A PROBLEM AND A SOLUTION FOR NEO-FREGEANISM[*]

MICHAEL GABBAY
Kings College London

Abstract

I argue that Benacerraf's famous objection to mathematical realism in his paper *"What Numbers Could Not Be"* can be adapted to present severe difficulties for the Neo-Fregean programme. I formulate an alternative abstraction principle and argue that there is no reason for the natural numbers to generated by one abstraction principle rather than the other.

Independently of this conclusion, the formal comparison of the two abstraction principles involves a result of interest to Neo-Fregeans: I offer a solution to the *bad company* objection.

1 BENACERRAF ON WHAT NUMBERS COULD NOT BE

In his celebrated paper, *"What numbers could not be"*, Benacerraf presents a challenge to theories identifying numbers with set theoretic constructs. He asks why the numbers should be identified with sequence (1), the Von Neumann ordinals, rather than sequence (2), the Zermelo ordinals.

$$\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}, \{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\} \tag{1}$$
$$\varnothing, \{\varnothing\}, \{\{\varnothing\}\}, \{\{\{\varnothing\}\}\} \tag{2}$$

Benacerraf concludes that there is no reason why the number 3 should be identified with an element from one construction rather than another. 3 cannot be identified with both as the constructs have incompatible properties. For example in (1) the fourth element has three members, but in (2) the fourth element has only one member. Since the number 3 cannot be both $\{\varnothing, \{\varnothing\}, \{\varnothing, \{\varnothing\}\}\}$ and $\{\{\{\varnothing\}\}\}$ and there is no fact of the matter whether it is one or the other, it is neither. Thus the attempted identification of numbers with sets has been refuted.

> … if the number 3 is really one set rather than another, it must be possible to give some cogent reason for thinking so. But there seems to be little to choose among the accounts *for the accounts differ at places where there is no connection whatever between features of the accounts and our uses of the words in question*. [Benacerraf 1965]

It is not hard to see that this objection generalises to any theory of numbers that has an ontology containing different sequences of objects that could serve as references of our number language. Realists may escape Benacerraf's argument either by finding a suitably miserly ontology of abstract objects (the ontology of sets is too vast), or simply refusing to get involved in the metaphysics of abstract objects.

I shall argue that Neo-Fregean ontology suffers from Benacerraf's objection in much the same way as the ontology of sets. I conclude, analogously to Benacerraf's original argument, that Neo-Fregean ontology is necessarily too rich and therefore does not provide a satisfactory foundation for arithmetic.

First I shall sketch the Neo-Fregean account of arithmetic, I shall assume that the reader is largely familiar with the formal concepts behind it (in particular, I assume the reader has some knowledge of the workings of *Frege's Theorem* [Wright 2000]).

2   NEO-FREGEANSIM ON WHAT NUMBERS COULD BE

2.1 Hume's principle

The aim of the Neo-Fregean programme is to provide a metaphysics of abstract objects together with an informative account of our epistemic link to them. According to Neo-Fregeanism, reference to the abstract objects that are *the numbers* derives from logic and definitions alone. Logic then entails arithmetic truths and, in this sense, arithmetic is *analytic*.

Neo-Fregeanism promises to provide a realist theory of number that can respond to Benacerraf's argument. According to Neo-Fregeanism, certain abstract objects exist, and we can know and refer to them via *abstraction principles*. The natural numbers are among those abstract objects we can know about via a particular abstraction principle, *Hume's Principle*:

The number of F = the number of G iff
the F and the G are in one-one correspondence   (3)

For each predicate $F,$ Hume's principle identifies or allows reference to, an object that is the number of F. This formalisation should be familiar to the reader:

$$\forall F \forall G[nx.Fx = nx.Gx \leftrightarrow F1{\sim}1G] \qquad (4)$$

Hume's principle is to be taken as a *definition*, in terms of one-one correspondence, of the binding term-former $nx.(\dots)$. Furthermore, the Neo-Fregeans argue that one-one correspondence is a fundamental application and concept of cardinal numbers. So the abstract entities, reference to which is generated by Hume's principle, really are the cardinal numbers (they are the only abstract entities tied appropriately to the application of counting).

