THE METAPHYSICS AND EPISTEMOLOGY OF ABSTRACTION*

CRISPIN WRIGHT
University of St. Andrews & New York University

1

Paul Benacerraf famously wondered[1] how any satisfactory account of mathematical knowledge could combine a face-value semantic construal of classical mathematical theories, such as arithmetic, analysis and set-theory – one which takes seriously the apparent singular terms and quantifiers in the standard formulations – with a sensibly *naturalistic* conception of our knowledge-acquisitive capacities as essentially operative within and subject to the domain of causal law. The problem, very simply, is that the entities apparently called for by the face-value construal – finite cardinals, reals and sets – do not, seemingly, participate in the causal swim. A refinement of the problem, due to Field, challenges us to explain what reason there is to suppose that our basic beliefs about such entities, encoded in standard axioms, could possibly be formed reliably by means solely of what are presumably naturalistic belief-forming mechanisms. These problems have cast a long shadow over recent thought about the epistemology of mathematics.

Although ultimately Fregean in inspiration, Abstractionism – often termed 'neo-Fregeanism' – was developed with the goal of responding to them firmly in view. The response is organised under the aegis of a kind of linguistic – better, propositional – 'turn' which I suggest it is helpful to see as part of the content of Frege's Context Principle. The turn is this. It is not that, *before* we can understand how knowledge is possible of statements referring to or quantifying over the abstract objects of mathematics, we need to understand how such objects can be given to us as objects of acquaintance or how some other belief-forming mechanisms might be sensitive to them and their characteristics. Rather we need to tackle directly the ques-

---

* This is a shortened version of Bob Hale's and my "The *Metaontology* of Abstraction" – Hale and Wright (2009). My thanks to Oxford University Press for permission to publish the present paper here.

[1] in Benacerraf (1973).

tion how propositional thought about such objects is possible and how it can be knowledgeable. And this must be answered by reference to an account of how meaning is conferred upon the ordinary statements that concern such objects, an account which at the same time must be fashioned to cast light on how the satisfaction of the truth-conditions it associates with them is something that is accessible, in standard cases, to human cognitive powers.[2]

Abstraction principles are the key device in the epistemological project so conceived. Standardly, an abstraction principle is formulated as a universally quantified biconditional – schematically:

$$(\forall a)(\forall b)(\mathrm{S}(a) = \mathrm{S}(b) \leftrightarrow E(a, b)),$$

where a and b are variables of a given type (typically first- or second-order), 'S' is a term-forming operator, denoting a function from items of the given type to objects in the range of the first-order variables, and E is an equivalence relation over items of the given type.[3] What is crucial from the abstractionist point of view is an epistemological perspective which sees these principles as, in effect, stipulative implicit definitions of the S-operator and thereby of the new kind of term formed by means of it and of a corresponding sortal concept. For this purpose it is assumed that the equivalence relation, E, is already understood and that the kind of entities that constitute its range are familiar – that each relevant instance of the right hand side of the abstraction, $E(a, b)$, has truth-conditions which are grasped and which in a suitably wide range of cases can be known to be satisfied or not in ways that, for the purposes of the Benacerrafian concern, count as

[2] For efforts to develop and defend this approach, see Wright (1983), ch.1; Hale (1987), chs.1,7; Hale & Wright (2001), Introduction sect.3.1, Essays 5,6; Hale & Wright (2002)

[3] More complex forms of abstraction are possible – see, for example, Hale (2000), p.107, where positive real numbers are identified with ratios of quantities, these being defined by abstraction over a four-term relation. One could replace this by a regular equivalence relation on ordered pairs of quantities, but this is not necessary – it is straightforward to extend the usual notion of an equivalence relation to such cases. It is also possible – and possibly philosophically advantageous, insofar as it encourages linking the epistemological issues surrounding abstraction principles with those concerning basic logical rules – to formulate abstractions as pairs of schematic introduction- and elimination-rules for the relevant operator, corresponding respectively to the transitions right-to left and left-to right across instances of the more normal quantified biconditional formulation.

unproblematic. In sum: the abstraction principle explains the truth conditions of S-identities as coincident with those of a kind of statement we already understand and know how to know. So, the master thought is, we can now exploit this prior ability in such a way as to get to know of identities and distinctions among the referents of the S-terms – entities whose existence is assured by the truth of suitable such identity statements. And these knowledge possibilities are assured without any barrier being posed by the nature – in particular, the abstractness – of the objects in question (though of course what pressure there may be to conceive of the referents of terms introduced by abstraction *as* abstract, and whether just on that account or for other reasons, is something to be explored independently[4].)

