A dilemma for neo-Fregeanism

Robert Trueman

Abstract

Neo-Fregeans need their stipulation of Hume's Principle — $NxFx = NxGx \leftrightarrow \exists R(Fx\, 1-1_R\, Gx)$ — to do two things. First it must implicitly define the term-forming operator '$Nx...x...$', and second it must guarantee that Hume's Principle as a whole is true. I distinguish two senses in which the neo-Fregeans might 'stipulate' Hume's Principle, and argue that while one sort of stipulation fixes a meaning for '$Nx...x...$' and the other guarantees the truth of Hume's Principle, neither does both.

1. Introduction

If we help ourselves to a strong enough background logic, we can derive all of the Peano Axioms from *Hume's Principle*:

(HP) $NxFx = NxGx \leftrightarrow \exists R(Fx\, 1-1_R\, Gx)$[1]

This result is the foundation of Hale and Wright's *neo-Fregean logicism* (or 'neo-Fregeanism' for short). They claim that we can stipulate (HP) as an implicit definition of the term-forming operator '$Nx...x...$'.[2] If they are right, all of arithmetic is entailed by logic and definitions alone. This is not quite enough to win us logicism in the classical sense, but it is a *very* good consolation prize.

Clearly the neo-Fregeans are asking their stipulation of (HP) to do two things. First, it must fix a meaning for '$Nx...x...$'; otherwise it will fail as an implicit definition. Second, it must guarantee the truth of (HP); otherwise it will give us no reason to take the derivation of arithmetic from (HP) to be sound.[3] In this paper I will distinguish two senses in which the neo-Fregeans might 'stipulate' (HP), and argue that while one sort of stipulation fixes a meaning for '$Nx...x...$' and the other guarantees the truth of (HP), neither does both.

[1] Unabbreviated: $NxFx = NxGx \leftrightarrow \exists R(\forall x(Fx \rightarrow \exists y(Gy \wedge Rxy)) \wedge \forall y(Gy \rightarrow \exists x(Fx \wedge Rxy)) \wedge \forall x\forall y\forall z((Rxy \wedge Rzy) \rightarrow (Fx \rightarrow (Fz \rightarrow (Gy \rightarrow x = z))) \wedge ((Rxy \wedge Rxz) \rightarrow (Fx \rightarrow (Gy \rightarrow (Gz \rightarrow y = z))))))$.

[2] This is clear in all of the neo-Fregean writings, but for an array of examples, see: Wright 1983: ch. 3, 1997: §I, 1999: 323–4; Hale 1997: 98; Hale and Wright 2000: esp. §6, 2009a.

[3] Some care is needed here. Although the stipulation of (HP) is meant to guarantee the truth of (HP), the neo-Fregeans do not claim to have certain, indefeasible, knowledge of the truth of (HP). That is because they cannot be certain that they have successfully stipulated (HP) — that will depend on a number of factors relating to bad company objections. For a neo-Fregean discussion of what sort of warrant we can have for the truth of (HP), see: Wright forthcoming.

2. Neo-Fregeanism vs. analytical logicism

The above characterisation of neo-Fregeanism is a thumbnail sketch, and at various points I will pause to fill in some further details. However, it is sufficiently rich to allow me now to distinguish neo-Fregeanism from what I will call *analytical logicism*. Neo-Fregeanism is often loosely characterised as the thesis that (HP) is an analytic truth. However, the neo-Fregeans have made it clear that their real contention is that (HP) can be stipulated as an implicit definition of '$Nx...x...$'; if that makes (HP) analytic then so be it, but that is at best a secondary matter.[4] The analytical logicist, on the other hand, puts the claim that (HP) is analytic front and centre. For her, (HP) is not a definition of any sort. Rather, we come to (HP) with an antecedent understanding of its two sides, and use that understanding to recognise that (HP) is an analytic truth. It seems to me that the line between neo-Fregeanism and analytical logicism has frequently been blurred; indeed, although Wright is aware of this difference, he has not always signalled it as clearly as he might.[5]

To be clear, this paper is targeted exclusively at neo-Fregeanism as I have laid it out. There is good reason to focus on neo-Fregean logicism rather than its close analytical cousin. Neo-Fregeanism is meant to speak to two of the most difficult questions in the philosophy of mathematics: how can we know anything about numbers, or for that matter, how can we even refer to them at all?[6] If the neo-Fregeans are right, these questions receive easy answers. No great mysteries surround our epistemic or semantic grasp of (HP). (HP) is true by stipulation, and serves to implicitly define our talk of numbers. We can, then, use (HP) to convert our unproblematic talk and knowledge of one property's being equinumerous with another into talk and knowledge of one number's being identical with another. Analytical logicism, on the other hand, can offer no immediate answers to these questions: it relies on an antecedent ability to talk about numbers and recognise the truth of number theoretic statements in order to spot (HP) as an analytic truth. It is because neo-Fregeanism makes these promises that it deserves to be singled out.

3. Subsentential stipulation

When the neo-Fregeans stipulate (HP) as an implicit definition of '$Nx...x...$', just what are they doing? On one understanding, they are issuing the following resolution:

> '$Nx...x...$' is to refer to whatever function it needs to refer to for (HP) to be true

[4] Wright 1999: 320–4.
[5] Wright 1983: 68–9 & 106.
[6] Wright 1983: §i; Hale and Wright 2001a: 10–1, 2002: 114–9, 2009b: 178–80.

As this stipulation primarily concerns a part of (HP) — '$Nx...x...$' — rather than the principle as a whole, I will call it the *subsentential* stipulation of (HP).[7] The subsentential stipulation of (HP) has (at least) two things going for it. First, it is an implicit definition of a very familiar form: it fits neatly into the general account of implicit definitions given by Lewis in his seminal (1970). Second, there can be no serious doubt that a subsentential stipulation of (HP) manages to fix a meaning for '$Nx...x...$': it gives us a criterion for deciding of a given function whether it is the function referred to by '$Nx...x...$'. Such a stipulation could no more fail to fix a meaning for '$Nx...x...$' than the stipulation,

> 'Jack the Ripper' is to refer to the person who committed the Whitechapel murders

could fail to fix a meaning for 'Jack the Ripper'.[8]

All of this speaks in favour of taking the neo-Fregean stipulation of (HP) to be a subsentential stipulation. But recall, the neo-Fregeans need their stipulation to do two things: first fix a meaning for '$Nx...x...$', and second guarantee that (HP) is true. And however well a subsentential stipulation does at the first of these tasks, it is utterly useless when it comes to the second. A subsentential stipulation of (HP) instructs us to assign whatever function we must to '$Nx...x...$' for (HP) to be true. In order to understand this stipulation, then, we must have a domain of functions in front of us. Moreover, this must be a domain of functions with which we are familiar *before* receiving the subsentential stipulation of (HP); we need to have a domain of functions in view in order to so much as understand the stipulation. What the subsentential stipulation of (HP) is telling us to do is look through that antecedently available domain until we find the function that '$Nx...x...$' must refer to for (HP) to be true. But when we stipulate (HP) in this spirit, we must be open to the possibility that *none* of the functions in the domain already available to us would make (HP) true.[9] Indeed, this is a point on which Hale and Wright themselves are very clear:

> It goes with the [subsentential] model that it must be at least initially intelligible that a principle proposed in this spirit fails to hit off reference to *anything.* It cannot just be a given that reference is secured, even if it is. (Hale and Wright 2009b: 206)[10]

[7]I will also say that someone who gives the subsentential stipulation of (HP) is stipulating (HP) *subsententially.*

[8]I say that there can be no serious doubt on this issue, but in fact Horwich (1998: 136–7) has expressed a general doubt about how stipulations of this form are meant to work. However, like Hale and Wright (2000: 124), I find myself unable to understand the nature of Horwich's worry.