Since I intend to dispute that the binding operator $nx.(\dots)$ really does refer to the numbers it would beg the question against my arguments to continue reading it as a shorthand for 'the number of ...'. It is therefore more convenient to regard Hume's principle as follows:

The Hume-abstract of F = the Hume-abstract of G iff
the F and the G are in one-one correspondence   (5)

which is formalised as

$$\forall F \forall G[hx.Fx = hx.Gx \leftrightarrow F1{\sim}1G] \qquad (6)$$

I now take Neo-Fregeanism to be a view that make the following two claims:

- Hume's principle implicitly defines a binding operator $hx.(\dots)$ that refers to certain abstract entities, Hume-abstracts.
- The Hume-abstracts (or at least some of them) are the cardinal numbers

2.2 Frege's theorem

I now sketch how Neo-Fregeans use Hume's principle to provide a realist foundation for arithmetic.

Following Frege, the strategy is to define suitable properties and relations that satisfy the second order Peano axioms of arithmetic. To distinguish the defined terms of this section with the defined terms of Section 3.2, I subscript them with H for 'Hume'.

First a successor/predecessor relation is defined:

$$Pre_H(t, t') \quad \text{means} \quad \exists F \exists z[t' = hx.Fx \wedge Fz \wedge t = hx.(Fx \wedge x \neq z)] \quad (7)$$

So t is the predecessor of t' when t' is the number of some property F and t is the number of the Fs that are not z, for some z that is F.

Zero is defined to be the number of any inconsistent property, e.g. $0_H = nx.(x \neq x)$, it does not matter which as all empty properties are in one-one correspondence.

A natural number is then defined as being any number in the transitive closure of the predecessor relation from 0_H. More formally, the transitive closure of any binary relation R may be defined as:

$$R^*(t, t') \quad \text{means} \quad \forall F \, [(Ft \wedge \forall x \forall y(Fx \wedge R(x, y) \rightarrow Fy)) \rightarrow Ft'] \quad (8)$$

And now we may define the natural numbers as all those objects in the transitive closure of the predecessor relation from 0_H:

$$Nat_H(t) \quad \text{means} \quad Pre_H^*(0_H, t') \quad (9)$$

So a natural number is any referent of an abstraction $nx.Fx$ that can be 'reached' from 0_H by following the relation Pre_H. As did Frege, Neo-Fregeans go on to define individual number terms:

$$\begin{aligned}
&0_H \quad \text{means} \quad hx.(x \neq x) \\
&1_H \quad \text{means} \quad hx.(x = 0_H) \\
&2_H \quad \text{means} \quad hx.(x = 0_H \vee x = 1_H) \ldots
\end{aligned} \quad (10)$$

From these definitions we can derive all of Second order Peano Arithmetic, which completely characterises a natural number structure.

3 AN ALTERNATIVE ABSTRACTION PRINCIPLE

3.1 Benacerraf's principle

Now I turn to the argument that the Neo-Fregean ontology contains too many abstract objects. I do this by presenting an abstraction principle that is similar to Hume's principle. This alternative abstraction principle can do the same work as Hume's principle and in a similar way. But, as with the (1) and (2) above, the two abstraction principles yield two distinct sequences of abstract objects. As was argued in the case of sets, I shall argue that there is nothing to decide which abstraction principle yields the 'true' natural numbers.

The new principle is simpler than Hume's principle, call it Benacerraf's Principle:

> the Benacerraf-abstract of F = the Benacerraf-abstract of G iff
> neither F nor G are singletons, or
> F and G have the same extension   (11)

Say that F is *unitary* if it has exactly one element in its extension. Then Benacerraf's principle identifies, for each predicate F, an object that is the *Benacerraf-abstract of F*. We can write 'the Benacerraf-abstract of F' as $bx.Fx$, and then Benacerraf's principle is:

$$\forall F \forall G[bx.Fx = bx.Gx \leftrightarrow ((\neg \exists!xFx \wedge \neg \exists!xGx) \vee \forall x(Fx \leftrightarrow Gx))] \quad (12)$$

Benacerraf's principle is to be taken as a *definition*, in terms of being unitary, of the binding operator $bx.(\ldots)$. The intuition for Benacerraf-abstracts is that one can abstract out of a unitary property its 'unitariness', or the way in which it is unitary. Any non-unitary properties are unitary in the same way: they are not. The way unitary properties are differentiated, in the spirit of Frege's Basic Law V (see (14)) is through their extension.