2

There are very many issues raised by this proposal. One might wonder, to begin with, whether, even if no other objection to it is made, it could possibly be of much interest merely to recover the means to understand and know the truth value of suitable examples of the schematised type of identity statement, bearing in mind the ideological richness displayed by the targeted mathematical theories of cardinals, real numbers and sets. The answer is that abstraction principles, austere as they may seem, do – in a deployment that exploits the collateral resources of second-order logic and suitable additional definitions – provide the resources to recover these riches – or at least, to recover theories which stand interpretation as containing them.[5] There then are the various misgivings – for example, about "Bad Company" (differentiating acceptable abstraction principles from various kinds of unacceptable ones), about Julius Caesar (in effect, whether abstraction principles provide for a sufficient range of uses of the defined terms to count as properly explaining their semantic contribution, or justifying the attribution of reference to them), about impredicativity in the key (second-order) abstractions that underwrite the development of arithmetic

[4] See Hale & Wright (2001), Essay 14, sect.4, for discussion of an argument aimed at showing that abstracts introduced by first-order abstraction principles such as Frege's Direction Equivalence cannot be identified with contingently existing concrete objects.

[5] At least, they do so for arithmetic and analysis. So much is the burden of Frege's Theorem, so called, and work of Hale and, separately, Shapiro. For arithmetic, see Wright (1983), ch.4; Boolos (1990) and (1998), pp.138–141; Hale & Wright (2001), pp.4–6; and for analysis, Hale (2000) and Shapiro (2000). The prospects for an abstractionist recovery of a decently strong set theory remain unclear.

and analysis, and about the status of the underlying (second-order) logic – with which the secondary literature over the last twenty-five years has mostly been occupied. For the purposes of the present discussion, I assume all these matters to have been resolved.[6] Even so, another major issue may seem to remain. There has been a comparative dearth of head-on discussion of the abstractionist's central *ontological* idea: that it is permissible to fix the truth-conditions of one kind of statement as coinciding with those of another – 'kind' here referring to something like logical form – in such a way that the overt *existential* implications of the former exceed those of the latter, although the *epistemological* status of the latter, as conceived in advance, is inherited by the former. Recently however there have been signs of increasing interest in this proposal among analytical metaphysicians. A number of writers have taken up the issue of how to "make sense" of the abstractionist view of the ontology of abstraction principles, with a variety of proposals being canvassed as providing the 'metaontology' abstractionists need, or to which they are committed.[7] I shall here describe what I regard as the correct view of the matter.

3

In what follows, I shall try to address two questions, one metaphysical and one epistemological, – questions that some may feel need answering before abstractionism should even be considered to be a competitive option. The metaphysical question is:

---

[6] Since the 'noise' from the entrenched debates about Bad Company, Impredicativity, etc., is considerable, it may help in what follows for the reader to think in terms of a context in which a first order abstraction is being proposed – say Frege's well known example of the Direction Principle:

Direction(a) = Direction(b) iff. a and b are parallel

in which range of 'a' and 'b' is restricted to concrete straight lines – actual inscriptions, for example – and of the listed concerns, only the Caesar problem remains. The pure ontological problems about abstraction – if indeed they are problems – arise here in a perfectly clean form.

Previous discussions of the more purely ontological issues are to be found in Wright (1983), chs.1–3; Hale (1987); Hale & Wright (2001), Essays 1–9 and 14.

[7] In particular, Eklund (2006), Sider (2007), Hawley (2007), and Cameron (2007) all discuss the abstractionist's (alleged) need for a suitable 'metaontology'.

(M) What does the world have to be like in order for (the best examples of) abstraction to work?

And the associated epistemological question is:

(E) How do we know – what reason can be given for thinking – that the transition, right to left, across the biconditional in instances of (the best examples of) abstraction is truth preserving?[8]

