Double Vision:
Two Questions about the Neo-Fregean Program*

John MacFarlane†

January 5, 2007

Abstract

Much of *The Reason's Proper Study* is devoted to defending the claim that simply by stipulating an abstraction principle for the "number-of" functor, we can simultaneously fix a meaning for this functor and acquire epistemic entitlement to the stipulated principle. In this paper, I argue that the semantic and epistemological principles Hale and Wright offer in defense of this claim may be too strong for their purposes. For if these principles are correct, it is hard to see why they do not justify platonist strategies that are not in any way "neo-Fregean," e.g. strategies that treat "the number of Fs" as a Russellian definite description rather than a singular term, or employ axioms that do not have the form of abstraction principles.

The philosophical heart of the neologicist program is the claim that by stipulating

HP $\forall F \forall G(N_x(Fx) = N_x(Gx) \equiv Eq_x(Fx, Gx))$,[1]

*This is a revised version of the paper I presented at an Author Meets Critics symposium on Bob Hale and Crispin Wright, *The Reason's Proper Study* (Oxford: Oxford University Press, 2001), at the 2003 Pacific Division APA meeting in San Francisco. I am grateful to Hale and Wright for their comments on that occasion and at a 2005 seminar in St. Andrews.

†Department of Philosophy, University of California, Berkeley. 314 Moses Hall, Berkeley, CA 94720–2390. E-mail: `jgm@uclink.berkeley.edu`.

[1] "$Eq_x(Fx, Gx)$" abbreviates a sentence of pure second-order logic stating that the Fs can be put in one-one correspondence with the Gs.

we can simultaneously fix a meaning for the "number-of" functor ("$N_x()$") and acquire epistemic entitlement to HP. I'll call this the Master Claim. If it is true, then the semantic and epistemological problems of platonism are solved. If "number-of" acquires a fully determinate meaning simply through its stipulated relation to the logical notion of equinumerosity, then we can explain how we manage to talk and think about numbers despite their causal isolation from us. And, if we are entitled to believe HP in virtue of its role as an implicit definition of "number-of," we can explain our entitlement to *all* of our arithmetical beliefs. For as Frege showed, the basic arithmetical primitives can be defined in terms of "number-of" in such a way that definitional equivalents of the Peano axioms can be derived from HP in second-order logic.[2]

Most critics of the neologicist program have sought (in one way or another) to discredit the Master Claim. I am going to take a somewhat different tack, one that remains neutral about the plausibility of the semantic and epistemological principles Hale and Wright offer in defense of the Master Claim. What I will argue is that *if* these principles are strong enough to vindicate the Master Claim, it is hard to see why they do not also vindicate platonist approaches to arithmetic that are in no sense neo-Fregean: approaches that take "the number of Fs" to be a quantifier rather than a singular term, or that directly stipulate the Peano axioms instead of deriving them from HP. Thus, when I try to focus on the philosophy of arithmetic advocated in *The Reason's Proper Study*, I see a double image. Depending on how I squint, I can see it either as a defensible form of platonism or as a distinctively neo-Fregean one, but I can't seem to see it as both defensible *and* distinctively neo-Fregean at the same time.

In what follows, I am going to articulate two questions for Wright and Hale—one about singular terms, one about implicit definitions—with the hope that their answers will help us get a single, clear image of the neo-Fregean program and its significance.

1 Numerical definite descriptions

Hale and Wright take definite descriptions like "the number of cows" to be *singular terms*: they put them in a semantic class with "Jake Barnes"

[2] I will assume, as Hale and Wright do, that such derivation preserves entitlement.

and "Australia."[3] Here they follow Frege, but after Russell this decision surely requires some defense. Russell argued that "the number of cows" is not a singular term at all, but a quantifier, like "all women," "a book," or "five men." This view is now held by many philosophers of language and linguists,[4] so it is remarkable that nowhere in RPS do Hale and Wright defend the view that definite descriptions are singular terms against the Russellian alternative. My first question is whether it is *essential* to their program to take "the number moons of Jupiter" to be a singular term, rather than a quantifier—and if so, why?

To be sure, there is quite a bit of discussion of singular terms in RPS. Hale and Wright point out that it is essential to the neo-Fregean strategy that singular terms be identifiable by a syntactic criterion, rather than as expressions whose semantic function is to refer to objects. For on the neo-Fregean view, the notion of *object* is posterior in the order of explanation to the notion of *singular term*: "... *objects*, as distinct from entities of other types... just are what (actual or possible) singular terms refer to."[5] Accordingly, the first two essays of the book are devoted to hammering out a syntactic and inferential criterion for singular termhood. But none of this either addresses or renders irrelevant the Russellian challenge. If we want to follow Frege's strategy of letting singular termhood be our guide to objecthood, then certain *reasons* for taking definite descriptions to be quantifiers are off limits. For example, our reason for denying that "the present King of France" is a singular term cannot be that there is no such object as the present King of France. But that does not mean that we are left with *no* grounds for taking "the present King of France" to be a quantifier. For there are plenty of purely syntactic and inferential motivations for the Russellian approach.