Benacerraf-abstracts can be seen as abstractions over the unit. If F is a unit property then its Benacerraf-abstract is unique to it. If any two properties are not units then they have the same Benacerraf-abstract which represents their non-unity. So for each unit property there is a distinct abstract unit, its Benacerraf-abstract, associated with it (compare this with Frege's discussion of Jevons and Schröder in [Frege 1953, §29–§44]).

The concept of a unit is at least as fundamental to our concept of number as one-one correspondences. After all, a one-one correspondence is a correspondence between unit objects; when we count, we count unit individuals; variables of first order quantifiers range over unit entities; the symbols of the language necessary to express even basic propositions are discrete, discernable units. To develop the notion of one-one correspondence we need to be able to conceive of the world as not being completely amorphous but as being made up of individuals, units.

Without the concept of a unit, a discrete thing, a single entity, we cannot even begin a logical enquiry let alone ground arithmetic in one-one correspondences.

In particular notice that Hume's principle (5) cannot be formulated without a formalisation of one-one correspondence which itself requires some formalisation of what it is for a property to be unitary. Thus, Benacerraf's principle uses no formal concepts or technical details *that are not required to understand Hume's principle*. In this sense Benacerraf's principle *is more elementary* than Hume's principle.

Hume's principle abstracts objects out of a formalisation of a relation of *being both in one-one correspondence*, Neo-Fregeans then argue that Hume-abstracts are numbers. Benacerraf's principle abstracts objects out of a formalisation of the relation of *being both unitary*, and as we shall see, we can develop an analogous theory of arithmetic if we take the Benacerraf-abstracts to include the natural numbers.

Frege himself discusses and rejects the possibility of developing a theory of arithmetic based on units. But his compelling refutations are aimed at theories of numbers as agglomerations or sums of (distinct) units [Frege 1953, §29–§44]. Frege objects that such accounts either make no sense, or fail to generate arithmetic. He did not consider the possibility that the unit, thought of as a property of properties, and derived by a similar abstraction method to Frege's own, could do the same work as his favoured theory of number.

3.2 An analogue of Frege's theorem

I now sketch how Benacerraf's principle can be used to define the numbers along Neo-Fregean lines. To distinguish the defined terms of this section with those of Section 2.2 I subscript them with B for 'Benacerraf'. We begin with zero:

$$0_B \quad \text{means} \quad bx.(x \neq x)$$
$$1_B \quad \text{means} \quad bx.(x = 0_B)$$
$$2_B \quad \text{means} \quad bx.(x = 1_B) \dots \tag{13}$$

It is not hard to show that the i_B are derivably distinct. For example suppose that $0_B = 1_B$, then $bx.(x \neq x) = bx.(x = 0_B)$. So by Benacerraf's principle either $\neg\exists!x(x \neq x) \wedge \neg\exists!x(x \neq 1_B)$ or $\forall x(x = 0_B \leftrightarrow x \neq x)$. Each of these is derivably false in even first order logic.

We can go on to define the predecessor relation as follows:

$$Pre_B(t, t') \quad \text{means} \quad \exists F[t = bx.Fx \wedge t' = bx.(x = t)] \tag{14}$$

A version of Frege's theorem now arises out of adopting Benacerraf's principle rather than Hume's principle. We use (6) to define the natural numbers to be exactly the entities in the transitive closure of Pre_B. This yields the second order Peano axioms.

$$Nat_B(t) \quad \text{means} \quad Pre_B^{*}(0_B, t') \tag{15}$$

I omit the remaining details here as they are almost identical to those of the proof of Frege's theorem in [Wright 1983].