[9]Equally, we may find that more than one function would make (HP) true. However, issues regarding uniqueness are not my concern here, and I will set them to one side.

[10]At this point I must issue a word of interpretative caution. In the above quotation, Hale and Wright are not concerned with exactly the same sort of stipulation of (HP) as I am concerned with. Rather, they (2009b: 206) are concerned with someone who takes the stipulation of (HP) to be "imposing a condition, viz. association with the elements in the field of the abstractive relation in a fashion isomorphic to the partition into equivalence classes which it effects, which it is then up to the world to produce a range of

Really, this should come as no surprise. As we already noted, the subsentential stipulation of (HP) fits into Lewis's general account of implicit definitions. Suppose we have a theory '$T(t)$' containing one theoretical term, 't'. According to Lewis's (1970: esp. 431–5), we can define 't' by stipulating that 't' should refer to the thing which satisfies '$T(x)$'. And it has *always* been understood that such a stipulation does not guarantee the truth of '$T(t)$'; rather, it guarantees the truth of T's *Carnap sentence*:

$$\exists x(Tx) \to Tt$$[11]

So, to take a concrete example, suppose our theory has one axiom, 'Jack the Ripper committed the Whitechapel murders', and we want to define 'Jack the Ripper' with this theory. Lewis would have us stipulate that 'Jack the Ripper' is to refer to the person who satisfies 'x committed the Whitechapel murders'. Such a stipulation obviously does not guarantee that 'Jack the Ripper committed the Whitechapel murders' is true. After all, it may be that no one person committed all of those murders. What the stipulation really guarantees is the truth of the following conditional:

> If someone committed the Whitechapel murders, then Jack the Ripper did

A subsentential stipulation of (HP) is a stipulation in Lewis's mould, and so is likewise powerless to guarantee the truth of (HP). What it really guarantees is the truth of the Carnap sentence for (HP):

$$\text{(HP}'\text{)}\quad \exists\eta\forall F\forall G(\eta xFx = \eta xGx \leftrightarrow \exists R(Fx\,1-1_R\,Gx)) \to \\ \forall F\forall G(NxFx = NxGx \leftrightarrow \exists R(Fx\,1-1_R\,Gx))$$

We obviously cannot derive the Peano Axioms from (HP′) alone; we can do so only with the help of an additional existential assumption telling us that some function satisfies '$\forall F\forall G(\eta xFx = \eta xGx \leftrightarrow \exists R(Fx\,1-1_R\,Gx))$'. And this assumption is by no means trivial: the existence of such a function requires the existence of infinitely many objects. So how, then, are we to justify this additional existential assumption?

There are a number of ways that we might try to convince ourselves that some function satisfies '$\forall F\forall G(\eta xFx = \eta xGx \leftrightarrow \exists R(Fx\,1-1_R\,Gx))$'. We could appeal to some background mathematical theory; the most obvious choice would be set theory. Alternatively, if we are physically minded, we could turn to the natural sciences; physics appears to make use of all sorts of functions, and we might hope that one of them will satisfy '$\forall F\forall G(\eta xFx = \eta xGx \leftrightarrow \exists R(Fx\,1-1_R\,Gx))$'. Or another alternative, if we are metaphysically minded then we could reach for a general ontological theory; for example, *maximalism*, the thesis that whatever could exist does exist, will make quick work of any and all ontological anxieties.

objects to satisfy". As far as I can tell, there is no important difference between this sort of stipulation and my subsentential stipulation of (HP).

[11] Again, I am disregarding issues surrounding uniqueness.

But none of these options is really very enticing for a neo-Fregean *logicist*. If the neo-Fregeans are to continue to deserve their 'logicist' title, whatever theory they appeal to to justify the assumption that some function satisfies '$\forall F\forall G(\eta xFx = \eta xGx \leftrightarrow \exists R(Fx\,1-1_R\,Gx))$' must admit of a logicist treatment. Some theories have a better a chance of being given a logicist treatment than others. I take it to be obvious that no logicist accounts of the natural sciences or of heavyweight metaphysical positions like maximalism are on the table, but some form of logicist set theory is still a going concern. However, the best logicist set theories are *neo-Fregean* set theories. They attempt to implicitly define the operator '$\{x : ...x...\}$' with some consistent version or other of *Basic Law V*:

(V) $\{x : Fx\} = \{x : Gx\} \leftrightarrow \forall x(Fx \leftrightarrow Gx)$

And if this stipulation is also given subsententially, i.e. if what is stipulated is that '$\{x : ...x...\}$' is to refer to whatever it needs to refer to for the consistent version of (V) to be true, then clearly the neo-Fregeans' problem has merely been delayed, not solved.

It is important to be completely clear on what the point is here. At one time it was common for philosophers to respond to neo-Fregeanism by denying that (HP) could be true by stipulation. Perhaps the most forceful presentation of this response was given by Boolos (1997). Boolos's thought was simple: (HP) is ontologically committed to infinitely many objects, and nothing which carries such heavy ontological baggage can be true by stipulation; what can be made true by stipulation is a conditional like (HP′), which by itself has no special ontological commitments.[12] Now, I have no desire to assess Boolos's objection, or Wright's (1999) response to it, here. I raise it only to distance myself from it. I am not just stubbornly insisting that Boolos was right and we can only make (HP′) true by stipulation. Rather, the point is that all that the *subsentential* stipulation of (HP) guarantees to be true is (HP′). It may be that there are *other* sorts of stipulation that the neo-Fregeans could offer to make (HP) true; indeed, the vast majority of this paper will be spent considering one candidate alternative.

We can make things clearer with the help of a distinction first introduced by Hale (1997: 105–6). Hale asks us to keep in mind the difference between the *import* of a stipulation — what is stipulated — and the *intended effects* of that stipulation. Of course, the neo-Fregeans hope that one of the effects of their stipulation will be that '$Nx...x...$' comes to refer to whatever function it needs to refer to for (HP) to be true. When we stipulate (HP) subsententially, we try to secure this effect in the simplest way possible: by making it the import of the stipulation. The problem with this direct strategy is that it is powerless to guarantee that (HP) is actually true. The reason for this is that the subsentential stipulation of (HP) tells us to look through a domain of functions which we already had available to us *before* receiving that stipulation. What the neo-Fregeans require, then, is that there be some way of stipulating (HP) which does not just refer us

[12]See especially: Boolos 1997: 250–4.

back to an antecedently accessible domain of functions, but which *itself* introduces us to the function which makes (HP) true. Again, this is a point on which Hale and Wright are themselves clear. They tell us that stipulating (HP) should be a matter of inventing a meaning for '$Nx...x...$',