Let's start with syntax. Like quantifiers ("some women," "most brown dogs," etc.), and unlike proper names (even complex ones like "Mr. George P. Willoughby III"), descriptions are composed of a determiner ("the") and a common noun phrase, which may be arbitrarily complex:

[3]See, e.g. RPS, 128. On p. 130 Hale and Wright are explicit that the result of completing a functional expression by supplying an argument is a definite description. On p. 154 Wright says that *many* ordinary definite descriptions will be singular terms, but it is not clear what exceptions are being allowed for here.

[4]For a powerful defense, see Stephen Neale, *Descriptions* (Cambridge: MIT Press, 1990).

[5]RPS, 8. See also Wright's *Frege's Conception of Numbers as Objects* (Aberdeen: Aberdeen University Press, 1983), 13, 53.

the/each castle
the/each castle in Spain
the/each castle in Spain where I wrote a novel
the/each castle in Spain where I wrote a novel that nobody read

Proper names, by contrast, typically lack syntactic structure or have rigid, non-recursive structures (first name + middle name + last name + suffix). Again, like quantifiers, and unlike proper names, definite descriptions have *scopes*. The sentence

> The professor who brought in the biggest grant in each of the last five years will be honored.

is ambiguous in a way that would be hard to explain without appealing to scope. It can mean either

> [each x: x is one of the last five years][the y: y is a professor who brought in the biggest grant in x](y will be honored)

or

> [the y: [each x: x is one of the last five years](y is a professor who brought in the biggest grant in x)](y will be honored).[6]

Similarly, the sentence

> The president of the United States will some day be Jewish.

exhibits a scope ambiguity that is precisely analogous to that of

> Most Scientologists will some day be rich.

A neo-Fregean might acknowledge these syntactic similarities between definite descriptions and quantifiers but insist that definite descriptions be classed with proper names on account of their *inferential* behavior. And it is true that there are important similarities in inferential behavior between

---

[6]To get the first reading, imagine the continuation "In 1999, this was Professor Brown; in 2000, it was Professor White; . . . ." To get the second, imagine the continuation "And her name is Sarah White." This example is lightly adapted from one used by Jeffrey King in *Complex Demonstratives: A Quantificational Account* (Cambridge: MIT Press, 2001), 10–11.

definite descriptions and proper names. But there are also important differences: for example, definite descriptions behave differently from proper names in modal and temporal contexts.[7] There are also inferential similarities between definite descriptions and quantifiers. For example, all natural language determiners, including "the," seem to obey the following principle:

> *Conservativeness:* $[Det\, x : Fx]Gx \Leftrightarrow [Det\, x : Fx](Fx \wedge Gx)$[8]

For example,

> A woman is angry ⇔ a woman is a woman and is angry.
>
> No American drivers have reflexes ⇔ no American drivers are American drivers and have reflexes.
>
> Most goldfish die of disease ⇔ most goldfish are goldfish and die of disease.
>
> The American driver is fast ⇔ the American driver is an American driver and is fast.

Proper names do not exhibit any analogous behavior.

None of this is meant to be decisive. My point is just that in light of these considerations, Hale and Wright's decision to class numerical definite descriptions as singular terms rather than quantifiers requires a defense—one that takes account of the way in which the theory of descriptions has been developed and motivated in the recent literature. But no defense is offered. In his long discussion of singular terms in Essays 1 and 2 of *The Reason's Proper Study*, Hale simply *presupposes* that definite descriptions are singular terms. The quantificational alternative is not even mentioned.

The decision to treat numerical definite descriptions as singular terms has consequences, not much noted in RPS, for the underlying logic. Though Hale and Wright do not say much in detail about what kind of referential semantics they accept for numerical definite descriptions,[9] it is clear that it must be one on which

(a) numerical definite descriptions can fail to refer, and

[7] See Saul Kripke, *Naming and Necessity* (Cambridge: Harvard University Press, 1980).

[8] See Gennaro Chierchia and Sally McConnell-Ginet, *Meaning and Grammar* (Cambridge: MIT Press, 1996), 425–30.

[9] For a survey of some of the options, see Stephen Neale, *Facing Facts* (Oxford: Oxford University Press, 2001), ch. 10.

(b) atomic sentences containing non-referring numerical definite descriptions cannot be true.

