of the auxiliary principles such as Tennant's Ratchet Principle or Rumfitt's suggested improvement. It seems to me that there is plenty to say on the subject of abstractions, and a discussion of auxiliary existence principles within a free logic should be treated as part of a separate project that need not concern us here. Thus for reasons of space, I will leave this kind of approach to one side.

62 See Tennant (1987), (1997)
63 See Rumfitt (1999), (2001)

To return to the issue at hand – assessing the epistemic status of HP and its consequences – let us give a more regimented reconstruction of Wright's argument:

1. Frege's Theorem: Arithmetic is a (recognisable) second-order logical consequence of HP.
2. If HP is analytically true, its recognisable logical consequences are analytically true.
3. HP is analytically true.
4. Therefore, arithmetic is analytically true.

This argument is clearly valid. (1) is a mathematical theorem that is not open to doubt. (2) is dependent on the notion of analyticity in play – if it is sufficiently epistemologically loaded then (2) may come out as trivially true, whereas if it is is strongly metaphysical then (2) may be far from obvious[64]. Despite this, let us for the moment grant (2) and focus in this chapter on assessing (3) – the claim that HP is analytic.

In this chapter I will evaluate Wright's argument for the analyticity of HP, and consider what epistemological significance such a result would have. I then discuss some (at least prima facie) counterexamples to HP, before exploring how HP might be modified to accommodate these troublesome cases.

§1: Wright's Argument for Analyticity

Wright's argument for the analyticity of HP has a number of strands to it; some positive, others more defensive. The reason for this is partly historical, as he investigates why Frege did not, upon the discovery of the inconsistency of BLV, simply adopt HP as an additional axiom. Frege's reason

64 And, indeed, may give rise to an epistemologically deficient conclusion!

was an inability to solve the Julius Caesar problem – that we cannot (merely thanks to our definition of number) resolve the truth of mixed identity statements. Wright takes it that this is one of the "several specific grounds on which it might be held that HP cannot function as... an implicit definition"[65], and that "much of the defensive work required of the neo-Fregean... consists in dealing with doubts and objections on this score."[66] However, I will set aside such a defensive task in order to concentrate on a more positive line of argument – that Fregean abstractions (and HP in particular) can have an epistemic status accorded to informative definitions.

Wright's argument has both a general and a more specific aspect to it. He notes that abstractions are being used as implicit, non-eliminative definitions of terms. He goes on to argue that the mere fact that we cannot guarantee being able to use such definitions to eliminate all instances of the definienda need not count against them. To move to the specific strand, he argues further that if the method of implicit definitions is acceptable then Fregean abstractions (that satisfy certain further constraints – the most obvious one being consistency) are often likely candidates for acceptable principles.

The relevant point for my discussion is the role of abstractions as implicit definitions. The key passage for Wright's framing of the issue goes as follows:

> "[HP] is inadequate as a definition in the purest sense since it fails to provide the eliminative facility which any genuine definition must... Nevertheless statements containing ineliminable occurrences of arithmetical vocabulary can still be apprehended as logical truths if they can be shown to follow logically from statements which do possess purely logical transcriptions and which are theorems of higher-order logic."[67]

65 Hale & Wright (2001), p.14
66 Loc. cit.
67 Wright (1983), p.139

He notes further that what is lacking is an 'explanation in purely logical terms' of those ineleminable occurrences of arithmetical vocabulary. He requires a principle that gives, via Frege's Theorem, a guarantee that arithmetic is “nothing other than a logical consequence of the very explanation of cardinal number itself.”[68] The idea, then, is to use HP as something akin to an informative definition of some notion that is already in our language.

At this point it should be noted that Wright does not always seem to have this kind of aim in mind. For instance, he writes that

> “An abstraction principle is not an attempt to state an antecedently determinate truth. It is intended, rather, to fix a new concept by determining the truth-conditions of certain statements involving it.”[69]

On this kind of reading of abstractions, they come out true by stipulation. As such, there is nothing in principle to constrain our stipulations; but this is clearly in tension with the idea of *explaining* the notion of cardinal number. If, as Wright periodically professes, we are looking to give an explanation of cardinal number, there must be some phenomenon or concept that requires explanation and that can only be our current concept of cardinality. As such we are constrained by our actual concept of cardinality in giving truth conditions for a statement governing identity of cardinal numbers. This point is implicit in the debate between Wright and Boolos about the analyticity of HP where Boolos quotes Wright as pointing the inadequacy of calling a principle analytic when “there is no *prior*, no intuitively entrenched notion, no notion given independently, which [our candidate principle] is analytic of.”[70]

Thus Wright seems to have in mind two distinct aims that he does not always manage to hold apart.

68 Ibid., p.154
69 Wright (1997), p.281
70 Boolos (1997), p.312

As indicated both above and in the introduction, it seems as though only the latter, explanatory, target can be the appropriate one – Wright is not free to explain our knowledge of arithmetic and our referential success in our use of arithmetical terms by hijacking our concept of cardinality and making it fit his new, stipulated, principle. Instead any principle that is supposed to define our concept of cardinal number must conform to the use of that very concept. As such, we should not read Wright as attempting to introduce, by way of a stipulative definition of a new term, a new notion into our language. Instead he must be seen as attempting, in some sense, to give a *retrospective* definition of cardinal number – a sharpening of a notion that is already well-entrenched. It seems that this captures both the explanatory aspect and the definitional aspect of what Wright requires of HP.

Where, then, is the link with analyticity? If such a retrospective definition is to count as analytic, Wright needs to be more explicit about what exactly is encompassed by analyticity. He concedes that on a Fregean construal of analyticity, HP will not come out as analytic. This is because it is not a truth of logic nor is it an explicit, eliminative definition. However, he points out that this construal is too narrow, as it excludes prima facie analytic truths (such as "whatever is yellow is extended"); what he requires is a notion that is Fregean in spirit but somewhat broader. He claims that such a notion exists, and moreover is the appropriate one by which HP should be judged. He argues by analogy to that notion governing the analyticity of statements that link the logical connectives with their rules of inference. For instance, he takes it that "the meanings of the logical constants should be regarded as implicitly defined by the stipulation of the usual rules for their introduction and elimination in inferential contexts."[71] He calls such stipulations *determinative* of the concepts that they are supposed to explain, and notes that abstractions such as HP could count as analytic on a similar model.

It is worth distinguishing two claims that the above reconstruction of Wright brings out. The first is

71 Hale & Wright (2000), pp.117-8. See also Hale & Wright (2001), pp.12-4.

a conjunction of two conditional claims; if HP is true, then it is analytically true and if it is false, it is analytically false. The second is that HP is true (and hence analytically true). I will call statements that if true are analytically true and if false are analytically false *quasi-analytic*. Both of Wright's claims here are doubted by Boolos[72], who argues that on natural construals of analytic statements there is insufficient reason to think that HP could count as one of them. He further argues that if some other notion of analyticity were right, then HP would still be inadequate as it is in fact false. Thus HP is not quasi-analytic nor is it true, whilst the reason for thinking that it is analytic is that there are (analytic) truths closely related to it.

My argument will be related but will have a different focus inasmuch as I grant the quasi-analyticity of HP but demonstrate that it is false. I then consider what would be involved in any attempt to exorcise the counterexamples to HP to yield an abstraction HP* – a restricted version of HP that is true – and ask whether any such principle could be analytic. My contention is that no restriction of HP adequate for Wright's purposes can be quasi-analytic, and hence Wright's proposal cannot be easily fixed.

The foregoing shows that I seek to establish a slightly different conclusion to that of Boolos, as he does not accept even the quasi-analyticity of HP. I grant that it is quasi-analytic but show that none of its relevant restrictions can be. However, my argument is not just a dialectical one – it is not simply asking for an account of analyticity and an example of HP* before showing that they do not combine in the required way. My point is that *any* reasonable construal of analyticity on which abstractions could count as analytic and *any* possible HP* will not mesh together. Having said this, such a contention is only relevant if HP is quasi-analytic *and false*. We have quasi-analyticity by hypothesis, and so I now turn to a discussion of purported counterexamples to HP.

72 Boolos (1997)

§2: Counterexamples to Hume's Principle

Whilst the Russellian reasoning that makes Frege's Basic Law V unacceptable is well-known, there is a strong inclination to think (as Frege and Wright did) that HP is true as well as consistent. However, HP does not seem to hold for all concepts (or even all concepts for which it is 'supposed to' hold). The purported counterexamples to HP are of the form of a counterexample to its consequence that all concepts have a unique cardinal number. The concepts that give rise to such cases can be split into three categories. The first are vague predicates, a class that for current purposes I can and will treat as contained within the second class of semantically paradoxical predicates. The third are indefinitely extensible concepts, a notion to be explicated later.

§2.1: Paradoxical Predicates

I will take a predicate to be vague only if it has fuzzy borders of extension, it is subject to borderline cases, and it is vulnerable to Sorites-like reasoning[73]. The relevant consideration here is that of there being no clearly-defined extension; if it is indeterminate whether or not something is F, then F will have an indeterminate extension. As such, there will be no determinate unique cardinal number that is the number of Fs. I will not attempt to treat such predicates separately; I will instead use the fact that they give rise to the Sorites Paradox[74] in order to put vague predicates in the same category as those that lead to alternative semantic paradoxes as outlined below.

73 All of these conditions should be understood as qualified by 'prima facie'.

74 It is of course possible to query this assumption. Two obvious ways to do so would be to reject classical two-valued logic or to endorse something like Williamson's epistemism. However, to go down the former route would be completely antithetical to a Fregean philosophy of mathematics. An assessment of the latter lies beyond the scope of this thesis; but it is clear that if it is a commitment of a neo-Fregean treatment of arithmetic to be unable to remain neutral on the correct treatment of vague predicates, this is to the neo-Fregean's detriment.

A well-known example of a predicate that gives rise to a semantic paradox is 'is a heterological adjective of English*', where English* is English with the word 'heterological' appended to it. A word is heterological iff it does not truly describe itself – for example, 'monosyllabic' is heterological as the word itself is not monosyllabic. The problem arises when one considers whether or not 'heterological' is a heterological adjective of English*. If it is, then it is heterological and hence does not truly describe itself. Thus heterological is not heterological. But if heterological is not a heterological adjective of English*, and hence does not truly describe itself, then it is heterological – a contradiction. Thus 'is a heterological adjective of English*' cannot have a definite extension containing all and only those adjectives that are heterological, and thus such a concept certainly cannot have a unique cardinal number.

There is a comparable argument using 'heterological' to construct disjunctive predicates. Take any controversial proposition (Wright's[75] example is "the universe is finite"), and take the predicate 'is heterological or "the universe is finite" is true'. By parallel reasoning, considerations about vocabulary leads to a proof that the universe is finite (as accepting the other disjunct leads to contradiction). If such armchair astronomy is to be ruled out, we must find a way to exorcise such predicates.