> where to invent a meaning, we suggest, can only be to bring it about that some expression have a novel but intelligible pattern of use, perhaps unmatched by any expression either already available in the language or explicitly definable by means of such. To invent a meaning, so conceived, is to fashion a concept: it is to be compared to making a mould and then fixing a certain shape-concept by stipulating that its instances comprise just those objects which fit the mould (or are of the same shape as something which does). (Hale and Wright 2000: 131)[13]

4. Sentential stipulation

The neo-Fregeans are trying to keep two plates spinning. They want their stipulation of (HP) to both fix a meaning for '$Nx...x...$' and to guarantee that (HP) as a whole is true. When we imagined the neo-Fregean giving a subsentential stipulation of (HP), we imagined them focusing on the first of these spinning plates; the difficulty for them was that in doing so, the second slipped and fell. What, then, if they change their focus, so that their primary concern is guaranteeing the truth of (HP)? Well, as far as I can tell, such a change of focus would involve taking the stipulation of (HP) to be the following resolution:

> '$NxFx = NxGx$' is to have the same truth-value as '$\exists R(F\ 1-1_R\ G)$' on every assignment of values to 'F' and 'G'[14]

Since this stipulation is primarily concerned with guaranteeing the truth of the whole sentence (HP), I will call it the *sentential* stipulation of (HP).[15]

I take it as given that the sentential stipulation of (HP) does guarantee that (HP) is true. Now the question is whether it also manages to fix a meaning for '$Nx...x...$'. And in fact, it can quickly become difficult to see how a sentential stipulation of (HP) could fix a meaning for '$Nx...x...$'. After all, that stipulation tells us *nothing* about the inner

[13] In fact, Hale and Wright are not primarily concerned with (HP) in this passage, but with implicit definitions in general. However, they (2000: 142) make it clear that they consider their discussion to apply to the contextual of definition of '$Nx...x...$' with (HP).

[14] At various points it seems as though this is precisely how the neo-Fregeans intend their stipulation of (HP) to be understood. Here is just one example: in their (2002: 117) they tell us that "the *import* of the stipulation [of an abstraction principle] is simply that the corresponding instances of the left and right sides [...] are to be alike in truth-value, i.e. materially equivalent".

[15] I will also say that someone who gives the sentential stipulation of (HP) is stipulating (HP) *sententially*.

workings of '$NxFx = NxGx$'; it just tells us that the instances of '$NxFx = NxGx$' as a whole are to share their truth-values with the corresponding instances of '$\exists R(F\ 1{-}1_R\ G)$'. Surely, then, we have no right to read any complexity into '$NxFx = NxGx$' beyond the occurrences of the variables 'F' and 'G'. That is, it seems that a sentential stipulation of (HP) leaves '$NxFx = NxGx$' as nothing but an unstructured open sentence with 'F' and 'G' as variables, a mere abbreviation of '$\exists R(F\ 1-1_R\ G)$'.[16] Following Wright (1983: 68), I will say that when we refuse to discern syntactic complexity in '$NxFx = NxGx$' beyond its free variables, we are giving it, and (HP) along with it, an *austere* reading; this is to be contrasted with a *robust* reading, on which '$NxFx = NxGx$' has precisely the form it appears to.

The neo-Fregeans have been alive to the worry that they are entitled only to an austere reading of (HP) since the very beginnings of their project. Their general response to such a concern is to remind their opponent that '$NxFx = NxGx$' is not made up of wholly alien symbols. It features 'F', 'G' and '$=$', all of which are familiar. And if we read these familiar symbols in the ordinary way, we will have no choice but to give '$NxFx = NxGx$' the robust reading. Hale and Wright put it this way:

> What a recipient of (HP) immediately learns is that whatever suffices for the truth of a statement of concept-equinumerosity is equally sufficient for the truth of the corresponding statement of number-identity. However, she also understands that she is to take the surface syntax of number-identity at face value. She already possesses the general concept of identity, and so is able to recognise that the expressions flanking the identity sign must be singular terms. Further, she already understands predicate-variables, and so can recognise that '$Nx...x...$' must be being introduced as a functional expression, denoting a function from concepts to objects. From this she is able to gather that the objects in question simply are objects for whose identity it is necessary and sufficient just that the relevant concepts be equinumerous. (Hale and Wright 2002: 117–8)[17]

The neo-Fregeans are here insisting that the recipient of (HP) take the surface syntax of '$NxFx = NxGx$' at face value. Obviously no one doubts that 'F' and 'G' appear in '$NxFx = NxGx$' as predicate-variables: after all, the very same variables appear in '$\exists R(F\ 1-1_R\ G)$'. The real work is being done by the insistence that '$=$' appears as the familiar identity sign, carrying the meaning it carries in ordinary identity sentences.

[16] There is an analogy between this objection and Frege's objection to the attempt to introduce number terms via numerically definite quantifiers in §56 of the *Grundlagen*. For a discussion of Frege's objection that makes this analogy apparent, see: Dummett 1991: 101–3. For expressions of similar concerns, see: Heck 1997a: 129–30; MacBride 2003: 118–9.

[17] This quotation originally concerned the abstractionist definition of 'the direction of'; I have modified it to deal with (HP). For more examples of this line of thought, see: Wright 1983: 67–8 & 81–2, 1998b: 270–1; Hale 1997: 98.

Indeed, Frege (1884: 63) himself was clear that this was how abstraction was meant to work: he tells us that we are to use the concept of 'equality, taken as already known, to obtain that which is to be regarded as being equal'.

Are the neo-Fregeans free to sententially stipulate (HP) and then simply insist that we read '=' in '$NxFx = NxGx$' as the identity sign? Plenty of philosophers have thought not. These philosophers have typically justified their negative answer by presenting the following reductionist argument:

> If the instances of '$NxFx = NxGx$' have the structure they appear to, i.e. an identity sign flanked by two singular terms, then they make reference to numbers. The instances of '$\exists R(F\ 1-1_R\ G)$', on the other hand, make no reference to numbers. Neo-Fregeans frequently claim that, as a consequence of their stipulation, instances of '$NxFx = NxGx$' share a meaning with the corresponding instances of '$\exists R(F\ 1-1_R\ G)$'.[18] But no two sentences can share a meaning and yet make reference to different things! Hence, the apparent structure of '$NxFx = NxGx$' must be *merely* apparent.

In the existing literature, the austere reading of (HP) has been entirely bound up with this reductionist argument. As a consequence, neo-Fregeans tend to think that all they need to do to show that they can indulge in more than the austere reading is show that the reductionist argument does not work.[19] And it really is quite easy to do that. There are two crucial premises to the reductionist argument: first, that by neo-Fregean lights, instances of '$NxFx = NxGx$' have the same meaning as the corresponding instances of '$\exists R(F\ 1-1_R\ G)$'; second, that no two sentences can share a meaning and yet make reference to different things. Now, there is a sense of 'meaning' on which the first premise is true,[20] and a sense of 'meaning' on which the second is true. But what is not clear is that these two senses coincide. In the end, it may be that the neo-Fregeans can content themselves with saying that the corresponding instances of '$NxFx = NxGx$' and '$\exists R(F\ 1-1_R\ G)$' share a meaning only in the sense of necessarily sharing a truth-value.[21] And there is no obvious reason to suppose that two sentences cannot make reference to different things and yet share a 'meaning' in this sense.[22]

[18] To be clear: this is meant to be an effect, not the import, of the neo-Fregean stipulation. The import of the stipulation is the humble biconditional. See: Hale 1997: 105–6.