To see why Hale and Wright are committed to (a), note that if all numerical definite descriptions could be presumed to refer—like the individual constants of standard first-order logic—there would be an easy proof of the existence of numbers that did not use HP (or any other nonlogical principle) as a premise:

1. $\forall x(x = x)$ [law of identity]

2. $N_x(Ax) = N_x(Ax)$ [1, universal instantiation, $N_x(Ax)/x$]

3. $\exists y(y = N_x(Ax))$ [2, existential generalization]

But Hale and Wright insist that the following, more complex proof is required.[10]

1. $\forall F \forall G(N_x(Fx) = N_x(Gx) \equiv Eq_x(Fx, Gx))$ [HP]

2. $N_x(Ax) = N_x(Ax) \equiv Eq_x(Ax, Ax)$ [2, second-order universal instantiation, A/F, A/G]

3. $Eq_x(Ax, Ax)$ [second-order logical truth]

4. $N_x(Ax) = N_x(Ax)$ [2, 3, truth-functional logic]

5. $\exists y(y = N_x(Ax))$ [4, existential generalization]

Thus they must have in mind a system in which singular terms are not presumed to refer, and in which universal instantiation and existential generalization are restricted accordingly. Indeed, they say explicitly that they want to leave room for "$\exists x(x = f)$" to turn out false.[11]

They are committed to (b) by their claim that a sufficient condition for a singular term to refer is its presence in a true, atomic, extensional sentence. They are not always careful to add the qualification "atomic": thus, for example, Hale says that "... it suffices for the existence of directions that there are true statements to be made, featuring terms which, if they have reference at all, refer to directions."[12] But without the qualification "atomic,"

[10]See e.g. RPS, 146 n. 48, 309–10; *Frege's Conception*, 147.
[11]RPS, 144. See also *Frege's Conception*, 147–8.
[12]RPS, 103.

this sufficient condition for existence implies that all singular terms in *false* sentences refer, too, since the negation of any false sentence is a true sentence containing the very same singular terms. This consequence should not be welcome to Hale and Wright: as noted above, they want to leave room for "$\exists x(x = f)$" to be false, and it can't be false if "f" refers. When they are careful, they restrict their referentiality criterion to "pure and applied arithmetical statements of identity and predication"—that is, atomic sentences in the language of arithmetic.[13]

Together, (a) and (b) demand a slight departure from classical logic. In light of (a), the inference rules for universal instantiation and existential generalization must be modified to block instantiation of first-order variables with non-referring singular terms. The obvious way to do this, given (b), is to restrict the instantial terms to those occurring in true atomic sentences:

$$\text{UI*}\quad \begin{array}{l} \forall x \Phi x \\ \Psi a \text{ [must be atomic]} \\ \hline \Phi a \end{array} \qquad\qquad \text{EG*}\quad \begin{array}{l} \Phi a \\ \Psi a \text{ [must be atomic]} \\ \hline \exists x \Phi x \end{array}$$

This gives us a *free logic*.[14] These rules vindicate the second proof given above while blocking the first one at step (2)—just what Hale and Wright need.[15]

Oddly, in RPS Hale and Wright do not seem fully aware that their decision to treat numerical definite descriptions as singular terms commits them

[13] RPS, 153.

[14] That Hale and Wright need a free logic has been pointed out by Stewart Shapiro and Alan Weir, "'Neo-Logicist' Logic is not Epistemically Innocent," *Philosophia Mathematica* 3 (2000), 160–89, and by Michael Potter and Timothy Smiley, "Abstraction by Recarving," *Proceedings of the Aristotelian Society* 101 (2001), 327–38. The version of free logic I am offering Hale and Wright here follows their own practice more closely than the versions discussed in these papers.

It is worth noting a contrast here with Frege's procedure. Hale and Wright need a free logic because they seek to show that terms of the form "$N_x(Ax)$" have referents by *using* such terms in proofs, and so must be able to use such terms without presupposing that they have referents. Frege, by contrast, sets out to show, *before* employing his concept-script in proofs, that every singular term that can be formed in it has a referent. See *The Basic Laws of Arithmetic*, trans. Montgomery Furth (Berkeley: University of California Press, 1964), sections 29–32, and Øystein Linnebo, "Frege's Proof of Referentiality," *Notre Dame Journal of Formal Logic* 45 [2004], 73-98.

[15] Note that in line (5) of the second proof can be obtained by EG* from line (4), because (4), being atomic, can instantiate both Φa and Ψa in the rule.

to using a free logic.[16] On a single page one can find them making claims that only make sense in a classical logic framework, and others that presuppose a free logic. For example, on p. 141 of RPS they suggest that instead of stipulating "$\#f$" directly (where "f" is a singular term and "$\#$" is an arbitrary context), we might stipulate "$\forall x(x = f \supset \#x)$." But in classical logic—where we assume that all terms refer—this is equivalent to $\#f$! So here they must be assuming a free-logical framework. But on the very same page, they characterize "$\#f \supset \exists x \#x$" as "logically true." This formula is logically true in classical logic, but certainly not in the free logic described above, except in the special case where "$\#f$" is atomic.