I want to suggest that the moral to draw from the above cases is not that HP is defective, but that our choice of concept with which to test (so to speak) HP is not a good one. Another way of looking at the alleged difficulty is to say that we take HP and 'plug in' various predicates, whereupon we discover that we get apparently false instantiations. In such a case, however, there are two possible diagnoses: that there is something wrong with the principle, or that there is something wrong with the predicates. Here there is a great temptation to say the latter. Something that might indicate that this is the way to go is that these predicates give rise to semantic paradoxes when used in more familiar settings – for example, using disjunctive predicates to generate a counterexample to HP is a

75 Wright (1997), p.286

close analogy of Curry's paradox. There is no readily available and uncontroversial way to resolve these semantic paradoxes, and as such we should not automatically take the fact that our system has not resolved these paradoxes – problems that our system is not designed to solve – to count against it. The objection is that adopting this method of concept formation leads to difficulties when considering a class of predicates that give rise to semantic paradoxes. The response is that this fact can only be used to show that abstraction principles are no better than our current, well-entrenched methods. Unless our current methods are held to be themselves defective, it does not follow just from these paradoxical or vague predicates that our abstraction is similarly defective.

This kind of argument seems to be what Wright is getting at when he writes that

> "We should not, presumably, conclude that the whole idea that properties can be defined by the stipulation of satisfaction conditions is misbegotten. Rather, some kind of restriction is warranted. Nor, crucially, pending such a restriction, should we suspend judgement about what appears to be perfectly innocent examples of such a procedure. The sensible response is rather that there is a distinction to be drawn which we are, perhaps, not clear how exactly to draw but which, once drawn, will safeguard the vast majority of cases where we fix a property by stipulating satisfaction conditions."[76]

However, it is not clear that it is sufficient to endorse the conditional: *if* HP is indeed a "perfectly innocent example", *then* what Wright says seems right. The point is that we have some suspicious predicates falsifying our principle; Wright's claim is only a defence of the *kind* of principle involved rather than the individual principle itself.

However, if Frege's Theorem is going to play the role of transmitting analyticity across second-

76 Ibid, p.288

order consequence, it had better be based on a *true* abstraction. Thus whilst Wright claims that drawing the relevant distinction is solely “in the interest of clarity and... intellectual good housekeeping”[77], it is not clear that this is quite fair to his objector. There is work to be done here, and it is perhaps rather more urgent than Wright would have us believe; it is also, however, somewhat less urgent than first appearances might suggest.

§2.2: Indefinitely Extensible Predicates

A different category of predicates include more interesting attempts to show that HP is false. Take the concept 'is self-identical'. HP implies that there is such a number as #[x: x=x], something Wright calls “anti-zero.” By definition of ≤, anti-zero is the number that is greater than any other number. The worry now is whether or not there is such a number. Both intuitively and according to ZF set theory, there is no such cardinal of all the things (or even sets) that there are. Thus Frege arithmetic is incompatible with ZF and, moreover, if HP is analytically true then ZF is analytically false.

There is a similar line of reasoning that works with respect to the concept 'set' or 'ordinal' – those that are too big. These are, in Dummett's terminology, *indefinitely extensible*. The extensions of such concepts are characterised by Wright as follows:

> “a totality of such a sort that any attempt to view it as a determinate collection of objects will merely subserve the specification of new objects which ought, intuitively, to lie within the totality but cannot, on pain of contradiction, be supposed to do so.”[78,79]

77 Loc. cit.
78 Wright (1998), p.316
79 This is importantly different from Dummett's own characterisation. He writes that “No definite totality comprises everything intuitively recognisable as an ordinal number, where a definite totality is one quantification over which

To take the case of 'ordinal', we obtain a counterexample to HP because to allow that there is such a thing as the number of all ordinals is incompatible with the conclusion of the Burali-Forti Paradox – that there is no set of all ordinals. Similarly, by Cantor's Paradox, there can be no number of all sets as there is no set of all sets. Thus instances of HP of such indefinitely extensible concepts gives a result that is incompatible with any acceptable set theories; as such, we have a class of concepts that give counterexamples to HP.

Wright notes that this need not be a reason to think that HP is worthless as an implicit definition. Rather than a wholesale rejection of HP, what is needed is a principled restriction to rule out these problem cases. These counterexamples show not that HP is not definitional of the concept of cardinal number, but that it is only a partial definition. Moreover, we can bolster the case for HP by noting that it is in some sense exceptional instances that seem to give rise to problems. However, if we accept that HP is not a complete definition, there are no longer any grounds for saying that the whole of arithmetic can be interpreted in an analytic theory.[80] I will return to this point later, but for now it is enough to consider that if HP must be restricted in an ad hoc way in order to give rise to a true theory in which arithmetic can be interpreted, it is going to come out as non-analytic.

The key qualifier, of course, in the above rendering of the objection to Wright's position is "ad hoc." Wright can accept the objection as it stands, but deny that a restriction of HP (HP*) need be ad hoc. For example, it does not seem ad hoc to restrict the domain of concepts to only *sortal* concepts. Wright explicates the notion thus:

always yields a statement determinately true or false. For a totality to be definite in this sense, we must have a clear grasp of what it comprises: but, if we have a clear grasp of any totality of ordinals, we thereby have a conception of what is intuitively an ordinal number greater than any member of that totality. Any definite totality of ordinals must therefore be so circumscribed as to forswear comprehensiveness, renouncing any claim to cover all that we might intuitively recognise as being an ordinal." (1991, p.316). On this understanding, it is open to regard totalities such as the natural numbers as indefinitely extensible, even though quantification over the natural numbers need not lead to contradiction. As such, Dummett's conception is broader than Wright's but it is Wright's that contains the relevant problem cases.

80 Whilst Boolos believes that even if HP were true it would not yield the desired result, this is not necessary for the point to go through.

> “[T]he usual intuitive understanding is that a sortal concept is one associated both with a criterion of application – a distinction between the things to which it applies and those to which it doesn't – *and* a criterion of identity: some principle determining the truth-values of contexts of the form “X is *the same F* as Y.”... [I]n general, purely qualitative predicates, predicates of constitution, and attributive adjectives – although syntactically admissible substituends for occurrences of the predicate letters in higher-order logic – are not [sortal concepts].[81]”

On such a rendering of sortal concepts, Wright provides an ingenious test as to whether or not a concept should count as sortal. He observes that non-sortal concepts, when conjoined with a sortal concept, give a conjunctive sortal concept. As such, if a non-sortal concept is conjoined with a concept of which it is not known whether or not it is sortal, it will be sortal only if the conjunctive concept is. For example, 'is brown' is non-sortal whilst 'is a horse' is sortal – as such, 'is a brown horse' is sortal. By these lights, 'is self-identical' is non-sortal; for any non-sortal concept, such as 'brown', the conjunctive concept 'brown and self-identical' is equivalent to 'brown' which, *ex hypothesi*, is non-sortal. Thus 'is self-identical' is non-sortal, and hence not admissible to the domain of concepts to which HP applies.

There are two issues relating to Wright's characterisation and treatment of sortal concepts that I wish to bracket. Firstly, it is not at all clear whether or not Wright's rendering of a sortal concept coheres with his claim that HP is supposed to characterise a sortal concept of 'natural number'. Secondly, why should HP need to have a sortal concept as both input and output, so to speak? Wright is attempting to develop Frege's view that numbers are numbers of concepts and is clearly claiming that numbers are numbers of *sortal* concepts, where “a concept is sortal if to instantiate it is to exemplify a certain general kind of object … which the world contains.”[82] Indeed, the bulk of his argument is to say that 'natural number' is one such sortal concept. It seems slightly strange and

81 Wright (1998), p.315
82 Wright (1983), p.2

restrictive to say that our principle of numerical identity should apply only to *kinds* of things, as the implication is that cardinality is a property only of such kinds of things. Nevertheless, it seems that we can grant that Wright has a coherent, consistent and plausible view of sortal concepts without our argument losing any of its bite.

The key point here is that Wright's appeal to such a view of sortal concepts as part of his argument to exorcise anti-zero simply does not work for all indefinitely extensible concepts. The concept of an ordinal, for instance, should surely count as a sortal concept. We have a clear criterion of application – x is an ordinal iff x is a transitive set with a well-ordering membership relation – and a clear criterion of identity – x is the same ordinal as y iff x and y are ordinals and there is an isomorphism between x and y. If one wishes to exclude it as non-sortal because there is no number of all ordinals, by the Burali-Forti Paradox, this is to beg a question that ought not to be begged. It seems clear that 'ordinal' has hallmarks of a sortal concept – if it is not to count as sortal, then that fact must be explained. My contention will be that once one undertakes such an explanatory project, the kind of justification of excluding such predicates from the domain over which HP admits quantification is going to invalidate claims that HP could count as a conceptual truth. I will return to this issue at a later stage, noting for now only that we have prima facie counterexamples and that there are alternative ways of attempting to resolve the difficulties raised by such examples.

§3: Status of Restrictions of Hume's Principle

A line of argument suggesting that HP cannot be analytic just because it is false can have any real power only if it is possible to demonstrate why some kind of restriction of an abstraction to give a true principle is not going to preserve pseudo-analyticity (and hence be analytic). This seems at first to be fairly plausible. In order to see this clearly, it is helpful to summarise the state of play so far.

We are granting that analyticity transmits across second-order logical consequence as demanded by (2). We are also conceding that it is an epistemological notion, as otherwise the conclusion that arithmetic is analytic does not help us resolve Benacerraf's challenge. Wright's substantive claim is that HP is analytic on such an understanding, but it turns out that HP is false if unrestricted. This suggests that the next step ought to be to refine HP to find a principled, non-ad hoc restriction such that it is true. But on this kind of approach it appears that HP is treated more like an empirical, inductive generalisation than any kind of conceptual truth. We refine the domain over which HP admits quantification in order to exclude problematic cases, but this is not going to help us to simply 'see' that there are no counterexamples just from understanding what is said by any candidate HP*. This seems to be a minimal requirement of an analytic truth, and hence HP* will not be analytic.

The response is to say that whilst our *method* of refining HP might resemble a refinement of an inductive generalisation, that resemblance alone is not sufficient to yield the result that HP* is non-analytic. What we are doing in making such refinements is paralleled by refinements in our understanding of the HP. It is not that we are changing what HP is saying; instead we are making it clear what the proper content of HP is. This kind of claim is supposed to be independently plausible on the grounds that such an explanation works in allegedly analogous cases. For example, take a general statement that is false, such as 'all men are unmarried.' We now alter this to 'all bachelors are unmarried' which gives an analytically true statement. Thus we have an example of a change in truth value that arises from a restriction of a general statement, and the latter (restricted) statement is analytic. Why should the same not be true of HP and its restrictions?

The first thing to point out at this juncture is that my argument only need work for abstractions – necessary equivalences that have a certain syntactic form. Indeed, it could be even less general than this, as I require only that it is a feature of HP to prevent Wright's argument from going through. I will therefore use either a general abstraction or HP, depending on the generality of my point at each stage of the argument and for the purposes of clarity. However, the thought here is that what is true of statements in general need not hold for abstractions, as abstractions have certain features that may allow for restrictions only at some epistemological cost.