3.3 The attack on Neo-Fregeanism

Let 0_H, 1_H, 2_H,… denote the entities abstracted and defined using Hume's Principle, call them the *Hume-numbers*. Let 0_B, 1_B, 2_B,… be the entities abstracted and defined using Benacerraf's principle call them the *Benacerraf-numbers*. It should be clear that an analogue of Benacerraf's original challenge arises. Benacerraf's original argument now applies, both the Hume-numbers and the Benacerraf-numbers serve as characterisations of the natural numbers. Furthermore there is no reason for the natural numbers to be identified with the Hume-numbers rather than the Benacerraf-numbers. Therefore Neo-Fregeanism is to be rejected alongside set theoretic reductionism by a variant of Benacerraf's original argument.

The penultimate claim, that there is no choosing between the Benacerraf and the Hume numbers, is in need of justification. I sketch a justification of it in Section 4 by comparing the systems obtained from the two abstraction principles and showing that there is little that can be done with Hume's principle that cannot also be done with Benacerraf's principle.

The reader should note that I am not arguing for Benacerraf's principle over Hume's principle. Neither am I arguing just that there is a mere burden of proof for Neo-Fregeans to justify their position further. My argument, like Benacerraf's, is that there is nothing in our arithmetic practices or concepts that determines which abstraction principle (out of many possible abstraction principles) really generates reference to numbers. I conclude that neither of them do and that numbers are therefore not founded in Hume's principle as the Neo-Fregeans argue.

4 A comparison of two abstraction principles

4.1 The concept of number

Perhaps an analysis of our concept of number will differentiate between the two principles. After all, it is Frege's discussion of the concept of number that initiated the whole debate.

There is a strong sense in which one-one correspondence is fundamental to our number concepts. This being the case we might argue in favour Hume's principle, over Benacerraf's principle, as follows.

- There is a conceptual connection between one-one correspondence and sameness of cardinal number.
- Hume's principle abstracts the referring expression 'the Hume-abstract of …' over sameness of number and Benacerraf's principle abstracts over something other than sameness of number.
- Therefore Hume's principle is the correct abstraction principle for reference to numbers and the Hume-abstracts really are the numbers.

The case for the first premise of this argument appears strong. There seems to be little, if anything, more to sameness of number than one-one correspondence. Furthermore it seems that someone who fails to use one-one correspondence in making judgements about cardinality has failed to understand the concept. The conclusion then looks inevitable that Benacerraf's principle offers the wrong conceptual analysis of cardinal number and so can be discounted.

A first response to this argument is that it conflates a necessary equivalence with a conceptual reduction. It is hard to deny that if F and G are

in one-one correspondence then that *entails* that their number is the same. But it is a separate matter entirely whether one-one correspondence is conceptually tied of sameness of number. Why could it not be that the concept of number is obtained from elsewhere in such a way that sameness of number and one-one correspondence are merely necessary equivalents? I think the Neo-Fregeans can meet this response reiterating the point that one makes a conceptual mistake if one thinks that two properties have the same number but are not in one-one correspondence.

The Neo-Fregean appears on strong ground in his case against Benacerraf's principle. For it seems correct that the only way of obtaining a simple referring term such as 'the number of …' that is appropriately conceptually linked to one-one correspondence is via Hume's principle.

But the considerations contain a hidden assumption that tends to be disguised by the usual formulation of Hume's principle as it is in (4) and (5). The assumption is that the sense of the expression 'the number of …', is a simple one has a simple logical form that is correctly analysed by a single term former such as $nx.(…)$ or $hx.(…)$. An alternative position is that 'the number of F' actually has a more complex logical form which describes an assignment of abstract objects (obtained from elsewhere) to concepts via one-one correspondence. Applied to Benacerraf's principle, the more complex logical form of 'the number of F' would be:

> The Benacerraf number the predecessors of which
> are in one-one correspondence with F $\qquad\qquad$ (16)

so that a sentence like 'there are n apples' is analysed as

> The Benacerraf number the predecessors of which
> are in one-one correspondence with the apples, is n_B $\qquad$ (17)

(where 'the predecessors of which' is formalised in terms of Pre_B^* the transitive-reflexive closure of the predecessor relation on Benacerraf numbers).

With this analysis there is not only a necessary connection between sameness of number and one-one correspondence, but also a conceptual one. The conceptual connection between sameness of number and one-one correspondence is to be found, not in the means of abstracting numbers, but in the complex logical form of the natural language expression 'the number of …'.