I began this section by asking whether it was essential to the neologicist program that numerical definite descriptions be classed as singular terms, and if so, why. One way of sharpening this question is to ask what (if anything) would be wrong with carrying out the program in another way, taking numerical definite descriptions as quantifiers and sticking with classical logic. On this approach, the basic arithmetical primitive would be a relation "$Num(\xi, \Phi)$," interpreted as "ξ numbers the Φs." Instead of writing "$N_x(Fx)$" one would write "$[the\ x : Num(x, F)]$", construing the description as a quantifier. HP would become

HP1 $\forall F \forall G([the\ x : Num(x, F)][the\ y : Num(y, G)](x = y) \equiv Eq_x(Fx, Gx))$,

or equivalently,

HP2 $\forall F \forall G(\exists! x Num(x, F) \wedge \exists! x Num(x, G) \wedge \forall x(Num(x, F) \equiv Num(x, G)) \equiv Eq_x(Fx, Gx))$.[17]

Is there any reason to favor HP over HP1 or HP2 as a neologicist foundation for arithmetic?

Hale and Wright certainly can't object that the left hand sides of instances of HP2 are existence claims. For the same is true of HP, in a free-logical framework. In a recent paper critical of the neo-logicist project, Michael Potter and Timothy Smiley point out that there are two possible identity predicates in a free-logical framework, a "strong" reading on which "$a = b$" can be true only if both "a" and "b" refer to existent objects, and a "weak" reading on which "$a = b$" can be true when neither a nor b refers to an existent

[16] Free logic is mentioned only briefly, in the Postscript (433). Hale and Wright now acknowledge (*p.c.*) that their program requires a free logic.

[17] Here "$\exists! x Num(x, F)$" abbreviates "$\exists x(Num(x, F) \wedge \forall y(Num(y, F) \supset y = x))$."

object. They criticize Hale for "[taking] for granted" the strong reading in formulating HP.[18] In his response, Hale acknowledges that he needs the strong reading, but questions why this should be taken to undermine his claim that HP can be stipulated without prior epistemological obligation:

> All that is stipulated is the truth of a (universally quantified) biconditional. In general, this will leave entirely open the question whether terms of the type provided for by the left hand side have reference or not—and it will do so, regardless of whether the identity predicate is understood as signifying a strong or rather a weak identity relation in Potter and Smiley's sense. There is therefore no good ground, for all we have seen so far, to insist that if an abstraction principle is to be the object of legitimate stipulation, it must be existentially bowdlerized by deploying the weak identity relation in the way Potter and Smiley suggest.[19]

But if Hale has no objection to formulating HP in a free-logical framework with strong identity, why should he object to formulating it as HP2? HP2 is a *conditional* existence claim in just the same way as HP is.

It might be objected that HP2 does not look like a "criterion of identity," and thus is not of the right form to impart competence with the sortal concept *number*. But this objection would not touch the logically equivalent HP1, which *does* look an identity criterion for numbers.

So my first question for Hale and Wright is this: What, exactly, would we be missing if we started with HP1 or HP2 instead of HP? Is it *essential* to the neologicist program to follow Frege in taking numerical definite descriptions to be singular terms?

2 HP and PA as implicit definitions

The second question I want to ask concerns our *entitlement* to lay down HP. Here I will accept, for the sake of argument, that we *are* entitled to stipulate HP without prior epistemic work, for the reasons Hale and Wright give. My question is this: why don't we have the *same* entitlement to lay down the

[18] "Abstraction by Recarving," 336. Shapiro and Weir make a similar criticism in their paper, 186–7.

[19] "A Response to Potter and Smiley," *Proceedings of the Aristotelian Society* 101, 339–58, at 347.

Peano axioms directly? This is something to which Hale and Wright have given a little attention in the book—but not enough, I think.

Why, then, do Hale and Wright think that we *are* entitled to "lay down" HP "without significant epistemological obligation"?[20] Not because HP is an abstraction principle, or because it offers criteria of identity—for there are plenty of formally similar principles that we are not entitled to "lay down" (Basic Law V, for one). And not because it purports to be an implicit definition of the functor "$N_x()$"—for it might purport falsely. Rather, we are entitled to lay down HP because HP satisfies certain general constraints that a putative implicit definition must satisfy in order to count as fixing a meaning for its constituent nonlogical terms (constraints that Basic Law V does not satisfy). It is the burden of Essay 5 of RPS—"Implicit Definition and the A Priori"—to articulate these constraints and argue that HP satisfies them.

It is notable that Hale and Wright resort to a general theory of implicit definition here, and not, say, a narrower theory of *acceptable abstraction principles*. In arguing that we can legitimately stipulate HP, they do not invoke features unique to abstraction principles. In particular, they do not invoke their claim that the right and left hand sides of instances of abstraction principles "carve up the same contents" or "reconceptualize the same states of affairs" as a *criterion* for distinguishing acceptable abstraction principles from unacceptable ones. Hale is admirably clear about this in his "Response to Potter and Smiley":

> . . . they [Potter and Smiley] are quite wrong in thinking that, according to the view they are criticizing, it is the possibility of viewing left hand sides as recarving the content of corresponding right hand sides which is the *criterion* for the goodness or otherwise of an abstraction. The criterion, to repeat, is simply whether the crucial constraints (**on implicit definition in general**) are satisfied. If, in the case of an implicitly definitional abstractive stipulation, those constraints are satisfied, then instances of its left hand side will recarve the content of corresponding instances of its right hand side—but this will be a *consequence* of the satisfaction of the criterion, not the criterion itself. (346, boldface emphasis added)[21]

[20]RPS, 321.