Despite this, there is a deeper issue here that is worth bringing out. Something that the 'all men are married' vs 'all bachelors are unmarried' example does not make explicit is *how* the restriction occurs. There are two options in such a case; either we change the antecedent of the conditional, or we restrict the domain of quantification. This can be made clear by the following formalisations:

'All men are unmarried' iff $\forall x\ (Mx \rightarrow Ux)$

'All bachelors are unmarried' iff $\forall x\ (Mx \rightarrow Ux)$, where $\forall$ ranges over all and only bachelors

'All bachelors are unmarried' iff $\forall x\ (Bx \rightarrow Ux)$

Thus either we change our domain such that we exclude potential counterexamples, or we make the antecedent of our conditional more demanding. There is a similar taxonomy of restrictions that can be used in the case of abstractions. Recall our general abstraction:

$$\forall F \forall G\ [\Phi F = \Phi G \leftrightarrow (F \approx G)]$$

We have a number of different options here, each of which I will discuss in turn. One might constrain the domain of quantification, change the abstractive relation in such a way as to allow the RHS to be satisfied by some other state of affairs, alter the abstraction so that it is the consequent of some conditional with an existential clause in the antecedent, or bring in additional theoretical commitments about our treatment of indeterminacy. A fifth and final option that can be quickly dismissed is to construe the functional expression as non-total; however, treating the functional operator on the LHS as non-total is impossible for Wright, as the proof of Frege's Theorem depends on the use of HP as a total function to guarantee that every natural number has a successor. As such, we cannot have grounds that derive from the status of HP for thinking arithmetic analytic if HP is not interpreted as a total function.

Of our four options to be considered below, we have it that the first option is a way to exclude the counterexamples whilst the final three are ways to allow for them. I will argue that whilst each way has its own more specific defects, there is a common flaw running through each; the kinds of extra theoretical commitments incurred in changing HP (a false general principle) to HP* (a true general principle) are not conducive to allowing HP* to be analytic.

§3.1: Domain Restriction

Our first option, then, is to restrict the domain to exclude certain concepts. We can understand this approach by way of an illustration; let us consider the relationship between HP and Cantor's Principle (CP). CP can be seen as a restricted form of HP in which the domain of quantification is restricted. CP states that

(CP): $\forall F \forall G\ [\#F = \#G \leftrightarrow \exists f\!: F \leftrightarrow G$ (i.e. f is a bijection between F and G)]

Where (crucially) F and G are *sets*. CP therefore states that cardinality is a property of sets, rather than of concepts. Thus the domain of quantification is restricted to a subset of that required by the unrestricted HP.

Leaving aside strict nominalist views, we can take it that CP is uncontroversially true; neither the semantic nor class paradoxes pose any problems for CP. As such CP is a truth of the same form as and in the same neck of the woods as HP. However, my point is that the logicist cannot avail himself of CP rather than the more controversial HP as to do so would undermine the epistemological aims of his programme. This is because the abstraction is not a purely logical abstraction; the explanation of the # operator is given in terms of relations between *sets*, rather than the intentionally individuated Fregean *concepts*. Thus if we take CP rather than HP as our foundational principle, it must be that an understanding of the notion of a set is epistemically prior to any understanding of arithmetic. As such the relevant variant of Wright's premise (2) secures only that arithmetic has the same epistemic status as that of set theory.

However, there is reason to think that CP is not *analytically* true. There are two aspects to this worry, the first of which is dialectic. Absent some independent reason to think that CP is non-analytic, one way to assess it would be to consider whether its logical consequences are analytic; if it has clearly non-analytic consequences, then by Wright's premise (2) the original principle had better not count as analytic. We should take it that reasoning using modus tollens gives us a more accurate assessment of analyticity than using modus ponens – to use the latter is to beg the question against the opponent of the analyticity of CP.[83]

But are there any independent reasons to think either that CP is analytic or that any of its consequences are not? It seems to me that there are two important points here. The first is that we lack a clear criterion of application for the predicate 'is a set', and it is a question that is only partially settled by adopting particular set theoretic axioms. This is going to be enough to render CP non-analytic. One reason that we can take 'bachelors are unmarried men' to be analytic is that there is no further question about the criteria of identity or of application of 'bachelor' once the relevant criteria are settled for 'man' and 'unmarried'.[84] In contrast this is not the case with respect to 'set'; the axiom of extensionality settles the question of identity, but the question of application remains open. One of the tasks of set theory is to attempt to settle this question, but that is part of set theory rather

83 This kind of reasoning parallels Boolos (1997), p.308
84 Here I am being deliberately ambiguous between a metaphysical and epistemological reading of 'settled'; the current point does not depend on which reading is thought to be appropriate.

than any part of logic or conceptual analysis. As such, the logicist cannot establish the analyticity of arithmetic by this kind of modification of Wright's argument.

The second issue is that Wright, following Frege, is seeking to establish a version of *logicism* about arithmetic that depends on treating cardinality as a property of concepts and not of sets. If cardinality were a property of sets (and could be taken to be a property of concepts only because of some connection between sets and the extension of those concepts), then it seems as though it could no longer be analytic unless one takes set theory more generally to be analytic. But to take set theory to be analytic seems to beg a number of important questions, and would entail a conclusion far stronger than the (already slightly implausibly strong) neo-logicism under scrutiny. Thus to treat cardinality as a property of sets may or may not be correct, but it is not an option for the neo-logicist.

To return to HP, there is already one relevant restriction place – a restriction to sortal concepts – but the question is whether this can be taken further in a principled way. One possibility is to do so with respect to indefinitely extensible concepts that, as noted above, are putative counterexamples to HP. It seems as though this is what Wright is tentatively suggesting when he writes

> "[i]f there's anything at all in the notion of an indefinitely extensible totality … one principled restriction on HP will surely be that F and G *not* be associated with such totalities."[85]

And, later:

> "*The Restriction Problem* [is] how best to restrict the range of concepts to which the cardinality operator is to apply if the generation is to be avoided of prima facie

85 Wright (1998), p.316

'rogue' numbers like anti-zero, ... the number of sets, the number of ordinals, and so on. It seems reasonable to expect that a well-motivated restriction will have to build on characterisations both of *sortality* and *indefinite extensibility*."[86]

In the former quotation there is no argument offered for the key qualifications "surely" and "principled." In the latter, he moots only that it seems reasonable that such qualifications will apply to restricted forms of HP. However, whilst we do not want to dispute his claim that only sortal concepts are appropriate for the domain of HP, there is nothing in what Wright says to persuade us that indefinitely extensible concepts should be treated in the same way.

Having said this, it seems to me that we can go further; we may object not just to Wright's treatment of indefinitely extensible concepts but more generally to the kind of illegitimate concept-mongering in play. Recall from §1 that Wright proposes two possible ways in which abstractions could be analytic; either analytically true by stipulation or analytically true by correctly characterising an antecedently understood concept. We saw that it is the latter that is the proper target for an account of the analyticity of an abstraction. However, even if we were to say that abstractions could be used "to fix a new concept by determining the truth-conditions of certain statements involving it"[87], the kinds of principles with a restricted domain that have been under scrutiny would not obviously count as analytic.

It seems as though in order to "fix a new concept," we take our domain and come up with some suitable equivalence relation on that domain. We then introduce an expression for our new concept into the LHS of our abstraction and state what it must be for the referents of the terms of the LHS to be the same, stated in terms of a biconditional and our equivalence relation. In the case of DE, then, we take as our domain 'lines'[88], we take an equivalence relation 'is parallel to', and introduce our

86 Hale &Wright (2001), p.427
87 Wright (1997), p.281
88 Here I ignore any issues as to whether only actual or possible inscriptions are permissible.

new functional operator 'the direction of', which has as its output the abstract objects 'directions'. If this definition is shown to be faulty in some way, it is surely not adequate to change the domain; what we are trying to do is define a concept on that particular domain, not on some other domain. If such a restriction were permissible, it seems that we could at best get a partial definition of our functional expression.

In the case of HP, it seems as though it is precisely this kind of situation that confronts us. The class of indefinitely extensible predicates cannot, without strong supplementary arguments, be excluded from the domain without threatening the comprehensiveness of our attempted definition. But if HP is only a partial definition of the cardinality operator, then (as Wright notes[89]) we cannot guarantee the kind of understanding of arithmetic that the neo-logicist programme requires. Thus more is required from the neo-logicist if these counterexamples are to be excluded in some principled way.

§3.2: Changes to Abstractive Relations

Let us move now to our second possibility – changing the abstractive relation. The main way in which one might do this is to make the equivalence relation disjunctive. A good example of this is Boolos' consistent variant of BLV, which he calls VE ('V enlightened'):

$$\text{(VE): } \forall F \forall G\, [\{x{:}Fx\} = \{x{:}Gx\} \leftrightarrow [(F \text{ is too big } \& \ G \text{ is too big}) \vee \forall x\, (Fx \leftrightarrow Gx)]]$$

Now the reasoning that leads to Russell's paradox no longer gives rise to a contradiction, but merely a proof that the relevant concepts are 'too big', where a concept is too big iff its extension is equinumerous with the universe.

89 Ibid., p.279

Wright[90] and Boolos[91] briefly attempt to develop a system that they dub SOLVE – Second-Order-Logic and VE. Wright, for one, is hesitant to endorse this kind of approach as it is not automatic that these sorts of principles give as strong a mathematical result as Frege's Theorem. SOLVE is not as powerful a system as FA^2 and is not equi-interpretable with it. However, systems based on such abstractions could, if those abstractions were to count as analytic, still be used to vindicate some sort of partial logicism – at least some important aspects of number theory could conceivably come out as analytic. As such, an investigation into the epistemic status of these principles is worthwhile.

There are two worries with taking these refined abstractions to be analytic. The first is to note that this approach does not seem to exorcise the semantically paradoxical counterexamples but only those that generate class paradoxes. However, this objection need not be decisive – we only have one example of an altered abstraction above, and for all that has been said so far there is nothing *in principle* to prevent us finding a suitable equivalence relation that can generate an abstraction that is not subject to the same kind of counterexample. I will not press this further, as it seems to me that there is a deeper worry for Wright or Boolos.

The initial attraction of this line of response is supposed to be that we are able to include, rather than exclude, potential counterexamples. We have a general principle that, on reflection, appears to be not as general as we took it to be. We have two concepts linked together – identity of cardinal number and equinumerosity – and have found that we cannot say that this link holds universally. The proposed solution of changing the abstraction relation is to say that it is not *those* two concepts that are linked, but two related ones; identity of cardinal number and some disjunctive concept. As such we can keep the domain of quantification the same, but show that what holds true in that domain with respect to our troublesome functional operator is not what we thought.

According to such a response, then, the moral to be drawn from the problematic cases of

90 Ibid, p.302
91 Boolos (1987)

indefinitely extensible concepts is that such concepts do in fact have cardinal numbers belonging to them – it is just that we need a broader notion of cardinality. This broader notion can be given by some appropriate (if somewhat ad hoc) disjunction. But this approach seems misguided. This is surely completely the wrong moral – we have it that whilst there is still no set of all sets, the concept 'is a set' has a cardinal number by default. But it is completely obscure what such a cardinal number could be, let alone how it behaves relative to any system of cardinal arithmetic. We do not want to say that we can put all these rogue concepts under one cardinal number that stands alone, independent of all other cardinal numbers – we instead want to say that such concepts do not have a cardinal number. If one wishes to introduce a new notion – call it 'is the schmardinal number of' – which is introduced by HP, then that is another matter. But clearly a concession that a new notion is needed is no good to the logicist – we have no interest in attempting to show that schmardinal arithmetic might be analytic. Thus this attempt to allow potential counterexamples is wrongheaded.