[19] One only needs to survey the literature to verify the truth of this claim, but for one example, see: Hale and Wright 2001b: 354.

[20] To give just a few examples: Hale 1994a: 192–7, 1997, 1999: 94; Wright 1997: 277–8; Hale and Wright 2000: 149–50, 2002: 118–23.

[21] In fact, Hale (2007: 110–4) seems tempted by an even more minimal conception of the shared meaning: instances of '$NxFx = NxGx$' and '$\exists R(F\ 1-1_R\ G)$' share a meaning only in the sense of sharing a truth-value. Importantly, the objection I develop (§5) against neo-Fregeanism thought of as involving the sentential stipulation of (HP) works on even this most minimal conception of shared meaning. For the extensive literature concerning the sense in which the two sides of (HP) share a meaning, also see: Dummett 1991: 168–76; Hale 1997, 1999: 89–98, 2001; Potter and Smiley 2001, 2002.

[22] Hale and Wright 2009b: §6.

However, this reductionist argument is not the only possible argument we might give for thinking that the neo-Fregeans cannot sententially stipulate (HP) and then airily insist that we read '=' in '$NxFx = NxGx$' as the identity sign. My major aim in the rest of this paper is to develop an alternative. But before actually doing so, it is important to mention just how limited my ambitions are. If the reductionist argument had worked, it would have shown that *any* position according to which the left and right sides of (HP) share a meaning, including analytical logicism, is false. Indeed, the ambitions of many reductionists run further still: they want to argue that *in general*, what appear to be terms referring to numbers are not really terms at all. I, on the other hand, have no quarrel with analytical logicism — perhaps (HP), understood as robustly as one likes, really is an analytic truth. And I certainly do not want to argue that there are no numerical terms! All I want to challenge is the thesis that we can sententially stipulate (HP) and then insist that we read '=' in '$NxFx = NxGx$' as the identity sign; that is, I will argue that *if* we sententially stipulate (HP), *then* we have no choice but to read '$NxFx = NxGx$' austerely.[23]

5. Truth and reference

To see why we cannot sententially stipulate (HP) and then insist that we read '=' as the identity sign in '$NxFx = NxGx$', we must consider the general relationship between sentential and subsentential meaning. For some reason it is much easier to think about what singular terms mean than it is to think about what predicates mean, so for ease let's start by examining the relationship between terms and sentences.

Suppose I introduce a new sentence, 'F(Socrates)', and fix its truth-value by simply stipulating that it is to be true. Are we now free to read 'Socrates' as it appears in 'F(Socrates)' as a name of Socrates? As Hale and Wright (2000: 134) note, there is one obvious problem with trying to do so. If 'Socrates' is a singular term, then we should be able to intelligibly substitute other terms for it. But our stipulation gives us no inkling of the truth-conditions for 'F(Plato)'.

However, that problem is a mere symptom of a much deeper lying one. Now suppose that I introduce the sentence 'G(Socrates)' by stipulating that 'G' is to be true of an object iff that object committed suicide. The truth-value of 'G(Socrates)' is in part determined by the reference of 'Socrates': it is true just in case the referent of 'Socrates', i.e. Socrates, committed suicide. In order to get from the truth-conditions of 'G(Socrates)' to 'G(Plato)', we need only swap out the reference to Socrates for a reference to Plato. But matters were wholly different in the case 'F(Socrates)'. No reference to Socrates

[23] I can, then, agree with Heck (2000: §8.3) when he argues that in order to understand the use of quantifiers like 'most' in relation to numbers, we must accept that there are genuine numerical terms. However, what Heck's argument does not show, and what I will challenge, is that we could introduce numerical terms by sententially stipulating (HP).

played any role in determining the truth-value of this sentence. Rather, the truth-value of ‘F(Socrates)’ was fixed outright. And this is why we cannot understand ‘F(Plato)’: when it comes to determining the truth-value of ‘F(Socrates)’, ‘Socrates’ is not a relevant unit ready to be replaced.

Why should this be considered a deeper problem than Hale and Wright’s? Because the link between reference and truth-value is non-optional. At least part of what it is for a given occurrence of ‘Socrates’ to refer to Socrates is for it to play a particular role in determining the truth-value of the sentence in which it features. So to admit that the apparent reference to Socrates in ‘F(Socrates)’ plays no role in determining the truth-value of that sentence is to admit that there is no such reference. ‘Socrates’ does not appear as a name of Socrates in ‘F(Socrates)’; in reality ‘F(Socrates)’ is nothing more than an oddly shaped sentence letter.[24]

There is an important lesson to draw from this discussion. When we fix a truth-value for a sentence, the way in which we do so settles what, if any, roles the parts of that sentence play in determining that truth-value. So if we want ‘Socrates’ to appear as a name of Socrates in a given sentence, we are thereby restricted in the ways in which we are free to fix a truth-value for that sentence. We must do so in a way that assigns the appropriate role to ‘Socrates’: the fact that ‘Socrates’ refers to Socrates must have a knock-on effect on the truth-value of the whole sentence. When we fixed a truth-value for ‘G(Socrates)’, we conformed to this restriction; when we fixed one for ‘F(Socrates)’, we did not.

So far we have been concerned with a toy case. What we are really interested in is whether the neo-Fregeans can insist that we read ‘=’ in ‘$NxFx = NxGx$’ as the familiar identity sign after sententially stipulating (HP). And there is an important difference between our real interest and our toy case. In the toy case we were wondering whether a given expression, ‘Socrates’, was functioning as a particular *singular term* in the sentence ‘F(Socrates)’; in the neo-Fregean case, we are wondering whether a given expression, ‘=’, is functioning as a particular *predicate* in the (open) sentence ‘$NxFx = NxGx$’. There are obviously important differences between the meanings of terms and the meanings of predicates, but there are also sufficient similarities for us to model our discussion of the neo-Fregean case on our discussion of the toy case.

To begin with, one problem with the neo-Fregeans’ attempt to read ‘=’ in ‘$NxFx = NxGx$’ as the identity sign jumps out. If ‘=’ appears as the identity sign in, for example, ‘$Nx(x \neq x) = Nx(x \neq x)$’, then ‘$Nx(x \neq x) =$’ appears as a complex predicate. And in that case we should be able to intelligibly substitute other predicates for it. But a sentential stipulation of (HP) gives us no inkling of the truth-conditions for ‘Julius

[24]The above argument is, I believe, what Wittgenstein (1974: 315) was getting at with his cryptic remark that it is a mistake to think that you can use a painting as a mirror as well, even if only for a single posture. Wittgenstein delivered this remark as part of an objection to Ramsey’s logicism. For a discussion of that objection, see: Trueman 2011.

Caesar $= Nx(x \neq x)$'. This is, of course, a version of the familiar Julius Caesar Problem.