So we have two theses that seem to account equally well for the conceptual connection between one-one correspondence and sameness of number.

I. 'the number of …' has a simple sense implicitly tied to one-one correspondence by Hume's principle.
II. 'the number of …' has a complex sense that ties explicitly one-one correspondence and Benacerraf's numbers.

The first thesis is the Neo-Fregean position, the second is the Benacerraf-number alternative. Both theses have cardinal number assertions conceptually tied to one-one correspondence. According to both theses a central, fundamental and standard use of numbers is to be assigned to concepts or properties according to how many things fall under them. The difference is that in Thesis I the assignment is implicit in the logical form of the number assertion, but in Thesis II it is explicit in the meaning of the phrase 'the number of …'.

The difference between I and II is sufficiently slight that an example is in order here. Consider a grocery stall at which different fruit are on sale. The manager of the stall has assigned numbers to each type of fruit for the purposes of carrying out an inventory. Now, there are at least two readings of the manager's assertion when he says:

The apples are 30 (18)

On one reading the manager is describing how many apples there are, on another reading he is describing what number they are assigned in his inventory. According to I the two readings have different logical forms, one standard, the other nonstandard. For example, according to Thesis I the standard reading has the form of a simple identity:

$nx.Apple(x) = 30$ (19)

Whereas the nonstandard reading has a more complex form like:

The number assigned by relation R to $Apple(x)$ is 30 (20)

According to II however, both readings have the same more complex form, perhaps like (17), but where the relation R is standard in the first (one-one correspondence) and nonstandard in the second (an inventory assignment).

This example should highlight the difference between I and II, it should also indicate that this difference is subtle and that intuitions in favour of one might as well favour the other. I doubt that any psycho-linguistic test indicate conclusively whether our standard assertions of cardinality really are straightforward identities like (16), or slightly more complex predications like (17). So I claim that both positions, I and II, are equally tenable and there is nothing to choose between them. Therefore I maintain that the attack on Neo-Fregeanism of Section 3.3 stands. There is nothing in our practices that fixes whether we obtain our number concepts and language using Hume's principle following Thesis I, rather than using Benacerraf's principle following Thesis II. So I conclude that, in fact, we do neither.

4.2 How to avoid bad company

Formally, Hume's and Benacerraf's principles are acceptable abstraction principles. I argue for this here by presenting a condition on good abstraction principles (i.e. I offer a solution to the bad company problem) and show that both abstraction principles satisfy it.

A famous worry for the Neo-Fregean project, called the 'bad company' problem, relates to the fact that not all abstraction principles are consistent. A famously inconsistent abstraction principle is Frege's notorious Basic Law V:

$$\forall F \forall G\, [\varepsilon x.Fx = \varepsilon x.Gx \leftrightarrow \forall x(Fx \leftrightarrow Gx)] \tag{21}$$

We can use (21) to derive Russell's paradox. An argument of Heck [Heck 1992] shows that there are many undesirable abstraction principles. For example, there are many Φ for which the abstraction principle:

$$\forall F \forall G\, [\varepsilon x.Fx = \varepsilon x.Gx \leftrightarrow \Phi \vee \forall x(Fx \leftrightarrow Gx)] \tag{22}$$

entails that Φ. It is not hard to find plenty of second order formulae Φ (some of which contain F and G) that are entailed by an abstraction principle like (19) which we would certainly think ought not to be true. Furthermore, different abstraction principles can be incompatible with each

other, although individually consistent; this is strange, as abstraction principles are supposed to be analytic and so ought to be true, and hence compatible, in any context. There is then a question whether some principle can be given to discern the acceptable abstraction principles from the unacceptable ones (see [Weir 2003] for many examples of unacceptable abstraction principles). I now present such a principle.