§3.3: Conditionalised Abstractions

Recall that the idea here is that the bracketed expression '$[\Phi F = \Phi G \leftrightarrow (F \approx G)]$' could be the consequent of a conditional with an existential statement as the antecedent. What we are trying to do is to change the form of the abstraction whilst maintaining as much of its spirit as possible. One point behind this is that the abstraction would be satisfied vacuously if there is referential failure of the terms of the LHS, but the antecedent could contain more than simply a guarantee of the right kind of objects; for instance, taking the example of HP, we could attempt to restrict it to say that:

$$\forall F \forall G\ [(\exists A \exists B\ \forall x\ (x \in A \text{ iff } Fx)\ \&\ \forall x\ (x \in B \text{ iff } Gx)\ \&\ A, B \text{ sets}) \rightarrow [\#F = \#G \leftrightarrow (F \approx G)]].$$

The content of such a principle would be to say that if the extensions of F and G form sets, then they have the same cardinal number iff they are equinumerous. Such a principle would be strongly reminiscent of CP, as it tells us about cardinality only inasmuch as cardinality is a property of sets; unlike CP, however, it leaves open the possibility that something other than a set can have cardinality. A similar proposal is considered in passing by Boolos. He writes that

> "I am moved to suggest... that the conditional whose consequent is HP and whose antecedent is its existential quantification might be regarded as analytic. The conditional will hold, by falsity of antecedent, in all finite domains. By the axiom of choice, the antecedent will be true in all infinite domains, but then, we may suppose, nothing will prevent the consequent from being true."[92]

However, there are two worries here. The first is that we no longer have an abstraction, and as such we cannot assume that it will be able to bear the same epistemological weight as an abstraction. Indeed, there is something like this going on in Wright – he takes it that abstractions can play the role of implicit definitions of some kind, and his arguments to this effect need not carry over to these conditionalised abstractions. It seems that this is Boolos' point – even if such principles could be made out to be analytic, this is going to be insufficient for the kind of programme envisaged by Wright.

The second difficulty is more straightforward and, given that it mirrors the discussion of CP in §3.1, it is perhaps more obviously decisive. In order to ascertain the truth of these principles, we require a certain amount of set theory – in my proposal it is just a criterion of application for 'is a set', in Boolos' it is such a criterion that includes the axiom of choice. Again, if we need it to be the case that some aspects of set theory that include the axiom of choice are to be analytic in order for logicism to have a chance, then prospects are poor. The point is not that there is anything controversial about the set theory involved, but that it cannot be used by the neo-logicist to establish

92 Boolos (1997), pp.306-7

the analyticity of any candidate principle.

There are two more promising alternatives for the neo-logicist, however. I will give a relatively brief assessment here, as I deal with each more fully in Chapter 3; despite this, there are a number of points worth making at this juncture. The first is what Wright and Hale call the 'inverse-Carnap conditional'[93] of HP, and can be thought of best as a development of Boolos' tentative suggestion given above. The idea is to reject the more usual kind of conditional used in giving implicit definition of (scientific) terms 'if anything satisfies HP, then numbers do' in favour of the (allegedly) more fruitful 'if there are such things as numbers, then their criterion of identity is given by HP'. More formally, this inverse-Carnap proposal can be given as follows:

(ICHP): $\forall F \forall G \forall u \forall v\ [(u = \#F\ \&\ v = \#G) \rightarrow (u = v \leftrightarrow F \approx_{1\text{-}1} G)]$

However, whilst this proposal is clearly more tentative than HP, it is perhaps too tentative. A claim of the form 'if there are such things as numbers of concepts, then any two such numbers are equal just when their concepts are equinumerous' is going to be necessarily true – perhaps even conceptually necessary – but that does not give the epistemological result that we need. The point is that whilst we might be able to know that ICHP is true, we do not know *why* it is true – by falsity of antecedent or by satisfaction of both antecedent and consequent. It justifies our talking of numbers in the way that we do, but not our mathematical knowledge. Thus such a principle, without some further *analytic* guarantee that the antecedent will be satisfied, does not help the neo-logicist.

The second, more interesting option is to adopt a principle given first by Heck[94] that he calls 'Finite Hume's Principle':

(FHP): $\forall F \forall G\ [(F \text{ finite} \vee G \text{ finite}) \rightarrow (\#F = \#G \leftrightarrow F \approx_{1\text{-}1} G)]$

93 Hale & Wright (2000)
94 Heck (1997)

FHP states that finite concepts have the same number just if they are equinumerous and has as a trivial consequence that no finite concept has the same number as an infinite concept; however, no more is said about the numbers of infinite concepts. As such, it is immune to the kinds of counterexamples discussed in §2. There are also two technical points to make; firstly, finititude can be defined within the system of second-order logic used by Frege and Wright. Secondly, Heck gives a slightly surprising strengthening of Frege's Theorem: that second-order arithmetic is also equi-interpretable with a system given by appending FHP to second-order logic. Thus it is worth taking the time to consider whether or not FHP could be analytic.

Heck attempts to argue that there is some plausibility in taking FHP to be the principle that we use in actual arithmetical reasoning. Whilst this hermeneutic claim seems somewhat far-fetched, it is importantly beside the point – what we need to address at this point is whether it can be both analytic of our concept of cardinal number and be used to explain the arithmetical knowledge that we in fact have. I will argue that FHP falls at the first hurdle but not obviously at the second; if the requirement of analyticity can be removed, then FHP is as good a candidate as any other of those available to the neo-logicist to explain our arithmetical knowledge. I will discuss only the former claim at this point, leaving the latter to Chapter 3.

What reasons do we have for taking FHP to be non-analytic? Recall Boolos' point from §1; that it is no good to appeal to a principle if "there is no *prior*, no intuitively entrenched notion, no notion given independently, which [our candidate principle] is analytic of."[95]. It seems to me that FHP fails on this account. A satisfactory account of *cardinality* must tell us how infinite cardinals (or at least small infinite cardinals) behave – otherwise it is not cardinal number of which the principle is analytic but some less refined (pre-Cantorian, perhaps) notion. Whilst on most other approaches (in particular, the set-theoretic and type-theoretic – see below – variants of HP*) it has been possible to argue as required that the notion of cardinality of which it could analytic is the right one, that seems

95 Boolos (1997), p.312

impossible in this case. Thus there is an interesting difference between FHP and other candidates for HP* – FHP in some sense says too little to be analytic. I will return to whether or not FHP can nevertheless underwrite our understanding of *elementary number theory*, even if not of *cardinality* more generally, in the next chapter.

§3.4: Indeterminacy

The final option I will consider is that of a treatment of indeterminacy. In order to see how such a treatment can be used to exclude indefinitely extensible and vague predicates, it is helpful to consider the rationale behind their being a counterexample. The reasoning that leads to $\forall F \exists! x\ (\#F = x)$ is to let F=G in HP, and obtain that the number of Fs is the same as the number of Fs iff the things that are F can be put into 1-1 correspondence with themselves. The RHS of the biconditional is trivially true – identity provides such a 1-1 relation – but the LHS is not true when '#F' has no referent. One time that this referential failure occurs is when F is non-sortal, but another is when #F is indeterminate. The thought, then, is that we can ensure that such a failure of reference does not occur by appealing to an account of how one treats indeterminacy.

One kind of account would be supervaluationist in spirit. The treatment of predicates with no determinate number because of their having no determinate size of extension would be that on any way of making such predicates precise, they will have a determinate extension and number. The clearest example amongst the indefinitely extensible predicates is 'set'; on the iterative conception[96], whereby one can always form more sets by going up a level in the cumulative hierarchy of sets, there is a determinate number of sets for all sets up to and including those of a given level of the hierarchy. The thought is that on any way of making the concept 'set' precise – something that is

96 See Boolos (1971) for a clear exposition of this.

identified with fixing a level in the hierarchy and taking the domain of discourse to be all sets formed up to this point – there will be a determinate number of sets. If something like this is right, indefinitely extensible concepts more generally and vague predicates need not give rise to counterexamples to HP. Thus whilst we would still need to find a way to solve the problems raised by 'heterological' and its semantically paradoxical cousins, we would have a partial patch to HP.

However, whilst this seems initially promising, difficulties arise when one attempts to put flesh on the bones. If the thought is to say that for any level of the cumulative hierarchy of sets, there will be a determinate number of sets, then our new HP* should read as follows:

$$\forall L \forall F \forall G\ [\#(F, L) = \#(G, L) \leftrightarrow (F \approx_L G)].$$

But at this juncture there is the same dialectical response as in the case of the previous appeals to set theory. One can perfectly well agree with Boolos that referencing levels of the set theoretic hierarchy is "entirely natural, free from artificiality, not at all ad hoc"[97] without being forced to acknowledge that a principle making essential use of this notion of a set can count as analytic. Again, we cannot take parts of set theory to be analytic in our attempts to render any candidate HP* analytically true.

Another way would be to introduce notions peculiar to type theory rather than set theory. The idea here would be to relativise the '#' function to whatever type theoretic notion plays the role of a given level of the set theoretic hierarchy. Thus we have a succession of $\#_L$ functions that correspond to and depend on each way of making 'set' or 'ordinal' determinate. The corresponding formal principle would be something like the following:

$$\forall F \forall G\ [\#_L F = \#_L G \leftrightarrow (F \approx_L G)]$$

97 Ibid, p.16

However, again there are difficulties with this approach. There is the usual problem of requiring too much theoretical apparatus to be analytic, but in this case there is a more pressing and more specific issue. There is a conflict between the insistence of the type theorist that such a definition (if that is what our latest incarnation of HP* really is) be predicative and the proof of Frege's Theorem; in order to prove that there is an infinity of natural numbers, it must be possible for the function to range over any and all concepts and not just those relativised to types. Thus it is by no means clear that this kind of principle would be sufficiently strong to generate Frege Arithmetic, and hence it is inadequate for the neo-logicist.

Of course, there is always the option of taking such a treatment of indeterminacy such as a supervaluationist approach as being in some way primitive; were this to be acceptable, there would be no need to further qualify HP, as indeterminate predicates would no longer give rise to counterexamples. However, this seems to be a fairly desperate and unsurprisingly unsuccessful last resort. It is by no means obvious that for any given treatment of indeterminacy that it is the one that ought to be adopted, but any given treatment is certainly not going to be analytic. There is nothing indefinite about the criterion of application of 'is an ordinal' – it is only that it gives rise to an indefinitely extensible totality. Hence it cannot be anything to do with the meaning of the terms that (say) a supervaluationist treatment of 'is an ordinal' is correct. Thus it seems that any suitable modification of HP to account for indeterminate predicates is going to be too theoretically loaded to be analytic.

§4: Conclusion

In this chapter, I have argued that Wright's argument that targets a vindication of a sort of neo-logicism relies on the false premise that HP is analytic. HP fails to be analytic because it is false, and the most natural attempts to patch HP such that the problematic cases are excluded give rise to principles that are not analytic on any reasonable construal of the notion. However, all that has been shown so far is that Wright has set his sights too high; he has attempted, like Frege, to show that arithmetic has a particular semantic and epistemological status that it does not have. This leaves open the possibility that to reject wholesale the neo-logicist approach is premature – thus Wright's original insight about Frege's philosophy of mathematics is taken a step further. It is therefore an open question whether or not there is some other way of obtaining the epistemological result that would vindicate neo-logicism as a resolution of Benacerraf's challenge. It is to this further question that I turn in the following chapter.