[21]In Essay 5 of RPS, Hale and Wright say that "the stipulation of Hume's Principle,

An abstraction principle can be legitimately "laid down," then, just in case it satisfies certain general constraints on implicit definition. What are these constraints? "When does a definition—of any kind—so fix a use that it genuinely explains a meaning?"[22] Hale and Wright list the following four conditions:[23]

1. *Consistency.* The principle must be logically consistent.

2. *Conservativeness.* When added to any theory with which it is consistent, the principle must not imply anything new about the old ontology.[24]

3. *Generality.* The principle must determine truth-conditions for a general *range* of contexts containing the expression it introduces. (How large a range is required is left rather vague.)

4. *Harmony.* If an expression is introduced by means of multiple implicit definitions, they must work together in a way that makes sense: for

and other abstraction principles, is tantamount to a resolution to *reconceive* the subject matter of their introductory components in a fashion determined by the overall syntax of and antecedently understood components in the type of identity statement introduced" (149). If the stipulation itself is tantamount to a resolution to reconceive the subject matter, then this reconceiving or recarving can't be the source of our *entitlement* to make the stipulation. The role of "content recarving" is simply to block an objection to the very idea that a principle like HP might function as an implicit definition. The objection complains: how can HP avoid being "arrogant" if the left-hand sides of its instances demand an ontology that the right-hand sides do not? And the answer is: the right-hand sides already demand the ontology of abstracts, since they describe the *very same states of affairs* as the left-hand sides (see RPS, 149, 277).

[22]RPS, 132.

[23]RPS, 132 ff. In Appendix 1 to Essay 13, Wright adds a fifth constraint. A putative implicit definition is *modest* if any implications for the enlarged universe must be grounded in what the principle implies about the objects for which it introduces means of reference. Wright thinks he needs this additional constraint to rule out some "bad company" proposed by Shapiro and Weir (324). I am going to ignore this fifth constraint here; it won't help Hale and Wright answer my question, because the Peano axioms certainly satisfy it as well as HP.

[24]For a more precise formulation, see RPS, 297 n. 49. Note that Conservativeness as Hale and Wright understand it is different from the usual proof-theoretic notion in at least two ways: it is not limited to derivability, and it allows that the new principle might imply some new truths stateable in the old vocabulary, provided they "[make] no demands on the previously recognized ontology, whatever it may have been, but [are] sustained by the objects . . . to which the Principle introduces means of reference" (133 n. 32).

example, elimination rules should not be weaker than is justified by the introduction rules.

Hale and Wright are not proposing that we must *prove* that a putative implicit definition satisfies these four constraints before we can be justified in laying it down. If that were the view, then we wouldn't be justified in laying down HP unless we could first prove that it was consistent. Of course, we can prove that HP is consistent *relative to analysis*, and that is enough to dispel any realistic worry that HP is inconsistent.[25] But Hale and Wright have a foundational aim: they want to show that our most basic mathematical knowledge, our knowledge of arithmetic, can be grounded in HP. It would frustrate this foundational aim to concede that our entitlement to HP rests on our entitlement to the claim that *analysis* is consistent. In proposing that we can lay down HP "without significant epistemological obligation," I take it, Hale and Wright are claiming that we can lay it down without *needing* to do the kind of epistemic work that would be involved in proving HP to be consistent, conservative, etc. On their view, our entitlement to HP is not something we have to *earn*; it's something we start with. In the absence of any positive reason to think that HP *fails* to meet the four constraints on implicit definitions, then, we are entitled to "lay it down" and take it to be true.[26]

We can now reformulate our question as follows: which of these four constraints do the Peano axioms fail to satisfy, and why? The question is an urgent one for the neologicists. For, if the Peano axioms turn out to satisfy all of the constraints, then either (a) satisfying these constraints is not sufficient to make a stipulated principle a successful implicit definition, in which case the case for HP's being such a definition has not yet been made, or (b) the Peano axioms themselves qualify as a successful implicit definition of the arithmetical primitives contained in them, in which case there is no evident *epistemological* advantage to founding arithmetic on HP, rather than the Peano axioms. Neither horn of this dilemma is a comfortable one for the neologicist.

[25]See George Boolos, "The Consistency of Frege's *Foundations of Arithmetic*," in *On Being and Saying: Essays in Honor of Richard Cartwright*, ed. Judith Jarvis Thomson (Cambridge: MIT Press, 1987), 3–20.

[26]Wright puts the point this way: "Explanations which seem to work well enough should surely be regarded as innocent until proved guilty. So it will be enough of a disanalogy if there is no extant reason to doubt the consistency of a second-order abstraction if the usual lines to contradiction do not succeed in its case" (RPS, 282).