But just as in our toy case, this problem is a mere symptom of a deeper lying one. When we sententially stipulate (HP), we are not *directly* stipulating which truth-values the various instances of '$NxFx = NxGx$' are to have. But this stipulation is meant to *indirectly* fix those truth-values: each instance of '$NxFx = NxGx$' is to have the same truth-value as the corresponding instance of '$\exists R(F\ 1-1_R\ G)$'. And when they are fixed in this way, the supposed fact that '$=$' appears as the identity sign in the instances of '$NxFx = NxGx$' simply plays no role in determining the truth-values of those instances. By the same token, the supposed complex predicate '$Nx(x \neq x) =$' plays no role in determining the truth-value of '$Nx(x \neq x) = Nx(x \neq x)$'. That is why we have no grasp of the truth-conditions for 'Julius Caesar $= Nx(x \neq x)$': when it comes to determining the truth-value of '$Nx(x \neq x) = Nx(x \neq x)$', '$Nx(x \neq x) =$' is simply not a relevant unit ready to be replaced.

Why should this problem be considered deeper than the familiar Caesar problem? Well, at least part of what it is for a given occurrence of '$=$' to play the role of the identity sign is for it to play a particular role in determining the truth-value of the sentence in which it features.[25] So to admit that '$=$' does not play any role in determining the truth-values of the instances of '$NxFx = NxGx$' is to admit that it does not appear in them as the identity sign.

It is important to see how this argument differs from the reductionist one outlined in §4. That argument went wrong because it assumed that the neo-Fregeans wanted '$NxFx = NxGx$' and '$\exists R(F\ 1{-}1_R\ G)$' to share a meaning in some quite demanding sense. However, the only notion of sentential meaning relevant to my argument is *truth-value*. Of course, the mere fact that the instances of '$NxFx = NxGx$' and '$\exists R(F\ 1-1_R\ G)$' share their truth-values does not suffice to to show that '$=$' does not appear as the identity sign in '$NxFx = NxGx$'. But that is not the point. The point is that if we want '$=$' to appear as the identity sign in '$NxFx = NxGx$', we are thereby restricted in the *ways* in which we are free to fix truth-values for those instances. We must do so in a way that assigns the appropriate role to '$=$'. When the neo-Fregeans fix the truth-values of the instances of '$NxFx = NxGx$' by sententially stipulating (HP), they do not conform to this restriction.

6. Inferentialism

It has often been pointed out that rather than thinking of Hume's Principle as a biconditional, we can think of it as a pair of introduction and elimination rules:[26]

[25] The difference between terms and predicates comes up in the different sorts of roles they play.

[26] In fact, Wright has recently (forthcoming: §13) insisted that, for neo-Fregean purposes, this is the best way to think of Hume's Principle.

$$
\begin{array}{l|ll}
i & \exists R(Fx\,1-1_R\,Gx) & \\
j & NxFx = NxGx & \text{HP-I}, i
\end{array}
\qquad
\begin{array}{l|ll}
i & NxFx = NxGx & \\
j & \exists R(Fx\,1-1_R\,Gx) & \text{HP-E}, i
\end{array}
$$

In this section I will quickly pause to show that the attempt to implicitly define '$Nx...x...$' by laying down these rules is open to exactly the same objection as the attempt to implicitly define it by sententially stipulating (HP). For my purposes, then, this inferentialist spin on neo-Fregeanism does not offer an interesting alternative to a neo-Fregeanism based on the sentential stipulation of (HP).

The idea that we can implicitly define certain expressions via their introduction and elimination rules is most familiar when it is applied to the logical constants. Take for example the material conditional, which has the following introduction and elimination rules:

$$
\begin{array}{l|ll}
i & \quad \begin{array}{|l} P \\ \hline \end{array} & \\
j & \quad \begin{array}{|l} Q \end{array} & \\
k & P \to Q & \to\text{I}, i\text{–}j
\end{array}
\qquad
\begin{array}{l|ll}
i & P \to Q & \\
j & P & \\
k & Q & \to\text{E}, i, j
\end{array}
$$

According to *logical inferentialism*, these two rules implicitly define '$\to$'. Hale and Wright (2000: e.g. 130) have been keen to emphasise the analogy between the neo-Fregean definition of '$Nx...x...$' and the inferentialist definition of '$\to$'. As we will see, however, this analogy gives out a crucial point. But before explaining exactly where, it will be useful to make a simplifying remark. According to inferentialism, we do not actually need to appeal to both ($\to$I) and ($\to$E) to implicitly define '$\to$'. ($\to$I) by itself will suffice; we can then infer that ($\to$E) is the elimination rule for '$\to$' on the grounds that it is harmonious with ($\to$I). Similarly, assuming that (HP-I) and (HP-E) are harmonious, which they had better be,[27] we can focus on (HP-I) and set (HP-E) to one side.[28]

($\to$I) tells us what the (canonical)[29] grounds are for introducing '$P \to Q$': we may introduce that sentence whenever we can infer Q on the supposition of P. There is a certain affinity between '$P \to Q$' and the grounds for introducing it. '$P \to Q$' is meant to be read as a sentential connective, '$\to$', flanked by 'P' and 'Q'. And in stating the grounds for introducing this sentence, we use 'P' and 'Q'; speaking loosely, in order to introduce the *connective* '$\to$' between 'P' and 'Q', we must establish a *connection* between P and Q.[30]

[27] Hale and Wright (2000: 133 fn. 32) are aware that (HP-I) and (HP-E) are not harmonious on certain standard definitions of 'harmony', and offer an alternative one.

[28] Equally, if we preferred we could focus on ($\to$E) and (HP-E), or continue to keep all the rules in mind. I have chosen to focus on introduction rules only to make the following discussion simpler.

[29] In what follows I will drop this qualifier.

[30] As should be plain, I am using 'P' and 'Q' schematically throughout this discussion.

(HP-I), on the other hand, tells us what the grounds are for introducing '$NxFx = NxGx$': we may introduce that sentence whenever we can show that F and G are equinumerous. But '$NxFx = NxGx$' and the grounds for introducing it do not share the affinity mentioned above. '$NxFx = NxGx$' is meant to be read as a relational predicate, '$=$', flanked by '$NxFx$' and '$NxGx$', but in stating the grounds for introducing this sentence we do not use '$NxFx$' or '$NxGx$'. In order to introduce the *relational predicate* '$=$' between '$NxFx$' and '$NxGx$' we do not need to establish a *relation* between $NxFx$ and $NxGx$; rather, we must establish a relation, equinumerosity, between F and G.

Of course, this difference does not show us that ($\rightarrow$I) is any better than (HP-I) as an inference rule. But it is an important difference when these rules are taken as implicit definitions. For the inferentialist, the grounds for introducing '$P \rightarrow Q$' serve to fix a meaning for that sentence. But importantly, it fixes that meaning in terms of the meanings of the sentences 'P' and 'Q': we use those sentences in the course of giving the grounds for introducing '$P \rightarrow Q$'. So 'P' and 'Q' are singled out as playing the roles of sentences in '$P \rightarrow Q$'. Consequently, '$\rightarrow$' must appear as a two-place sentential connective. This is why, for the inferentialist, ($\rightarrow$I) is properly seen as implicitly defining a two-place connective.