Chapter 3

Recall that Benacerraf's challenge is to explain how it is possible that we have mathematical knowledge and an ability to refer to mathematical objects. Wright's neo-logicist's explanatory programme consisted of showing that the truths of arithmetic, at least, are analytic and hence are epistemically accessible if the axioms of arithmetic are. The discussion in the foregoing chapter demonstrated that the most natural ways of putting flesh on these bones are inadequate. However, the conclusion to draw need not be that the strategy is entirely misguided – the truths of arithmetic need not be *analytic* in order for us to be able to know them in the ways indicated by the neo-logicist. It can be easy to lose sight of the epistemological target: knowing the truths of arithmetic.

The viability of this kind epistemological programme has not been investigated. As such, this chapter is exploratory in nature and raises two issues. The first is whether or not this kind of epistemic neo-logicism is really viable – how well does taking a non-analytic principle on which to base our knowledge of arithmetic fit with the linguistic approach suggested in chapter 1? If this issue does not constitute an insurmountable obstacle, then we can raise the question of how such an epistemic neo-logicism would best be formulated. In what follows I give an overview of some possible options and highlight associated difficulties with each.

§1: New Neo-Logicist Strategies

Wright's argument for the analyticity of arithmetic was given as follows:

1. Frege's Theorem – arithmetic is a (recognisable) second-order logical consequence of HP.

2. If HP is analytically true, its recognisable logical consequences are analytically true.
3. HP is analytically true.
4. Therefore, arithmetic is analytically true.

In the previous chapter, I showed that (3) is false as HP is not true and its modifications are not analytic. But it is possible to see (2) and (3) as being specific instances of the following argument schema:

1. Arithmetic is a (recognisable) second-order logical consequence of HP*.[98]

2'. If HP* has **A**, its recognisable logical consequences have **A**.

3'. HP* has **A**.

4'. Therefore, arithmetic has **A**.

Where **A** stands in place of an expression for an epistemic status that entails knowability, such as analyticity.

Evaluating this argument as it stands is not a fruitful exercise, as its soundness is relative to the choice of epistemic status to replace **A**. In this chapter I will consider two ways of filling out the argument, and assess each option according to two criteria. Firstly, there is the substantive question as to whether any candidate epistemic status that is weaker than analyticity will thereby fail to be powerful enough to solve our original problem. The point is that even if the argument is sound, it may not be able to underwrite any satisfactory resolution of Benacerraf's challenge. Secondly, there is the issue of soundness – does the epistemic status in question transmit across recognisable second-order logical consequence? And does our choice of HP* (a true principle used as an alternative to HP that depends on our choice of status) have the relevant status?

98 Note that HP is of no use if it is false; as such, we must use a suitable HP* that has the same resources as HP. This result will be equivalent to Frege's Theorem, and the ensuing argument will be formally valid.

The two approaches that I will develop in this chapter are the claim that *HP* is knowable a priori*[99] and the claim that *HP* explains our concept of cardinal number*. These should be contrasted with Wright's stated aim to give *an implicit definition of cardinal number*. However, whilst Wright's aim is clear enough given the general thrust of the corpus of his work on this matter, the following suggestive passages can be seen as an adumbration of each of these alternatives:

> "Frege's Theorem will still ensure the truth of the following: 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 the concepts of second-order logic. If such an explanatory principle, in company with 'implicit definitions' generally, can be regarded as *analytic*, then that should suffice at least to demonstrate the analyticity of arithmetic. Even if that term is found troubling … it will remain that HP [gives rise to] one clear a priori route into a recognition of the truth of … the fundamental laws of arithmetic."[100]

> "If [HP] may be viewed as a complete explanation – as showing how the concept of cardinal number may be fully understood on a purely logical basis – then arithmetic will have been shown up by Frege's Theorem not [to be] part of logic, it is true, but as transcending logic only to the extent that it makes use of a *logical* abstraction principle – one whose right-hand side deploys only logical notions. … Such an epistemological result, it seems to me, would be an outcome still worth describing as logicism, albeit rather different from the conventional, more stringent understanding of the term."[101]

99 I will (to aid clarity and readability) at times abbreviate 'knowable a priori' to simply 'a priori'; thus a truth is a priori iff it is a truth that is knowable a priori.
100 Wright (1997), p.279
101 Ibid., pp.279-80

Wright thus attempts to vindicate a version of logicism by appealing to three distinct epistemological notions: analyticity, apriority and explanation of a concept. He runs these three alternatives together, but it seems to me that this is a mistake. If the target (as indicated by the first passage) is epistemological, we need only to consider how best to achieve such an epistemological result. One way to do this would be through implicit definition and showing arithmetic to be analytic and hence knowable a priori, but (as we have seen) such a claim is unsustainable – it is an implausibly strong claim. However, it is still an open question whether or not one or more of the more modest proposals of seeking to only to explain our arithmetical knowledge either in virtue of the apriority of a suitable HP* or by treating HP* as an explanation in logical terms of cardinal number are more plausible.

However, before addressing each of these options it is worth exploring a natural objection to the separation of each of these epistemological notions. The objection is not a denial that these notions come apart, but instead that the relevant differences between the notions cannot be usefully exploited by the neo-logicist: any neo-logicist position based on some weaker epistemological notion would be unstable. The thought is that there is a tension between an endorsement of the Context Principle – first and foremost a *linguistic* principle – as the basis of a solution to Benacerraf's challenge on the one hand, and of a *non-linguistic* epistemological target on the other. The Context Principle, as a principle governing meaning, can correspondingly only license an appeal to those principles that operate at the level of giving the meanings of the terms that they contain – in other words, analytic principles.

The thought is that we are perfectly entitled to appeal to analytic principles to do the required epistemological work, but that is because analyticity has both a semantic and an epistemological dimension to it. We are faced originally with a problem about reference, and particular analytic principle can help us solve it; the fact that they also help us solve a parallel problem about knowledge is a significant bonus, but it is no more than that. The neo-logicist way of framing the issue precludes taking the knowledge problem to be central, as doing so licences the use of tools that are not fit for purpose. Once we start appealing to principles with some weaker epistemological status that do not help to fix meanings, we can no longer assume that the Context Principle will do the same kind of work.

To put the point another way: consider that in chapter 1, all the justification of the Context Principle and abstractions took place under the assumption that abstractions could operate as implicit definitions. Wright and Hale, for instance, talked about how abstractions could fix "by stipulation" the truth conditions of their left- and right-hand sides, whilst Fine's discussion of the Context Principle was in terms of "creative definitions" and of linguistic practice being "partly constitutive of reference". Thus it is clear how the solution to both the knowledge and reference problems are supposed to run given the proviso that the relevant abstractions to which one appeals are analytic. The challenge is that it is at the very least unclear whether rejecting the assumption is sufficient to stymie the line of argument offered in those discussions.

This kind of objection is pressing the thought that only analytic principles are acceptable for the neo-logicist. Given the conclusion of the previous chapter, this result would be tantamount to sounding the death-knell for neo-logicism. Moreover, perhaps an assumption of this kind would help to explain why Wright would not seriously consider retreating from the position that HP is analytic: if it were not, it would be of no use to him. However, it is not clear to me that this objection is well-founded. It is certainly correct to point out that there is tension here, but such tension seems to be less serious than first appearances might suggest.

To attempt to give a decisive rebuttal to the objection would require more space than I have here, but I will present two considerations that could lead one to think that the foregoing line amounts to much less than a knock-down argument. The first is to point out that the Context Principle was predominately (and most controversially) used to bolster the case for adopting the Ontological

Plenitude Thesis about certain abstractions. This was the view that if we believe an abstraction to be true, we should believe in the existence of all the objects (abstract or otherwise) to which there is apparent explicit reference. This might most naturally be used to solve the reference problem, but we are certainly not forced to see things in this way. That the apparent reference of abstraction principles is genuine reference is something that could be ascertained in some other way. The point is that whilst it is very tempting to try to kill two birds with one stone by appealing to analytic abstraction principles, we need not do so; we are free to treat abstraction principles as being best used to solve the knowledge problem.

The second is that even if the thought that underlies the objection is a good one, it would be decisive only given the implicit assumption that a solution to the knowledge problem that does not give rise to a parallel solution to the (presumed to be parallel) reference problem is in some way unsatisfactory. If we reject this assumption, then we are free to attempt to solve one problem at a time. We can deny that the problems are parallel or that their being parallel problems would be sufficient to force us to adopt parallel solutions. We might agree that giving parallel solutions to parallel problems is a desideratum of any potential line of response, but it need not be an inviolable methodological principle. Thus whilst we have a reason to proceed with caution, rather more must be done to show that the whole strategy of using an epistemically weaker principle is misbegotten in some way. This is not to say that our putative objector could not have any more to add, and indeed it may well be something of a success to persuade a defender of a view to be suspicious of it; it is merely to point out that the objection cannot be as decisive as it might appear. Nonetheless, the issue raised by the objection is crucial and requires further attention if some suitable epistemic status is to be accorded to arithmetic and justified by the kind of argument put forward in this chapter.

§2: Cardinal Number Explained

The first option I will look at is the idea of HP*'s serving as an explanation of the concept of cardinal number. As it is difficult to frame this in terms of the argument schema given above, it is better to make a small modification as follows:

1. Arithmetic is a (recognisable) second-order logical consequence of HP*.
2''. If HP* has **A**, its recognisable logical consequences have **B**.
3''. HP* has **A**.
4''. Therefore, arithmetic has **B**.

Where **B**, like **A**, is a schematic variable standing for an expression for an epistemic status that entails knowability. We can then run an argument based on HP* explaining our concept of cardinal number by a straightforward substitution of an epistemic status for each of our schematic variables:

1. Arithmetic is a (recognisable) second-order logical consequence of HP*.
2-ECN. If HP* explains our concept of cardinal number, our understanding/knowledge of its recognisable logical consequences can be explained by HP*.
3-ECN. HP* explains our concept of cardinal number.
4-ECN. Therefore, our understanding/knowledge of arithmetic can be explained by HP*

We cannot treat 2-ECN and 4-ECN as instantiations of our original (2') and (4'), as it seems somewhat unclear what 'the consequences of HP* are explained by HP*' could mean. However (2-ECN) and (4-ECN) are clearly instances of our latter, weaker, argument schema, and are further filled out in a very natural way. We can explain what we know by appealing to HP* and Frege's Theorem, or we can explain how we are able to think about and refer to abstract number terms by appealing to HP* and Frege's Theorem.

The key idea here is that HP* is designed to be in some sense a *purely logical* explanation: an abstraction principle has an expression containing the explanandum as its LHS, and some kind of explanans as the RHS; crucially, this explanans is a formula of pure second-order logic. Abstractions give necessary and sufficient conditions for the obtaining of some fact about abstract objects, stating that whether or not such a fact obtains is solely a matter of whether some second-order logical formula is true. Thus if it is possible to explain particular truths about particular abstract objects by giving such a necessary and sufficient condition, then it seems plausible to say that we can explain our understanding or knowledge of such truths in terms of our understanding or knowledge of the relevant logical formulae and the abstraction principle – or, more concisely, (2-ECN).