Let λ be any infinite cardinal, then the *consistency constraint* for λ is the condition that any abstraction principle should have the form:

$$\forall F \forall G \, [\varepsilon x.Fx = \varepsilon x.Gx \leftrightarrow \Psi_\lambda(F, G)] \tag{23}$$

Where

i. $\Psi_\lambda(F, G)$ is a second order sentence containing no free variables other than F and G, and also does contain the 'new' abstraction operator εx.

ii. Ψ_λ is a reflexive, transitive and symmetric relation on *unary* predicates. That is:
   – $\Psi_\lambda(F, F)$
   – $\Psi_\lambda(F, G)$ implies $\Psi_\lambda(G, F)$
   – $\Psi_\lambda(F, G)$ and $\Psi_\lambda(G, H)$ implies $\Psi_\lambda(F, H)$

iii. For any model M of cardinality λ, there are at most λ many valuations σ such that $\sigma(\Psi_\lambda(F, G)) = \bot$.[1]

Note that the familiar examples of 'bad' abstraction principles (e.g. in [Weir 2003]) violate this condition. For example in Frege's Basic Law V has the form

$$\forall F \forall G \, [\varepsilon x.Fx = \varepsilon x.Gx \leftrightarrow \forall x(Fx \leftrightarrow Gx)]$$

which clearly violates this condition for any λ. Note also that Benacerraf's principle and Hume's principle satisfy the consistency constraint for any infinite λ. Now we can show that any abstraction principle satisfying the consistency constraint for λ can be interpreted in *any* second order model

---

[1] This says that the (second order) property represented by Ψ_λ groups the properties of the domain into at most λ many different equivalence classes. In other words, Ψ_λ is only allowed to distinguish up to extensionality, all properties (i.e. subsets of $|M|$) of cardinality $< \lambda$. Ψ_λ must be unable to distinguish all but λ of the 2^λ properties of cardinality λ.

of cardinality λ. Let M_λ be a model (that can interpret the language of Ψ) with domain $|M_\lambda|$ of cardinality λ. Let R be a relation on properties (i.e. a relation on subsets of $|M_\lambda|$) such that

$$R(P, Q) \text{ iff } \sigma(\Psi_\lambda(F, G)) = T$$

for any valuation σ such that $\sigma(F) = P$ and $\sigma(G) = Q$.[2] In other words, R is the interpretation of Ψ_λ in the model M_λ. Since $\Psi_\lambda(F, G)$ contains no free first or second order variables other than F and G, R does not depend on σ.

If $P \subseteq |M_\lambda|$ then let $P_R = \{Q: R(P, Q)\}$. Clearly, P_R is an equivalence class. Now consider the set $A = \{P_R: P \subseteq |M_\lambda|\}$ and let μ be its cardinality. If $\mu > \lambda$, then there would be more than λ many valuations σ that falsify $\Psi_\lambda(F, G)$ (at least one for each of the μ-many pairs of different equivalence classes in A). This would violate condition (iii) of the consistency constraint for λ. So $\mu \leq \lambda$, i.e. the cardinality of A is less than or equal to the cardinality of $|M_\lambda|$. It follows then, that there is an injection f from $\{P_R: P \subseteq |M_\lambda|\}$ into $|M_\lambda|$.

We may use f to identify elements $e_P \in |M_\lambda|$:

$$e_P = f(P_R) \tag{24}$$

It is now a straightforward matter to check that

$$e_P = e_Q \quad \text{iff} \quad R(P, Q)$$

It follows that we can extend any second order model M_λ of cardinality λ, with an abstraction principle satisfying the consistency constraint for λ: we define the e_P as in (24) and then extend the language of M_λ to include the abstraction operator $\varepsilon x.(\dots)$; then we extend the inductive defition of how valuations σ assign elements to terms using (25):

$$\sigma(\varepsilon x.\Phi) = e_{\{m:\ m \in |M_\lambda|\ \text{and}\ \sigma_{[x/m]}(\Phi) = T\}} \tag{25}$$

(where Φ is any second order formula).

---

[2] σ assigns elements of M_λ to first order variables and subsets of M_λ to second order variables; $\sigma_{[x/m]}$ is a valuation that agrees with σ on all variables except that it maps the variable x to m; σ also assigns, by means of an inductive definition, elements of M_λ to complex terms and truth values for formulae.

Let me describe this interpretation in English: Ψ_λ forms equivalence classes of properties; the conditions on Ψ_λ guarantee that there is a one-one function f from these equivalence classes into the domain $|M_\lambda|$ of M_λ; we interpret the referent of $\varepsilon x.Fx$ under valuation σ as the element e which the function f assigns to the equivalence class of properties that Ψ_λ forms from the extension of F.