For the neo-Fregean, (HP-I) fixes a meaning for '$NxFx = NxGx$', but nothing in that rule tells us that '$=$' is to be read in '$NxFx = NxGx$' as the identity sign or that '$NxFx$' and '$NxGx$' are to be read as terms. When we take (HP-I) as an implicit definition, we fix the meaning of '$NxFx = NxGx$' purely in terms of 'F' and 'G': they are the only parts of '$NxFx = NxGx$' which appear in the grounds for introducing that sentence. Consequently, (HP-I) does not license anything other than the austere reading of '$NxFx = NxGx$', according to which it is an unstructured open sentence with 'F' and 'G' as variables.[31]

Perhaps some neo-Fregeans may be tempted to reply to this objection as follows: 'Sure, (HP-I) does not *by itself* license anything but the austere reading of "$NxFx = NxGx$". But we are also meant to recognise the occurrence of the identity sign in that sentence. And when we do, we will be pushed into the robust reading!' Unfortunately, however, we have already seen (§5) what is wrong with this sort of response. If we want '$=$' to appear as the identity sign in '$NxFx = NxGx$', we are thereby restricted in the range of ways we are free to fix a meaning for that sentence; we must do so in a way that assigns the appropriate role to '$=$' in determining the truth-values of the instances of '$NxFx = NxGx$'. (HP-I) assigns no such role to '$=$'. The neo-Fregeans cannot now just insist that '$=$' is to play the role of the identity sign nonetheless.

[31]See (Więckowski 2011: 222) for a similar criticism of Prawitz's attempt to define atomic formulae with inferential rules.

7. The Syntactic Priority Thesis

So far in my discussion I have not touched on the neo-Fregeans' *Syntactic Priority Thesis*, which is often put as follows:

(SPT) If an expression behaves syntactically like singular term in (an appropriate range of) true sentences then it refers to an object[32]

One might think that (SPT) is the key to defending neo-Fregeanism, understood as involving the sentential stipulation of (HP), from my attack. Once the neo-Fregeans sententially stipulate (HP), there are bound to be true sentences featuring expressions of the form '$NxFx$'. Take, for example, '$Nx(x \neq x) = Nx(x \neq x)$': '$\exists R(x \neq x \ 1-1_R \ x \neq x)$' is a truth of second-order logic and so, once we have sententially stipulated (HP), '$Nx(x \neq x) = Nx(x \neq x)$' must be true. So long as '$Nx(x \neq x)$' behaves syntactically like a singular term in '$Nx(x \neq x) = Nx(x \neq x)$', (SPT) will guarantee that it behaves semantically like a term too.

However, this line of response misunderstands both Hale and Wright's position and my objection to it. First, although Hale and Wright do seem to have endorsed the above version of (SPT) at various points,[33] they (2003: 254) are clear that they do not do so now. Rather, they claim merely that if an expression behaves syntactically like a singular term in appropriate true sentences, then we have good, *but defeasible*, reason to think that it refers to an object.

Second, and more importantly, I have been arguing that when we sententially stipulate (HP), we have no choice but to give '$NxFx = NxGx$' an austere reading. And on such a reading, '$NxFx = NxGx$' is a *syntactically* unstructured open sentence with 'F' and 'G' as variables. Consequently, the upshot of my argument is that when we sententially stipulate (HP), '$Nx(x \neq x)$' does not even appear as a *syntactic* singular term in '$Nx(x \neq x) = Nx(x \neq x)$'. (SPT) simply comes in too late to do anything to overturn this argument. Indeed, it is worth noting that Wright (1983: 68) himself has emphasised that on the austere reading, '$Nx(x \neq x)$' does not appear even syntactically as a singular term in '$Nx(x \neq x) = Nx(x \neq x)$'. Moreover, to the best of my knowledge, Hale and Wright have never tried to respond to the austere reading with an appeal to (SPT).

It may be useful to go through this point a little more slowly. We begin with a question: Just what is it for an expression to be have syntactically like a singular term? It is obviously not enough that the the expression have a certain shape. 'Frege' is not a term because it is written with a capital letter. Rather, behaving syntactically like a

[32]There is a question here of how to characterise the 'appropriate range' of true sentences mentioned in the above principle. However, for neo-Fregean purposes it suffices that identity contexts fall within that range.

[33]For example: Wright 1983: 51–2.

singular term is a matter of fulfilling a certain kind of inferential role. Unsurprisingly, it has proven very hard to give an adequate description of that role.[34] Nonetheless, at a bare minimum we expect to be able to generalise into the place of a singular term. That is, the following seems like a good starting point for any syntactic account of singular terms:

> 't' behaves syntactically as a singular term in a (extensional, atomic) sentence '$A(t)$' only if that sentence entails 'There is something such that A(it)'

So now the question is: Can we existentially generalise for '$Nx(x \neq x)$' in '$Nx(x \neq x) = Nx(x \neq x)$'? And my answer is: Not when we sententially stipulate (HP)! When we sententially stipulate (HP), '$Nx(x \neq x)$' is not a significant unit in '$Nx(x \neq x) = Nx(x \neq x)$': '$Nx(x \neq x) = Nx(x \neq x)$' is really the result of substituting '$x \neq x$' into both the argument places in the unstructured '$NxFx = NxGx$'. And if '$Nx(x \neq x)$' is not a significant unit, we have no right to replace it with a variable.[35]

There are three issues to be clear on here. First, I am not claiming that in English there is no operator, 'the number of Fs' or whatever, such that we can legitimately infer 'Something is identical to the number of non-self-identical things' from 'The number of non-self-identical things is identical to the number of non-self-identical things'. To seriously entertain that opinion is to indulge in the sort of reductionism I tried to distance myself from at the end of §4. Rather, my claim is that *if* we sententially stipulate (HP), *then* we cannot existentially generalise for an occurrence of '$Nx(x \neq x)$' in '$Nx(x \neq x) = Nx(x \neq x)$'. Neo-Fregeanism is meant to serve as a logicist reconstruction of our arithmetical practices; the problem with a sentential stipulation of (HP) is that it falls short of offering a foundation for those practices.

Second, this problem would obviously not have come up if the neo-Fregeans had been allowed to sententially stipulate (HP) and then insist that we read '=' in '$NxFx = NxGx$' as the identity sign. If they had then '$Nx(x \neq x)$' would be a singular term and so our ordinary practice of existential generalisation could go to work on it. But as we saw in §5, when we sententially stipulate (HP), we are barred from reading '=' as the identity sign.

Third, I am not denying that the neo-Fregeans could define a meaning for '$\exists y(y = Nx(x \neq x))$' in such a way as to guarantee that it is entailed by '$Nx(x \neq x) = Nx(x \neq x)$'. An obvious starting point would be the following (sentential) stipulation:

$$\exists y(y = NxGx) \leftrightarrow \exists F \exists R(Fx\, 1 - 1_R\, Gx)^{36}$$

[34] For attempts to develop such a description, see: Dummett 1981: ch. 4; Wright 1983: §ix; Hale 1994b, 1996.

[35] Again, there is a strong analogy between my point here and Frege's in (1884: §56).

[36] Wright (1983: 29–30) offers an analogous stipulation for directions. This stipulation would obviously have to be generalised to cover all of the existential generalisations that the neo-Fregeans might want to make.