Wright has plenty to say about a programme along these lines at the end of his [1983] in his discussion of Number-Theoretic Logicism (III):

> "[I]t is possible, using the concepts of higher-order logic with identity, to explain a genuinely sortal notion of cardinal number; and hence to deduce appropriate statements of the fundamental truths of number-theory, in particular the Peano Axioms, in an appropriate system of higher-order logic with identity to which a statement of that explanation has been added as an axiom."[102]

One striking feature of this thesis is that it is set up in contrast to a relevantly similar version of logicism – his Number-Theoretic Logicism (II) – that couches the relevant epistemological notions as definitions of arithmetical terms in logical terms, rather than explanations. However, Wright's subsequent development of his logicism demonstrates that he had not abandoned all commitment to

102 Wright (1983), p.153

working at the level of (implicit) definition – raising the question of whether he ought instead to have pursued a line of argument working on the level of explanations after all.

The reason that availing ourselves of HP as an explanatory axiom can give rise to something that its supporters might call logicism is, as noted, that the explanans is given in purely logical terms. As such “there will still be a route to an apprehension of the basic truths of number-theory to follow which will require knowledge only of the concepts, truths and techniques of second-order logic with identity.”[103] The idea is that HP provides an explanation of the concept of cardinal number, which (when supplemented by understanding of second-order logic) in turn provides an explanation of why an arithmetical truth is a truth rather than a falsehood.

Whilst this kind of proposal is superficially attractive, much of this evaporates under closer scrutiny. Wright's discussion of this kind of solution assumes that the relevant notion of an 'explanation of a concept' is fairly intuitive and clear, but this is far from obvious. Wright explicates it as “a statement whose role is to fix the character of a certain concept”[104]; but this, as Boolos points out, seems to be an attempt to tacitly “assign it an epistemological status importantly similar to the one it was once thought analytic judgements, including definitions, enjoy.”[105] As we are trying to distance ourselves from this kind of assimilation of explanation and analyticity, we would need to say that an explanation of a concept is something different. However, trying to give some kind of substance to such a notion of explanation that does not rule out the possibility of its having the kind of epistemic significance required by the neo-logicist is not easy and perhaps not even possible (again, as noted by Boolos).

There is an associated worry as to whether such a notion of explanation of a concept is going to give rise to a solution to the sceptical problem with which we started. The more general problem is

103 Ibid., En 18.2
104 Ibid., p.153
105 Boolos (1997)

to explain how it is possible for us to have the knowledge that we in fact have, and the more general solution is the suggestion that we explain it in terms of our knowledge of some foundational principle – HP* – that has second-order arithmetic as a logical consequence. However, we must be able to give an account that is acceptable to the sceptic of why it is that we are warranted in claiming to know HP*. To say that 'it explains our concept of cardinal number' is only enough if the notion of explanation in play entails knowability, but it is far from obvious that there is any reasonable construal of an explanation of a concept (distinct from analyticity) that has the requisite entailment. For instance, one possible reading of the claim that 'it explains our concept of cardinal number' is as a sociological point – that we (actually) understand cardinality because of our knowledge of HP*. But, clearly, this would be of no philosophical use. Again, we need some reading of an explanation of a concept that naturally gives rise to a reason to accept HP*, but it is not clear how this could be done without retreating to a position that has already been considered and dismissed.

§3: A Priori Arithmetic

I will finish this chapter by exploring the case for treating HP* as a priori. The argument that is being considered here runs as follows:

1. Arithmetic is a (recognisable) second-order logical consequence of HP*.

2-AP. If HP* is a priori, its recognisable logical consequences are a priori.

3-AP. HP* is a priori.

4-AP. Therefore, arithmetic is a priori.

As we have seen, at some level Wright took the real target to be the apriority of (at least) arithmetic, an *epistemological* result that he believes would follow from the *semantic* result of the analyticity of (at least) arithmetic. The idea is that by relying on an epistemological reading of analyticity, we can deduce the apriority of arithmetic from the stronger semantic thesis that it is analytic. However, it is worth noting at this point that the questions of analyticity and apriority come apart – even if it may be safely assumed that analytic statements are a priori, it need not be the case that we *must* establish analyticity in order to justify treating statements as a priori. It is open to us to say that a given statement can be analytic but somehow obscurely so, and hence some other route to a priori knowledge is required; another alternative is to say that a given statement can be non-analytic but nonetheless a priori. Having said this, the 'take-home message' is that analyticity is first and foremost a semantic notion (albeit one that can carry heavy epistemological weight), apriority is an epistemological notion, and the two need not be coextensive.

With this in mind, then, let us begin to evaluate the above argument by considering (2-AP). The question is whether apriority transmits across recognisable – knowable – logical consequence. It is not quite right to see this as an instance of an epistemic closure principle – if S knows (a priori) that p and S knows that p implies q then S 'is in a position to know (a priori)' that q – but the idea is much the same. Whilst this is loosely formulated, the point is clear enough; we have a clear a priori route to knowledge via knowable logical implication. In endorsing such a principle we take it that it can both explain the knowledge that we already have – we could have come to know it by our knowledge of HP* and by following the proof of any given theorem of Peano arithmetic[106] – and of any knowable but as yet unknown arithmetical truth (for the same reasons).

In order to assess (2-AP) we should consider whether or not there are important differences between this kind of restricted epistemic closure principle and the semantic closure principle (2). Prima facie, it might seem as though semantic status is something that transmits across logical consequence rather better than epistemic status, but this is perhaps not true in the case of apriority.

106 The existence of such a proof follows from Frege's theorem.

In the case of an unrestricted epistemic closure principle – that knowledge simpliciter is closed under known entailment – there are a number of purported counterexamples, but all depend on there being some unknown proposition q that is known to follow from a known proposition p. In the case that we are interested in, the relevant q must be *unknowable* (a priori) rather than merely unknown. But if p is a priori and the entailment to q is known, then that q can be deduced seems to follow; moreover, it seems to follow that q would then be knowable a priori. As such, it seems to be a rather unpromising line of response to put pressure on (2-AP); instead it is better to focus on (3-AP) – the claim that the relevant choice of HP* could be known a priori.

Moving on to (3-AP), the plausibility of the case for HP*'s being a priori is of course dependent on the choice of HP*. For instance, in cases where HP* is some kind of restriction of HP, HP* will be a priori only if we can know something – precisely what will depend on the choice of HP* – about the kinds of things to which we restrict HP. In what follows I will consider various options for the neo-logicist, but there is an essential preliminary issue to address first. In order to answer whether any given HP* is a priori, it will be important to bring into focus what would count as being knowable a priori. The obvious and natural characterisation of a priori knowledge – a proposition "which can be known to be true without any justification from the character of the subject's experience"[107] – is unfortunately not much help, as what is at issue is not whether HP* is knowable a priori rather than by justification from experience but whether it is knowable at all. What is required, then, is to suggest a way in which we might come to know HP* that seems to fit with how we might come to know a priori truths more generally. This would fit very well with Benacerraf's original generality requirements – that how we know purely mathematical statements must accord with a more general explanation of how we know natural language statements.

It seems worth pausing at this point to note that this way of framing the issue makes it clear that this kind of position is likely to be fairly unstable. One strand of Quine's famous critique of the analytic

107 Boghossian & Peacocke (2000), p.1

and the a priori[108] seeks to tie the two together in order to reject both. Boghossian and Peacocke's reconstruction provides us with a helpful taxonomy of ways to justify a priori knowledge:

> "Quine may, in rough outline, be represented as having reasoned as follows. Unless we are to resort to postulating occult faculties of knowledge, a priori knowledge will be explicable only if grasp of meaning – understanding – is somehow sufficient for knowledge of truth. Understanding will only suffice for knowledge of truth, however, if there are sentences that are true purely by virtue of their meaning. But there can be no such sentences, and so a priori knowledge is not explicable."[109]

Thus in order to defend the view that a particular truth is a priori, one must either establish its analyticity or separate the analytic from the a priori. Whilst Quine would see at least the former strategy as doomed to failure, it is part of the background to the discussion of chapter 2 that there are propositions "true purely by virtue of their meaning"; what was disputed there was that any appropriate HP* could be analytic. Thus the task of one seeking to demonstrate that HP* is a priori is to show that the a priori outruns the analytic and explain why HP* is an example of such a phenomenon.

However, many (admittedly somewhat nascent) strategies for pursuing such an explanation are ruled out by the overall project of the neo-logicist[110]. To give just one example, Yablo aims in his [2000] to demonstrate that a view that commits us to the truth of "a priori bridge principles"[111] allows us to prove the existence of almost anything we want "from a priori and/or empirical banalities"[112]; he then attempts to explain our commitment to such principles as being adequately

108 See Quine (1936) and (1962)

109 Boghossian & Peacocke (2000), p.5

110 Boghossian & Peacocke (in their (2000), Introduction) give an extremely brief summary of each approach; they use the term "non-factive" for strategies that do not commit to the truth of a priori principles to explain a priority, a term that seems both appropriate and illustrative of the unavailability of each of these to the neo-logicist.

111 Yablo (2000), p.199

112 Loc. cit.

modelled by our commitment to certain metaphorical statements. But recall that the neo-logicist's aim is to explain our mathematical knowledge; if this is to be a successful explanation, we presumably cannot appeal to principles to whose truth we are not committed. Such a view would not be consonant with Benacerraf's semantic requirement that mathematical truth is not fundamentally different from truth in other domains.

Despite this somewhat bleak prognosis, this kind of scepticism is more general than specific. It is an undoubted implicit assumption of neo-logicism that a priori knowledge is possible and, presumably, that it is not just analytic truths that are a priori. As such, whilst there is a substantial issue here, the sceptic is not addressing the concerns of the neo-logicist on his own terms. A more damaging approach would be to grant the neo-logicist his assumptions but demonstrate that there is nonetheless no adequate way of filling out the details of his programme. With this in mind, then, I turn to an assessment of the different options for the neo-logicist.

§3.1: Set-Theoretic Hume's Principle

First of all, let us try to explain the apriority of arithmetic in terms of a a version of HP where the domain is restricted to all and only those concepts whose extensions form a set. Call such a principle HP-Sets. The first thing to note is that the analogue to Frege's theorem based on HP-Sets rather than HP should follow, as we have a total function on an altered domain; however, the change of domain in this fashion is irrelevant to the proof as (crucially) 'is a natural number' has a set extension. This means that we can prove, on the basis of HP-Sets, that there is an infinity of natural numbers. Secondly, recall that this is equivalent to Cantor's Principle (and hence presumably true), but cannot be considered to be analytic if, as required by the neo-Fregean, cardinality is a property of *concepts* rather than of *sets*. The important question at this point is whether HP-Sets is a priori, even if not analytic.

One potentially fruitful starting-point in seeking to discover how we know HP-Sets at all is to consider why we need to restrict to concepts whose instances form a set. The problematic concepts were those in which there was no definite collection of things that constituted the extension of that concept. In the case of semantically paradoxical predicates I argued that the difficulty arose at the semantic level, whereas there was a more fundamental, metaphysical problem with indefinitely extensible predicates. Whilst there were differences in how the indefiniteness was explained, the point is that in each case there could be no cardinal number of the problematic concept because the concept is in some way indefinite.