There is surprisingly little in RPS that bears directly on this question. The idea that the Peano axioms themselves might count as an implicit definition just does not appear on Wright and Hale's menu of alternatives for explaining our knowledge of them:

> . . . if the question is raised, how do we know that the natural numbers constitute an infinite series of which the Dedekind-Peano axioms hold good, the available answers would seem to be, crudely, of just three broad kinds: that we don't actually know any such thing—it's a fiction or a groundless stipulation; or that we just do, primitively and immediately, know it; or that we know it in a manner informed by deeper principles of some sort. Our proposal is an answer of the third kind: the infinity of the number series may be known by knowing that it follows from the constitutive principle for the identity of cardinal numbers.[27]

The options here—fiction, primitive grasp, or inferential knowledge—do not seem to leave room for the kind of knowledge we might have (on Hale and Wright's own account!) of a principle we lay down in order to fix the meanings of its constituent terms. That there is a missing alternative here comes out very clearly in Hale's "Response to Potter and Smiley." Our knowledge of the Peano axioms *must* be inferential, he argues:

> For the **only remaining alternative**—holding that the infinity of the numbers or the truth of the usual axioms is apprehended directly and immediately—is epistemologically completely unilluminating. There is nothing intrinsically wrong with the idea that some knowledge is—perhaps even must be—direct and non-inferential. But **the obvious candidates—sense-perception and introspection**—provide no satisfactory model either for arithmetical knowledge in particular or for a priori knowledge of necessary truth in general. (boldface emphasis added)[28]

[27]RPS, 147-8.

[28]"Response to Potter and Smiley," 349. There is a very similar passage in RPS: ". . . anyone sympathetic to the . . . thought . . . that the infinity of the natural numbers—and indeed the truth of the Dedekind-Peano axioms—is part of our most basic knowledge, should be receptive to the idea that it is *inferential* knowledge, grounded ultimately in deeper principles of some kind determining the nature of cardinal number. For the only alternative which takes it seriously—the idea that the truth of the usual axioms is somehow

If our only two choices for explaining how we know the Peano axioms to be true are *inferentially* or *through a direct and immediate grasp*, then perhaps we must answer *inferentially*. But Hale and Wright themselves have put a third option on the table: we know some principles to be true because we stipulate them as implicit definitions and they satisfy the constraints on acceptable implicit definitions. We do not need to *earn* our entitlement to these principles (either by inference, or through some kind of primitive and immediate grasp); they enjoy a default entitlement that persists as long as we have no reason to think they do *not* satisfy the constraints. If our entitlement to HP has this character, why can't our entitlement to the Peano axioms be like this, too?

Presumably Hale and Wright think that it is *obvious* that the Peano Axioms cannot count as an implicit definition of the arithmetical primitives. The question is why not. Let PA be the conjunction of some standard (second-order free-logical) Peano axioms, with three non-logical expressions: a one-place predicate "N", an individual constant "0", and a first-level relation "P" for predecession.[29]

1. $N0$
2. $\forall x \forall y (Nx \wedge Pxy \supset Ny)$
3. $\forall x \forall y \forall z (Pxy \wedge Pxz \supset y = z)$
4. $\forall x \forall y \forall z (Pxz \wedge Pyz \supset x = y)$
5. $\neg \exists x Px0$
6. $\forall x (Nx \supset \exists y Pxy)$
7. $\forall F [F0 \wedge \forall x \forall y (Fx \wedge Pxy \supset Fy) \supset \forall x (Nx \supset Fx)]$

How well does PA meet Hale and Wright's constraints on implicit definitions?

---

apprehended primitively and *immediately*—is not only epistemologically utterly unilluminating but flies in the face of the historical fact that the grasp and practice of the theory of the finite cardinals did not originate with the Dedekind-Peano axiomatization but antedated and informed it" (147).

[29] These are taken from Richard Heck, "Finitude and Hume's Principle," *Journal of Philosophical Logic* 26 (1997), 589–617, at 592 (system PAS). '$A \wedge B \supset C$' abbreviates '$(A \wedge B) \supset C$'.

- *Consistency:* We have at least as much reason to think that PA is consistent as we have to think that HP is consistent, because any proof of a contradiction from PA could be transformed into a proof of a contradiction from HP. (Just use Frege's proofs to derive definitional equivalents of the Peano axioms from HP, then proceed from there.)

- *Conservativeness:* Does stipulating PA "...introduce fresh commitments (i) which are expressible in the language as it was prior to the introduction of its *definiendum* and (ii) which concern the previously recognized ontology of concepts, objects, and functions, etc., whatever in detail they may be"?[30] No. PA has no new implications for the non-arithmetical part of the universe. (And if it did, so would HP.)