We now have in (20) a general criterion for the legitimacy of abstraction principles. This criterion legitimates Benacerraf's principle *as well as* Hume's principle. The only difference between them being that extending a model to validate Benacerraf's principle is slightly more straightforward than Hume's principle. Say that an abstraction principle is *almost analytic* if it satisfies the consistency constraint for any infinite λ. It is now a matter of dispute whether the fact that λ has to be infinite detracts from the analyticity of Hume's principle and Benacerraf's principle. A point in favour of the Neo-Fregean programme is that we can give an independently motivated formal reasons for treating Hume's principle as analytic and rule out principles like Basic Law V. However, a point against the Neo-Fregean programme is that Benacerraf's principle also comes out as analytic, and the argument of Section 3.3 stands.

4.3 Distinguishing the abstracts

Perhaps Neo-Fregeans should not try to rule out the Benacerraf-numbers as legitimate references of our number language, but embrace them. There is nothing to stop a Neo-Fregean accepting that in the abstract realm there are at least two number-like sequences of abstract objects. A good line for a Neo-Fregean to take might be that the Hume-numbers are the referents of our numbers-as-cardinals language, whereas the Benacerraf-numbers are the referents of our numbers-as-ordinals language. A Neo-Fregean could then argue that there are two main uses of number language, perhaps even two concepts of number (cardinal and ordinal) and so see no reason to be worried if there are two collections of abstract entities associated with them. Indeed, such a result could be regarded as a success of the Neo-Fregean programme.

The problem is that the Benacerraf-numbers are *not* ordinals: Benacerraf's principle involves no characterisation of ordering or any criterion of position correspondence. To understand Benacerraf's principle we need nothing that is not needed to understand Hume's principle. There is nothing about Hume-numbers that rules them out as being ordinals, and there

is nothing about Benacerraf-numbers that rules them out as being cardinals. The concept of a unit is no less important to that of cardinality than the concept of one-one correspondence. Benacerraf's principle and Hume's principle each could be taken as allowing reference to the natural numbers *as cardinals*. But then Neo-Fregeanism must account for why our arithmetic language refers to the Hume-numbers rather than the Benacerraf-numbers (or vice versa). I have been arguing that that we stand in no significant relation to Hume's principle that we do not also stand in to Benacerraf's principle. So if the Neo-Fregean accepts that the two abstraction principles allow reference to different abstract entities, then he has made no progress overcoming the objection of Section 3.3.

5 CONCLUSION

I have argued that there is no particular abstraction principle that we can associate with the natural numbers. At least two similar, but formally distinct, abstraction principles are capable of lying at the heart of the Neo-Fregean programme. The principles are distinct enough that there is no natural way of equating the abstract objects they give reference to. The principles are however sufficiently similar that there is no principled criterion that identifies one over the other as 'the correct' abstraction principle for elementary arithmetic. I conclude that numbers are not the abstract objects referred to by either abstraction principle, or of any other abstraction principle. The point to emphasise here is that neither the Hume-numbers nor the Benacerraf-numbers are really *the natural numbers*. The whole Neo-Fregean framework of abstraction principles is just another way of generating sequences that *encode* the natural numbers. This conclusion is independent of questions regarding the metaphysics of abstraction and whether abstraction principles really refer to any abstract objects at all.

REFERENCES

Benacerraf, Paul 1965 "What Numbers Could Not Be", *The Philosophical Review*, 74(1).
Frege, Gottlob 1953 *The Foundations of Arithmetic*, tr. by J.L. Austin, Blackwell, Oxford.
Heck, Richard 1992 "On the Consistency of Second-order Contextual Definitions", *Noûs*, 26.

Weir, Alan 2003 "Neo-Fregeanism: An Embarrassment of Riches", *Notre Dame Journal of Formal Logic*, 44(1).

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

Wright, Crispin 2000 "Neo-Fregean Foundations for Real Analysis: Some Reflections on Frege's Constraint", *Notre Dame Journal of Formal Logic*, 41.