Rather than talking in terms of definite and indefinite concepts, it is more perspicuous to frame the discussion in terms of definite and indefinite *extensions* of concepts. We can define a definite extension as follows:

> (Definition of 'definite extension'): A concept C has a definite extension iff for any entity E, exactly one of the following is true: {E is an instance of C; E is not an instance of C}.

With this in mind, we should note that sets, in marked contrast to concepts, are paradigmatically definite. By the axiom of extensionality, sets have their identity just in virtue of their elements. Moreover if we have a definite collection of things, we can 'lasso' them to form a set – it is this kind of idea that underpins the iterative conception and cumulative hierarchy of sets on which, say, ZF is based. This is captured by the following conceptual truth:

> (Principle of definiteness of sets): For any collective S, S is a set iff for any entity E, exactly one of the following is true: {E is a member of S; E is not a member of S}.

From this it follows that a concept has a definite extension iff there is some set that contains all and only those entities that are instances of C. Thus we can take as an auxiliary principle the very plausible 'a concept has a definite extension iff its instances form a set.'

If this is true, then the next step in establishing the apriority of HP-Sets is a way of knowing the further principle that 'a concept has a cardinal number iff it has a definite extension.' Were we to justify such a claim, we would then have it that a concept has a cardinal number iff it has a definite extension, and a concept has a definite extension iff its extension forms a set. Therefore a concept has a cardinal number iff its extension forms a set. But this is just the principle that we need to warrant restricting HP to concepts with set extensions. If there is nothing untoward about such a restriction, then it seems that we have a natural, non-ad hoc way to alter HP that excludes the problematic concepts and incurs minimal epistemological costs. It seems to me that this is the kind of result that the neo-logicist is looking for, as we have at least "one clear a priori route into a recognition of the truth of … the fundamental laws of arithmetic". Thus if HP-Sets is a priori and its apriority transmits across second-order logical consequence, arithmetic is a priori; therefore the truths of arithmetic are knowable even if what we know are facts about abstract objects.

But what are we to make of the latter auxiliary principle – that a concept has cardinality iff it is definite? One way to evaluate it is to split the biconditional into its left-to-right and right-to-left directions. The left-to-right direction seems fairly plausible – if a concept has a cardinal number, then it is definite. One way to bring out this plausibility is to consider putative counterexamples to the claim. This would be some indefinite concept with a particular cardinal number attached to it; but such a concept would have to be a concept such that no matter which things fell under it, the number of things that fell under it would not change.

But this kind of case seems incoherent. This can be brought out by considering a small cardinality example in which a concept C has cardinality 1 but no definite extension. Thus there are entities E_1

and E_2 such that exactly one is an instance of C but it is not determinate which of E_1 and E_2 is that instance. But by the former condition, we have two possibilities: that E_1 is an instance but E_2 is not, or E_2 is an instance and E_1 is not. But in either case it is determinate which of E_1 and E_2 is an instance, contrary to the latter condition.

The right-to-left direction is rather more problematic. It presumably requires something like the following: if a concept has a definite extension, then there is a fact of the matter as to whether or not any given thing falls under that concept. This much follows from our definition. However, it is by no means straightforward to take the further step and establish that this result can help us to show that such a concept has a cardinal number. We have merely identified a necessary but insufficient condition. It may be sufficient if we further restrict to *enumerable* concepts, as such concepts are equinumerous with some initial segment of the numerals. Indeed, it is hard to see what more could be required for a concept to have a cardinal number beyond it being definite and being equinumerous with some initial segment of the numerals.

However, it is not clear that such a restriction is warranted; recall that a restricted version of HP can be a priori only if we can know a priori something about that restriction. We now need to know something a priori about both sets and sizes of concepts, and this latter restriction may be difficult to justify even on its own terms. Furthermore (as we shall see) a restriction to enumerable concepts is a very strong constraint, inasmuch as a version of HP* restricted to enumerable concepts allows us to prove everything that we want to prove. We need not appeal to any principles about sets at all if we have such a limitation of size. Hence such a restriction seems at best somewhat out of place, even if it can be justified on independent grounds. Therefore as things stand, *from a neo-logicist viewpoint* our original conditional – that a concept has cardinality if it is definite – lacks sufficient motivation.

§3.2: Conditional Hume's Principle

We can try a different tack by noting that the difficulties with HP and its variants seem to arise with their right-to-left halves. The thought is that whilst we take it that numbers and their ilk actually exist, we cannot presuppose such an existence; but endorsing the right-to-left half without further justification is to make just such a presupposition. Instead, then, let us reconsider ICHP

$$\text{(ICHP): } \forall F \forall G \forall u \forall v\ [(u = \#F\ \&\ v = \#G) \rightarrow (u = v \leftrightarrow F \approx_{1\text{-}1} G)]$$

as discussed in chapter 2: if there are such things as numbers of concepts, then any two such numbers are equal just when their concepts are equinumerous. As this principle is essentially a conditional rather than a biconditional, we do not have the troublesome presupposition of the right-to-left half of HP. Moreover the principle is not subject to the same kinds of counterexamples as HP itself, as whenever there is no number of Fs or Gs the conditional will still be true by falsity of antecedent.

This kind of principle is originally supposed by Wright and Hale to be without much merit precisely because it lacks the existential commitments of HP; this in turn derives from their belief that caution is unnecessary at this point. We can unpack the thought that underlies such a thesis by considering a suggestive passage from Wright and Hale in which they discuss the relationship of a priori knowledge and the stipulated conditionals of the form of HP or ICHP:

> "How in principle might the infinity of the series of natural numbers ever be recognised? Of course there's the option of simply denying that it can be... but anyone sympathetic to the opposing thought, viz. that the infinity of natural numbers – and indeed the truth of the Dedekind-Peano axioms – is part of our most basic

> knowledge, should be receptive to the idea that it is inferential knowledge, grounded ultimately in deeper principles of some kind determining the nature of cardinal number."[113]

The point here is that all parties to the debate – including the Benacerrafian sceptic – can agree that we do have knowledge of the relevant arithmetical facts. We can further agree that we can infer this knowledge from knowledge of some 'deeper principle'. Indeed, it can be conceded (following on from the discussion of chapter 1) that such a 'deeper principle' can play the role of fixing a reference of number terms and be an explanatory route to our knowledge of arithmetic. Wright and Hale go on to say that this 'deeper principle' can itself be stipulated fairly freely, and that this does not in turn make the existence of the natural numbers and the obtaining of arithmetical facts a matter of stipulation. They argue for this by

> "urging... that the stipulation of HP should be seen first and foremost as a meaning-conferring stipulation – one providing for the introduction and elimination of contexts of numerical identity – of which it is a relatively un-immediate, interesting and welcome consequence that there is a population of objects of which the Peano axioms are true."[114]

However, it is at this point that the disagreement becomes a little sharper. The thought that Wright and Hale wish to dispute is that our epistemic warrant to a given principle is only as good as our warrant to its consequences; as such on their view, to stipulate HP without further justification or explanation is not simply to help oneself to its consequences (and hence, a fortiori, it is not to stipulate into existence particular objects). This is why caution is supposed to be unnecessary. However, the upshot of chapter 2 is that even to allow that stipulation of HP were to be admissible if it were true, there is no acceptable true substitute principle for stipulation. As such, we must

113 Hale & Wright (2000), p.147
114 Ibid., p.148

accept that caution *is* necessary, but this need not be a barrier to tentatively endorsing a conditional such as ICHP (with no consequences that entail the existence of particular objects) and asking what kind of epistemic status it has. Do we need to know it, and how can we know it (if we can know it at all)?

One attractive feature of ICHP is that it seems to serve as a partial explanation of the concept of cardinal number. It is an instance of a claim of the form 'if anything is an x, then it is F, G, …', a statement that can help to tell us what is distinctive about the concept of cardinal number. Moreover, as noted above, it seems to escape the problems associated with the consequence of HP that all concepts have a unique number; there is no entailment either to there being a (problematic) number of all sets or ordinals, nor is there a stipulation of the existence of infinitely many particular objects. Thus it seems as though it is a candidate for being necessarily true and knowable a priori.

However, such a proposal seems a little lightweight for the neo-logicist's purposes. In ICHP, we have a principle that we can know to be true but not one that we know what makes it true. There seems to be little that is disanalagous from the Benacerrafian sceptic's point of view between an arithmetical theory based on ICHP, and a scientific theory based on the inverse-Carnap conditional 'if there is such a thing as phlogiston, then it is removed during a chemical reaction'. In both cases, the idea is to treat the existence of numbers or phlogiston as part of a theory which may or may not be true; as such, if there is in fact no such thing as a number or phlogiston, then the theories have precisely nothing to say. But this means that these kinds of principles might well be true, but are epistemically ineffectual. Whilst they help give some kind of partial explanation of a concept, they do not help answer the sceptic. What we need is a reason to think not only that the conditional is true but that the antecedent is true, and that is not given by such a conditional.

§3.3: Finite Hume's Principle

The final versions of HP* that I will consider are Heck's FHP and a variant of it. Recall that Heck's own principle is

$$\text{(FHP): } \forall F \forall G\, [(F \text{ finite} \vee G \text{ finite}) \rightarrow (\#F = \#G \leftrightarrow F \approx_{1\text{-}1} G)].$$

The variant is HP where the domain is restricted to finite concepts. Call such a principle HP-Weak[115]. HP-Weak is weaker than FHP as the latter, unlike the former, rules out any infinite concept having the same number as a finite concept. However, each avoids being committed to the problem cases for HP itself that arise when it is instantiated by concepts that are indefinitely extensible, in the sense that for any set S of instances of the concept there is a set of instances which is a proper extension of S. Furthermore, the technical questions of whether there is some appropriate version of Frege's Theorem underwritten by each principle are answered in the affirmative by Heck's [1997].

These preliminary considerations give us no reason to prefer one approach over the other. However, it seems to me that there is likely to be rather more mileage in the conditional principle rather than the restricted biconditional. This can be seen by comparing a purported justification of the restriction required by HP-Weak with that required by HP-Sets. The strategy pursued in justifying the restriction to concepts with definite extensions was to find some auxiliary principles that could be known a priori that would justify the restriction of the domain. However, it is hard to see what the equivalent could be in this case. But without some further basic principle to which we can appeal to justify our restricting the domain, it is hard to envisage how we could come up with the requisite justification. Given this, I will focus on FHP.

115 The terminology echoes Heck (1997)

Heck's principle is a conditional rather than a biconditional, but it does not have the same defects as ICHP. It is not vacuously true by falsity of antecedent in all (relevantly) empty domains, as the condition imposed by the antecedent is a condition on concepts rather than objects. Of course, as FHP is not an abstraction it cannot be assumed to have whatever privileged epistemic status abstraction principles are required by the neo-logicist to have; however, there may well be some other way of establishing that it has the relevant status.