- *Harmony:* It is a little hard to know how to apply this constraint outside the domain of logical introduction and elimination rules. But the Peano axioms work very well together indeed, and it would be surprising if at this point we found grounds for thinking them "disharmonious."

- *Generality:* Does the stipulation of PA succeed in fixing truth-conditions for a sufficiently wide range of contexts involving the arithmetical primitives "N," "0", and "P"? That depends on how wide a range of contexts counts as sufficient. PA does not fix truth-conditions for sentences like "0 = Julius Caesar", "N(Julius Caesar)", or "P(Nero, Claudius)." But Hale and Wright concede that perhaps all expressions have a

  > ...limited *range of significance*—a limited range of sentential matrices in which it so much as makes sense to introduce them—so that the proper demand imposed by the Generality Constraint on the definition of an expression is only that it bestow understanding of any sentence resulting from combining the definiendum with an understood matrix encompassed in its range of significance.[31]

  If any contexts are outside the range of significance for "0", "N", and "P", the ones quoted above would seem to be prime candidates. Further discussion would require consideration of the infamous "Caesar

[30]RPS, 133.
[31]RPS, 135.

problem." I will forego this here, because Wright and Hale never suggest that the problem with direct stipulation of PA has anything to do with the Generality constraint.

It is not clear, then, how direct stipulation of PA as an implicit definition of the arithmetical primitives would violate any of Wright and Hale's four constraints on acceptable implicit definitions. Nor do Hale and Wright say that they would. Instead, they criticize the direct stipulation approach on the grounds that it would be *arrogant*: "...the stipulation of the axioms would directly call for the existence of an appropriately large range of objects...and would therefore be arrogant."[32] *Arrogance* is defined (earlier in the essay) as follows:

> Let us call *arrogant* any stipulation of a sentence, '$\#f$', whose truth, such is the antecedent meaning of '$\#\_$' and the syntactic type of 'f', cannot justifiably be affirmed without collateral (a posteriori) epistemic work.[33]

It is hard to see how this helps. Surely Hale and Wright do not want to argue that PA cannot be justifiably affirmed without collateral *a posteriori* epistemic work, since on their own account, PA can be justified *a priori.* But if we excise the parenthetical "a posteriori" from the definition of arrogance, then to say that stipulating PA is arrogant is just to say that we are not entitled to lay down PA without significant epistemic work. That is the *conclusion* Hale and Wright need, not an argument for it.

The reason Hale and Wright offer for thinking that direct stipulation of PA would be arrogant is that it "would directly call for the existence of an appropriately large range of objects...." But what does "directly" mean here? PA does not include an axiom that *says* "there are infinitely many natural numbers." This is of course a *consequence* of the axioms. But it is also a consequence of HP, which, according to Hale and Wright, does not directly call for the existence of anything. What Hale and Wright seem to mean when they say that HP does not directly call for the existence of anything is that its existence claims are *conditional.* HP says that *if* the objects falling under a concept F can be put into one-one correspondence with themselves, then the number of Fs exists. It does not say outright that the number of Fs exists:

[32]RPS, 147.
[33]RPS, 128.

> The truth-*value* [as opposed to truth-*conditions*] of instances of the abstraction's left-hand side is never itself a matter of direct stipulation. ... The existence of referents for [abstract]-terms is therefore never part of what is stipulated—and implicit definition through Fregean abstraction is accordingly never arrogant *per se*.[34]

The view, then, is that "in order to avoid arrogance, legitimate implicit definitions must have an essentially conditional character."[35] This amounts to a fifth constraint, independent of the other four. Even if PA is consistent, conservative, general, and harmonious, its stipulation would amount to an *unconditional* existence claim. It doesn't just set conditions for the existence of numbers; it says outright that these numbers exist. That is why it cannot be legitimately stipulated.

It would be wrong to object that of the seven axioms of PA, only the first fails to have the form of a quantified conditional. For the point is that the existence claims should be made conditional on formulas that do not contain any of the vocabulary being "implicitly defined." So Axiom (6)—"if x is a natural number, then there is something it precedes"—does not count as "conditional" in the relevant sense.

But what is the motivation for this new Conditionality constraint? In Essay 5 of RPS, Hale and Wright point out that it would be "presumptuous" to lay down

J Jack the Ripper is the perpetrator of this series of killings.

as an implicit definition of "Jack the Ripper," because "we could have no a priori entitlement to the presupposition that 'the perpetrator of this series of killings' refers at all."[36] What we *can* lay down, they say, is the conditional

CJ If anyone singly perpetrated these killings, it was Jack the Ripper.

But this example does not motivate adding Conditionality as a *separate* constraint, in addition to the other four. The Conservativeness constraint already suffices to explain why J is illegitimate and CJ is okay. J implies something new about the prior ontology—that there were not multiple

[34] RPS, 146. Compare RPS, 129, 144–5.
[35] RPS, 146, cf. 129.
[36] RPS, 127.

murderers—and so fails to be conservative.[37] So we have still not seen any reason to demand Conditionality *in addition to* the other constraints.