But the point is that if we do rig up a meaning for '$\exists y(y = Nx(x \neq x))$' in this sort of way, then the inference from '$Nx(x \neq x) = Nx(x \neq x)$' to '$\exists y(y = Nx(x \neq x))$' will not be an instance of (first-order) existential generalisation. We already have an understanding of such generalisations, and that understanding is already wholly general: it allows us to generalise in the place of *any singular term*. The fact that that rule is not already sufficient to allow the inference from '$Nx(x \neq x) = Nx(x \neq x)$' to '$\exists y(y = Nx(x \neq x))$' when we sententially stipulate (HP) shows that in that case, '$Nx(x \neq x)$' does not appear as a singular term in '$Nx(x \neq x) = Nx(x \neq x)$'.[37]

8. Tolerance and predicativity

The neo-Fregeans want their stipulation of (HP) to kill two birds with one stone: it should implicitly define '$Nx...x...$', and it should guarantee (HP) to be true. I have challenged the suggestion that *one* stipulation could do *both* of these things: the subsentential stipulation of (HP) fixes a meaning for '$Nx...x...$', but does not by itself ensure that (HP) is true; the sentential stipulation guarantees that (HP) is true, but it leaves us with an austere reading of '$NxFx = NxGx$' on which '$Nx...x...$' is not a significant unit at all. Of course, this challenge is not conclusive. There may be some third way of stipulating (HP) that I have not considered. However, I hope to have done enough to make the challenge pressing.

Through the course of this paper, we have encountered three familiar objections to neo-Fregeanism: Boolos's claim that a principle as ontologically rich as (HP) cannot be made true by stipulation; the reductionist argument purporting to show that the apparent logical form of '$NxFx = NxGx$' must be merely apparent; and lastly, the Julius Caesar objection. I hope that by presenting these objections in this new setting, I have gone some way to getting at what was right about them (without wishing to deny that there may be a great deal wrong with them). I would like to end the paper by connecting my argument with one last response to neo-Fregeanism: Dummett's *tolerant reductionism*.

According to the tolerant reductionist, neither the austere nor the robust readings of (HP) are appropriate. Rather, we should give that principle an intermediate reading on which the instances of '$NxFx$' are referring singular terms, but terms whose reference is not 'realistic' or 'robust'. Dummett attempts to explain what he means by this in a number of places,[38] but I think the clearest statement is contained in the following passage. He tells us that what it is for a notion of reference for a class of singular terms to be 'robust' is for

> their reference [to] be semantically operative. Whether a notion of reference for terms of a given range is semantically operative or semantically idle depends

[37] As Tim Button has pointed out to me, there is an interesting intersection between this point and a criticism Sellars (1956: §§8–9) offered of a certain attempt to introduce terms for sense-data.

[38] Dummett 1981: 494–505, 1991: 189–99, 1998: 384–6.

> on the sense we attach to a sentence containing a term of that range. In grasping its sense, we have a conception of the way in which its truth-value is determined. If the determination of the truth-value of any such sentence goes through the identification of the referent of the term, the notion of reference, as applied to it, is semantically operative; if it does not, that notion, even though legitimate, is semantically idle. If we follow Frege's stipulations of the *Bedeutung* of value-range terms, as stated in Part I of *Grundgesetze*, the notion of *Bedeutung*, as applied to them, is semantically idle. In order to determine the truth-value of a sentence containing such a term **t**, we have to invoke Frege's stipulation of the *Bedeutung* of the smallest term (or the truth-value of the smallest sentence) properly containing the term **t**; determining the *Bedeutung* of **t** itself is no part of the process, because no stipulation was made stating what that was to be. The same would hold good of numerical terms if they were introduced by a direct stipulation of (HP). Even a satisfactory vindication of such a use of an abstraction principle would not justify a realistic interpretation of the terms introduced by means of it. (Dummett 1998: 385)

Now, I said that this was Dummett at his clearest on this issue, but plainly, that is not very clear. There are at least two questions to ask. First, what exactly is Dummett's reason for thinking that when we introduce numerical terms in the neo-Fregean way, they are semantically idle? And second, does it make sense to talk about semantically idle expressions referring at all?

Let's start with the first question. Wright (1998b: 270) took Dummett to be arguing as follows: When we introduce numerical terms by stipulating (HP), we can determine the truth-values of instances of '$NxFx = NxGx$' without considering the semantic structure of those instances; all we have to do is determine the truth-values of the corresponding instances of '$\exists R(F\ 1-1_R\ G)$'. Wright responded to Dummett by first acknowledging that we can indeed determine the truth-values of the instances of '$NxFx = NxGx$' in this way, but then insisting that this datum alone is not enough to establish that numerical terms are semantically idle.

> To establish that, it ought to be necessary, rather, to show that the assignment of reference to [numerical] terms plays *no role in determining the content of the sentences in which they occur*—it's not enough just to consider sentences in which they *don't* occur and, trading on their stipulated equivalence in truth-conditions to ones in which they do, merely to make the point encapsulated in the datum. (Wright 1998b: 270)

Perhaps this is an adequate response to Dummett's point as Wright understood it. But I am far from sure that Wright's understanding is correct. He takes Dummett to be making a point about how we *can* determine the truth-values of the instances of '$NxFx = NxGx$'.

But in the passage I quoted, Dummett talks about how we *have to* establish the truth-values of those sentences. It is by no means obvious what point Dummett was making, but it seems plausible to think that Dummett assumed that the the neo-Fregeans were stipulating (HP) sententially, and was then issuing a condensed version of my objection to such a stipulation. When we stipulate (HP) sententially, we do not assign any role to '$Nx...x...$' or '$=$' in determining the truth-values of the instances of '$NxFx = NxGx$'. Consequently, the only way we have of determining the truth-value of those sentences is to check the truth-values of the corresponding instances of '$\exists R(F\ 1-1_R\ G)$'. And if this is right, if we cannot determine the truth-values of the instances of '$NxFx = NxGx$' by considering their inner structures, then this does seem enough to show that instances of '$NxFx$' are semantically idle.

We move now to the second question: Does it make sense to talk about semantically idle reference? Hale and Wright have argued forcefully that it does not,[39] and this is an issue on which we agree. Indeed, the argument I presented in §5 turned on the assumption that it does not (although that was not how I put things). There are all sorts of ways of trying to show that the idea of semantically idle reference is incoherent, but here is just one. Dummett (1998: 378) seems to think that so long as every sentence in which a given expression occurs has a determinate truth-value, then that expression will refer in *some* sense; the question then is just how robust that sense is. When we have such a thin notion of reference, we are forced give up on a number of platitudes, for example: in order to know what a given descriptive term refers to, it suffices to know what kind of thing it refers to and what a thing of that kind has to be like in order to qualify as its referent. But as Wright (1998b: 269) observes, this platitude goes 'so deep that nothing should count as a notion of reference at all, not even a thin one, unless similarly constrained by' it.

However, agreeing with Hale and Wright on this point, that there is no such thing as semantically idle reference, is not to agree that '$NxFx = NxGx$' should be given a robust reading. Indeed, it goes the other way. If the likes of '$Nx(x \neq x)$' really are semantically idle in the instances of '$NxFx = NxGx$', then we should give those instances austere readings. And the whole burden of §§4–7 of this paper was showing that when we sententially stipulate (HP), '$Nx(x \neq x)$' and its fellows are indeed semantically idle.