As indicated in chapter 2, FHP cannot be taken to be analytic because there is no prior notion of cardinality of which we can say that FHP is analytic. However, all that we require at this point is an a priori principle that can explain our knowledge of arithmetic – specifically of finite arithmetic. It seems to me that FHP is a strong candidate for such a principle. Its a priori status need not be informed by any deeper principle – it seems that if one sets aside more general misgivings about the possibility of a priori knowledge, then FHP is a likely principle to be known to be true without requiring any justification from experience. It is surely a knowable principle about counting and finite arithmetic that any two finite concepts will have the same number of things falling under them iff they are equinumerous; but equally, any justification for that need not make reference to any kind of experience.

The issue pressed in chapter 2 was that FHP failed to characterise any pre-existent notion; as such, the objection to treating it as analytic was that it failed to say enough, so to speak. However, that need not be a barrier to taking it to be a priori; the constraint that there be some extant notion characterised by our principle that comes from the demand for analyticity simply does not apply. The only kind of adequacy constraint in operation here comes from the need for our choice of HP* to serve as a foundation for arithmetic – in other words, that some version of Frege's theorem with HP* replacing HP goes through. As demonstrated by Heck, FHP satisfies this constraint. Therefore there seem to be no *specific* objections to treating FHP as a priori. This is of course not to rule out the possibility that there may be more *general* reasons for taking FHP to be an inadequate foundation for arithmetic; but unlike HP there is nothing specifically wrong with it. The point here is that FHP is a better auxiliary principle than its competitors, and that as such if neo-logicism is to be vindicated at all, its best chance is by way of an adoption of FHP.

§4: Conclusion

I have aimed in this chapter to clarify two points that play an important role in any more general assessment of the neo-logicist strategy. The first is that there is a tension of sorts between the linguistic nature of the resolution of Benacerraf's challenge envisaged in chapter 1 and the purely epistemological target of showing that an abstraction (or something relevantly similar) is a priori. But if this tension can be shown to be relatively unproblematic, as my assessment suggests, then the second (more positive) conclusion is that the most plausible way to develop the a neo-logicist epistemology of arithmetic is by explaining our knowledge of arithmetic in terms of our a priori knowledge of FHP.

Conclusion

In this thesis I have investigated some underexplored options for the neo-logicist about arithmetic. I have argued that taking Benacerraf's challenge to the acceptability of any philosophy of mathematics is a good starting point, as it helps to bring into focus the problems of knowledge and reference that beset any philosophy of arithmetic that posits the existence of abstract objects. The core of the neo-logicist proposal is, I have argued, the Context Principle; it is that which offers a linguistic solution to the knowledge and reference problems. We can use it to explain the ability that we in fact have to refer to mathematical abstract objects and the knowledge of facts about those objects that we in fact have.

A natural development of this line of thought made essential use of abstraction principles; in particular, Wright argued that at least one such principle – Hume's Principle – can be analytic. The thought is that if Hume's Principle is analytic, then its analyticity will be conferred on its recognisable logical consequences. By Frege's Theorem, such consequences include elementary number theory. Thus it is possible to explain our knowledge of the truths of (at least) arithmetic by appealing to our understanding of the meaning of the terms which, by analyticity, is sufficient for knowledge to be possible. Modulo various worries about the coherence of any such notion of analyticity, this argument stands or falls with whether or not Hume's Principle is analytic. I demonstrated in chapter 2 that this key premise is false, and hence that Wright's conclusion – that arithmetic is analytic – is not sufficiently well supported by his argument.

Nonetheless, given the *epistemological* nature of the challenge to a platonist theory, it seems as though there is room to say that the same structure of argument could be used to confer some other epistemological status on at least basic arithmetic. In the more exploratory final chapter, I discussed whether there really was space for such a solution. I then investigated two possible attempts to instantiate this kind of epistemological neo-logicist solution – treating the principle on which arithmetical knowledge was to be founded as an *explanation of our concept of cardinal number*, and treating it as *knowable a priori*. I indicated that the latter attempt seemed more promising as a strategy, and that in turn it was best fleshed out by relying on Finite Hume's Principle.

In giving an assessment of different neo-logicist strategies, I have raised a number of issues (if only to set many of them to one side) which arise at varying levels of generality. Chapter 1 contained a discussion of a very general form of scepticism and a correspondingly general technique for circumnavigating the problem. In contrast, chapters 2 and 3 looked at more specific possible developments of this general strategy. As such, the viability of the relatively positive suggestions of chapter 3 are open to question in both a general and a specific manner. Three particularly important examples of a kind of more general scepticism are as follows; firstly there are serious doubts brought to prominence by Dummett about the inadmissability of abstractions as explanations of any kind. Secondly, one can argue that the neo-logicist strategy of appealing to the Context Principle as a principle governing reference is misguided. Thirdly, one could query Frege's view (inherited by neo-logicists) that numbers are numbers of concepts – for instance, one could instead take it that Cantor's set-theoretic framework is superior. Whilst these issues are interesting and important for a neo-logicist, they can be and have been set aside in the foregoing discussion.

The specific grounds for doubt seem to me rather more interesting, as they query whether or not there is any room for a solution of the kind envisaged in the final chapter. One issue (albeit still a relatively general one) is exactly what it would take for the candidate principles to serve as a foundation for arithmetic to be a priori but non-analytic. More specifically still, there is more to say about whether or not neo-logicism can abandon analytic principles so lightly – does the appeal to the Context Principle rule out a later appeal to anything weaker than analyticity?

I have offered considerations in favour of treating the neo-logicist strategy for resolving the

Benacerrafian challenge as a viable option, but have pointed out a particular shortcoming of an extant attempt to make the strategy work. However, if it is legitimate to retain as many of the insights of the rejected programme as possible by developing a neo-logicist theory of a priori arithmetic (rather than of analytic arithmetic), then such a theory based on Finite Hume's Principle has clear epistemological advantages over its competitors.

References

Benacerraf, P., (1973), 'Mathematical Truth', *Journal of Philosophy* 70: 661-80. Reprinted in Benacerraf & Putnam (1983), pp.403-20

Benacerraf, P. & Putnam, H., (1983), *Philosophy of Mathematics,* Cambridge University Press

Boghossian, P., & Peacocke, C., (2000), *New Essays on the a Priori*, Clarendon Press

Boolos, G., (1971), 'The Iterative Conception of Sets', *Journal of Philosophy* 68: 215-32. Reprinted in Boolos (1998), pp.13-29

---- (1987), 'The Consistency of Frege's *Foundations of Arithmetic*', in Jarvis Thompson, Judith (ed.): *On Being and Saying*. Reprinted in Boolos (1998), pp.183-201

---- (1997), 'Is Hume's Principle Analytic?', in Heck, R. Jr. (ed.): *Logic, Thought and Language*. Reprinted in Boolos (1998), pp.301-14

---- (1998), *Logic, Logic and Logic*, Harvard University Press

Burgess, J., (1983), 'Why I am not a Nominalist', *Notre Dame Journal of Formal Logic* 24: 93-105

Dummett, M., (1973), *Frege: Philosophy of Language*, Duckworth

---- (1991), *Frege: Philosophy of Mathematics*, Duckworth

Ebert, P., (2006), 'What Mathematical Knowledge Could Not Be', *St. Andrews Undergraduate Philosophy Society Journal*, 1: 46-70. Reprinted at http://web.mac.com/philipebert/Site/Online_Papers/Entries/2006/12/6_%E2%80%9CWhat_Mathematical_Knowledge_could_not_be%E2%80%9D_files/WhatMathematicalKnowledgeCouldNotBe.pdf, pp.1-19

Field, H., (1984), 'Platonism for Cheap? Crispin Wright on Frege's Context Principle'. *Canadian Journal of Philosophy* 14: 637-62. Reprinted in Field (1989), pp.147-170

Field, H., (1989), *Realism, Mathematics and Modality*, Basil Blackwell

Fine, K., (2002), *The Limits of Abstraction*, Clarendon Press

Frege, G., (1884), *Die Grundlagen der Arithmetik*, W. Koebner. Reprinted with English translation by Jacquette, D. (2007) as *The Foundations of Arithmetic*, Longman

---- (1893), *Die Grundgezetze der Arithmetic I*, H.Pohle. Partially reprinted with English translation by Furth, M. (1964) as *The Basic Laws of Arithmetic*, University of California Press

---- (1903), *Die Grundgezetze der Arithmetic II*, H.Pohle. Partially reprinted with English translation by Geach, P. and Black, M. (1952) as *Translations from the Philosophical Writings of Gottlob Frege*, Blackwell

Hale, R., (1994), 'Singular Terms (2)', in McGuinness, Brian and Olivieri, Gianluigi (eds.): *The Philosophy of Michael Dummett*. Reprinted in Hale & Wright (2001), pp.48-71

---- (1995), 'Dummett's Critique of Wright's Attempt to Resuscitate Frege', *Philosophia Mathematica* 3: 122-47. Reprinted in Hale & Wright (2001), pp.189-213

---- (1996), 'Singular Terms (1)', in Schirn, Matthias (ed.): *Frege: Importance and Legacy*. Reprinted in Hale & Wright (2001), pp.31-47

---- & Wright, C., (2000), 'Implicity Definition and the A Priori', in Boghossian & Peacocke (2000), pp. 286-319. Reprinted in Hale & Wright (2001), pp.117-152

---- ---- (2001), *The Reason's Proper Study*, Clarendon Press

---- ---- (2002), 'Benacerraf's Dilemma Revisited', *European Journal of Philosophy* 10: 101-129

Heck, R. Jr., (1997), 'Finititude and Hume's Principle', *Journal of Philosophical Logic* 26: 598-61

MacBride, F., (2000), 'On Finite Hume', *Philosophia Mathematica* 8: 150-9

---- (2002), 'Could Nothing Matter?', *Analysis* 62: 125-35

---- (2003), 'Speaking with Shadows: A Study of Neo-Logicism', *British Journal for the Philosophy of Science* 54: 103-63

Peacocke, C., (1999), *Being Known*, Clarendon Press

Quine, W., (1936), 'Truth by Convention' in O. Lee (ed.): *Philosopical Essays for Alfred North Whitehead*. Reprinted in Benacerraf & Putnam (1983), pp.329-54

---- (1962), 'Carnap and Logical Truth', in Kaezemeir B., Vuysje D. (eds.): *Logic and Language: Studies Dedicated to Professor Rudolf Carnap on the Occasion of His Seventieth Birthday*. Reprinted in Benacerraf & Putnam (1983), pp.355-76

Rumfitt, I., (1999), 'Frege's Logicism', *Proceedings of the Aristotelian Society*, Supplementary Volume 73: 151-80

---- (2001), 'Hume's Principle and the Number of all Objects', *Noûs* 35: 515-41

Tennant, N., (1987), *Anti-Realism and Logic: Truth as Eternal*, Clarendon Press.

---- (1997), 'On the Necessary Existence of Numbers', *Noûs* 31: 307-36

Wright, C., (1983), *Frege's Conception of Numbers as Objects*, Aberdeen University Press

---- (1997), 'On the Philosophical Significance of Frege's Theorem', in .Heck, R. Jr. (ed.): *Logic, Thought and Language*. Reprinted in Hale & Wright (2001), pp.272-306

---- (1998), 'Is Hume's Principle Analytic?', *Notre Dame Journal of Formal Logic* 40. Reprinted in Hale & Wright (2001), pp.307-32

Yablo, S., (2000), 'Apriority and Existence', in Boghossian & Peacocke (2000), pp.197-228