Later in Essay 5, Hale and Wright say that in empirical scientific cases, conditional stipulations are required "in order to keep open the possibility of empirical disconfirmation."[38] But in the mathematical case, there is presumably no possibility of *empirical* disconfirmation. Indeed, no mathematical principle that satisfies the Conservativeness constraint can have any empirical consequences, so no such principle can be empirically disconfirmed.[39] If the reason for insisting on Conditionality is to leave room for empirical disconfirmation, why should we insist on it in *mathematical* cases?

It is hard to see why Conditionality matters, then, except where it helps secure satisfaction of the other constraints. But if it does matter, it is easy enough to secure in the case we are discussing. Instead of stipulating PA, we can stipulate

CPA $\forall x(x = x) \equiv PA$.[40]

It should be clear that if PA satisfies the other four constraints, so does CPA. And CPA makes a conditional existence claim. Granted, it may seem conditional in a Pickwickian sense, since it makes the existence of numbers conditional on a logical truth. But this is hardly an objection that a neo-logicist can make! HP, too, makes the existence of numbers conditional on logical truths: that is precisely why it can serve as the basis of a kind of logicism.

Hale and Wright might object that CP can be ruled out as a legitimate implicit definition on the grounds that its right hand side has ontological commitments of which its left hand sides is innocent. But of course they face a similar challenge to HP. Their response to that challenge is that the two sides of instances of HP are merely different ways of reconceptualizing

[37] I suppose Hale and Wright might claim that (J) merely *presupposes*, and does not *imply*, the existence of a unique killer. If they go this way, they will need to be much more specific about the semantics of definite descriptions than they have been so far. Even in this case, the obvious move seems to be an extension of Conservativeness to include presuppositions as well as implications, not a new Conditionality constraint.

[38] RPS, 144.

[39] Hale and Wright point this out themselves in RPS, 145.

[40] Hale and Wright are careful *not* to say that stipulation of the truth of a non-conditional sentence is *always* arrogant: after all, they note, the sentence may be "equivalent to a conditional" (RPS, 130 n. 25).

the same state of affairs, and thus have the same ontological commitments.[41] Recall, however, that the doctrine of content-recarving is not meant to help distinguish legitimate implicit definitions from illegitimate ones.[42]

Pending clarification on this point, then, I see no reason to suppose that we should be any less entitled, on Wright's and Hale's principles, to "lay down" PA than we are to "lay down" HP. And if that's right, then there is little distinctively *neo-Fregean* left to the neo-Fregean program. The epistemological theory required to make sense of the idea that we can lay down HP without epistemic work seems to allow us to lay down PA directly. But then Frege's Theorem plays no *essential* role in the neologicist story about our knowledge of arithmetic. The real work is being done by the theory of implicit definition—a theory Frege himself would have abhored[43]—and the logicist trappings are irrelevant.[44] Thus, my double vision. I can see a defensible version of neologicism or a distinctively neo-Fregean one, but I can't seem to get both in focus at the same time.

---

[41] See RPS, 148–9, 276–7. Note that Hale's account of "content-carving" in the Postscript to Essay 4 seems to rule out saying that PA reconceptualizes the same content as "$\forall x(x = x)$". The only way two necessary sentences can count as recarving the same content, on this account, is if they can be obtained by uniform bilateral substitution from two contingent sentences that recarve the same content (RPS, 113–14).

[42] See page 10 and note 21, above.

[43] See especially his correspondence with Hilbert.

[44] They are irrelevant, that is, to the *epistemology* of arithmetic, which is the payoff Hale and Wright have emphasized the most in their work. It might still be argued that there are other, non-epistemological reasons for preferring HP to PA as a basis for arithmetic. For example, logicists used to argue that starting with PA leaves us without any understanding of *applied* arithmetic. But as Quine points out, we can define the fundamental notion of applied arithmetic, *number of*, in terms of logic and pure arithmetic: "That there are n so-and-sos can be explained simply as meaning that the so-and-sos are in one-to-one correspondence with the numbers up to n" (*Ontological Relativity and Other Essays* [New York: Columbia University Press, 1969], 44). So there is a symmetry between the two approaches. The Fregean approach starts with the fundamental notion of applied arithmetic (number of) and the axiom that governs it (HP), then defines the fundamental notions of pure arithmetic (0, predecession, natural number) and derives the principles that govern them (PA). The Peano approach, by contrast, starts with the fundamental notions of pure arithmetic (0, predecession, natural number) and the axioms governing them, then defines the fundamental notion of applied arithmetic (number of) and derives the principle that governs it (a form of HP restricted to finitely instantiated concepts). Still, it might be argued that one set of primitives is more fundamental to our understanding than the other: see e.g. Richard Heck, "Cardinality, Counting, and Equinumerosity," *Notre Dame Journal of Formal Logic* 41 (2000). Such an argument would suggest a very different understanding of the significance of neo-Fregean logicism than the one articulated in RPS.