Of course, giving '$NxFx = NxGx$' an austere reading does not preclude tolerance of any kind. We can continue to use '$NxFx = NxGx$' and its instances. And as I acknowledged in the previous section, we can even use what *look* like existential generalisations in relation to them: we can rig up a meaning for '$\exists y(y = Nx(x \neq x))$' in such a way as to guarantee that it is entailed by '$Nx(x \neq x) = Nx(x \neq x)$'. We can even say things like ' "$Nx(x \neq x)$" refers to something', so long as that is taken as just a typographical

[39] Wright 1983: §x, 1998b: 269–70; Hale 1994a: §4.

alternative to '$\exists y(y = Nx(x \neq x))$'.[40]

But this tolerance does not extend indefinitely far. We can see its limits by reflecting on Dummett's other great concern about neo-Fregeanism: its impredicativity. (HP) is strong enough to entail all of arithmetic only when it is understood impredicatively. On such an understanding, we take it that the objects introduced to us by the left hand side of (HP) fall within the first-order quantifiers in the (unabbreviated) right hand side. It is by no means easy to pin down all of Dummett's worries about this aspect of neo-Fregeanism.[41] However, it is undoubtedly true that he saw a relation between his opposition to an impredicative understanding of (HP) and tolerant reductionism. Wright (1998a: 247–55) attempted to render the impredicative understanding of (HP) unproblematic by telling a story in which a character, Hero, comes to have such an understanding. In the course of giving that story, Wright appealed to the platitude mentioned above: in order to know what a given descriptive term refers to, it suffices to know what kind of thing it refers to and what a thing of that kind has to be like in order to qualify as its referent. Dummett (1998: 385–6) responded by insisting that when terms refer in a 'thin' sense, this platitude does not apply.

We have already remarked that there seems to be something incoherent in Dummett's position: that platitude runs so deep that any notion of reference should be constrained by it. However, we also noted that the way to resolve this incoherence is to collapse tolerant reductionism into the austere reading. On the austere reading, expressions like '$Nx(x \neq x)$' are not really terms, and sentences like '$\exists y(y = Nx(x \neq x))$' are not really first-order generalisations. Consequently, when we read the left hand side of (HP) austerely it does not introduce us to any objects to be fed into the quantifiers in the right hand side. An austere reading of (HP) is tediously predicative. So while a sentential stipulation of (HP) may guarantee the truth of that principle, it also drains the principle of the strength it needs to ground a neo-Fregean logicism.[42]

References

Boolos G. 1997. Is Hume's Principle analytic? In Heck 1997b, 245–61.

Chalmers D., Manley D. et al (eds). 2009. *Metametaphysics*. Oxford: Oxford University Press.

Dummett M. 1981. *Frege: Philosophy of Language*. 2nd edition. Londond: Duckworth.

———. 1991. *Frege: Philosophy of Mathematics*. London: Duckworth.

[40] Hale 1994a: 204.

[41] Dummett 1991: e.g. 226–7, 1998.

[42] This article went through a long and torturous development, and so there are a great many people to thank: Arif Ahmed, Dan Brigham, Tim Button, Philip Ebert, Owen Griffiths, Bob Hale, Luca Incurvati, Fraser MacBride, Steven Methven, Walter Pedriali, Michael Potter, Lukas Skiba, Peter Sullivan, Nathan Wildman, Crispin Wright, everyone at the 2011 *Logicism Today* conference, and enough anonymous referees to make plain just how long and torturous the development was.

———. 1998. Neo-Fregeans: In bad company? In Schirn 1998, 369–87.
Feigl H. & Scriven M. (eds). 1956. *Foundations of Science and the Concepts of Psychology and Psychoanalysis*. vol. I of *Minnesota Studies in the Philosophy of Science*. Minneapolis, MN: University of Minnesota Press.
Frege G. 1884. *Die Grundlagen der Arithmetik*. Breslau: Koebner.
Hale B. 1994a. Dummett's critique of Wright's attempt to resuscitate Frege. In Hale & Wright 2001a, 189–213.
———. 1994b. Singular terms (2). In Hale & Wright 2001a, 48–71.
———. 1996. Singular terms (1). In Hale & Wright 2001a, 31–47.
———. 1997. Grundlagen §64. In Hale & Wright 2001a, 94–116.
———. 1999. Intuition and reflection in arithmetic II. *Proceedings of the Aristotelian Society Supplementary Volume* 73: 75–98.
———. 2001. A response to Potter and Smiley. *Proceedings of the Aristotelian Society* 101: 339–58.
———. 2007. Kit Fine on *The Limits of Abstraction*. *Contemporary Perspectives on Logicism and the Foundations of Mathematics, Travaux de Logique du Centre de Recherches Sémiologiques* 18: 103–29.
Hale B. & Wright C. 2000. Implicit definition and the a priori. In Hale & Wright 2001a, 117–50.
———. 2001a. *The Reason's Proper Study*. Oxford: Oxford University Press.
———. 2001b. To bury Caesar. In Hale & Wright 2001a, 336–96.
———. 2002. Benacerraf's dilemma revisited. *European Journal of Philosophy* 10: 101–29.
———. 2003. Responses to commentators. *Philosophical Books* 44: 245–63.
———. 2009a. Focus restored: comments on John MacFarlane. *Synthese* 170: 457–82.
———. 2009b. The metaontology of abstraction. In Chalmers, Manley & Wasserman 2009, 178–212.
Heck R. 1997a. The Julius Caesar Objection. In Heck 2011, 127–55.
Heck R. (ed). 1997b. *Language, Thought and Logic: Essays in Honour of Michael Dummett*. Oxford: Clarendon Press.
Heck R. 2000. Syntactic reductionism. In Heck 2011, 180–99.
———. 2011. *Frege's Theorem*. Oxford: Oxford University Press.
Horwich P. 1998. *Meaning*. Oxford: Oxford University Press.
Lewis D. 1970. How to define theoretical terms. *The Journal of Philosophy* 67: 427–46.
MacBride F. 2003. Speaking with shadows: A study of neo-logicism. *The British Journal for the Philosophy of Science* 54: 103–63.
Potter M. & Smiley T. 2001. Abstraction by recarving. *Proceedings of the Aristotelian Society* 101: 327–38.
———. 2002. Recarving content: Hale's final proposal. *Proceedings of the Aristotelian Society* 102: 351–4.
Rossberg M. & Ebert P. (eds). Forthcoming. *Abstraction*. Oxford: Oxford University Press.
Schirn M. (ed). 1998. *The Philosophy of Mathematics Today*. Oxford: Oxford University Press.
Sellars W. 1956. Empiricism & the philosophy of mind. In Feigl & Scriven 1956, 253–329.
Trueman R. 2011. Propositional functions in extension. *Theoria* 77: 292–311.
Więckowski B. 2011. Rules for subatomic derivation. *The Review of Symbolic Logic* 4: 219–36.

Wittgenstein L. 1974. *Philosophical Grammar*. Oxford: Blackwell.

Wright C. 1983. *Frege's Conception of Numbers as Objects*. No. 2 in Scots Philosophical Monographs. Great Britain: Aberdeen University Press.

———. 1997. On the philosophical significance of Frege's Theorem. In Hale & Wright 2001a, 272–306.

———. 1998a. On the harmless impredicativity of $N^{=}$. In Hale & Wright 2001a, 229–55.

———. 1998b. Response to Dummett. In Hale & Wright 2001a, 256–71.

———. 1999. Is Hume's Principle analytic? In Hale & Wright 2001a, 307–32.

———. forthcoming. Abstraction and epistemic entitlement: on the epistemological status of Hume's Principle. In Rossberg & Ebert Forthcoming, .