Very different conceptions are possible of what it is to give a satisfactory answer to question (E); that is, to justify the thought that a good abstraction is truth-preserving, right-to-left. One such conception, which I reject, has it, in effect, that it is, in some sense, *possible*[9] – something we have initially no dialectical right to discount – for any abstraction to fail right-to-left unless some relevant kind of independent, collateral assurance is forthcoming from the metaphysical nature of the world. There are, that is to say, possible situations – in some relevant sense of 'possible' – in which an abstraction which actually succeeds would fail, even though *conceptually*, at the level of explanation and the understanding thereby imparted, everything is as it is in the successful scenario. Hence in order to make good that the right-to-left transition of an otherwise good abstraction is truth-preserving, argument is needed that some relevant form of metaphysical assurance is indeed provided. This is, plausibly, the way that the metaphysicians who call for a suitably supportive 'metaontology' are thinking about the issue. The 'possible' scenario would be one in which not everything that could exist does exist – in particular, the denoted abstracts do not exist. And the requisite collateral assurance, to be sought within the province of metaphysics, would be that this 'possibility' is not a genuine possibility – because, for instance, maximalism[10] is true (and is so, presumably, as a matter of metaphysical necessity.) Although the idea is by no means as clear as one would like, I reject this felt need for some kind of collateral metaphysical assistance. The kind of justification which I acknowledge *is* called for is precisely justification for the thought that no such collateral assistance is necessary. There is no hostage to redeem. A (good) abstraction *itself* has the resources to close off the alleged (epistemic metaphysi-

[8] I will hence generally omit the parenthetical qualification "the best examples of". But except where stated otherwise, it is to be understood.
[9] – perhaps this modality is*: epistemically* [*metaphysically possible*]!
[10] See Eklund, op, cit. n.7

cal) possibility. The justification needed is to enable – clear the obstacles away from – the recognition that the truth of the right-hand side of an instance of a good abstraction is *conceptually* sufficient for the truth of the left. There is no gap for metaphysics to plug, and in that sense no 'metaontology' to supply. This view of the matter is of course implicit in the very metaphor of *content* recarving. It is of the essence of abstractionism, as I understand it.

4

Question (M) is: What does the world have to be like in order for (the best examples of) abstraction to work? A short answer is that it is at least necessary that the world be such as to verify the associated *Ramsey sentences*: the results of existential generalisation into the places occupied by tokens of the new operators. So for any particular abstraction,

$$(\forall a)(\forall b)(\mathrm{S}(a) = \mathrm{S}(b) \leftrightarrow E(a, b))$$

the requirement is that this be true:

$$(\exists f)(\forall a)(\forall b)(f(a) = f(b) \leftrightarrow E(a, b))$$

More generally, the minimum requirement is that each equivalence relation suitable to contribute to an otherwise good abstraction be associated with at least one function on the members of its field that takes any two of them to the same object as value just in case they stand in the relation in question.

A world in which abstraction works, then – a world in which the truth values of the left-and right-hand sides of the instances of abstraction principles are always the same – will be a world that displays a certain ontological richness with respect to functions. Notice that there is no further requirement of the existence of values for these functions. For if 'S' is undefined for any element, c, in the field of E, then the instance of the abstraction in question, $\mathrm{S}(c) = \mathrm{S}(c) \leftrightarrow E(c, c)$, will fail right-to-left. This brings us sharply to the second question, (E). To know that the transition right to left across an otherwise good abstraction principle is truth-preserving, we need to know that the equivalence relation in question is indeed associated with a suitable function. Here is George Boolos worrying about

the latter question in connection with Hume's principle ("octothorpe" is a name of the symbol, '#', which Boolos uses to denote the cardinality operator, "the number of …"):

> … what guarantee have we that there is such a function from concepts to objects as [Hume's Principle] and its existential quantification [Ramsey sentence] take there to be?
>
> I want to suggest that [Hume's Principle] is to be likened to "the present king of France is a royal" in that we have no analytic guarantee that for every value of "F", there is an object that the open definite description[11] "the number belonging to F" denotes …
>
> Our present difficulty is this: just how do we know, what kind of guarantee do we have, why should we believe, that there is a function that maps concepts to objects in the way that the denotation of octothorpe does if [Hume's Principle] is true? If there is such a function then it is quite reasonable to think that whichever function octothorpe denotes, it maps non-equinumerous concepts to different objects and equinumerous ones to the same object, and this moreover because of the meaning of octothorpe, the number-of-sign, or the phrase "the number of." But do we have any analytic guarantee that there is a function which works in the appropriate manner?
>
> Which function octothorpe denotes and what the resolution is of the mystery how octothorpe gets to denote some one particular definite function that works as described are questions we would never dream of trying to answer.[12]

Boolos undoubtedly demands too much when he asks for "analytic guarantees" in this area. But the spirit of his question demands an answer that at least discloses some reason to believe in the existence of a function of the relevant kind. So: what, in general, is it to have reason to believe in the existence of a function of a certain sort?

If, as theorists often do, we think of functions as *sets* – sets of pairs of argument-tuples, and values – then standard existence postulates in set theory can be expected to provide an answer to Boolos's question in a wide range of cases: there is whatever reason to believe in the existence of the functions required by abstraction principles as there is to believe in the existence of the relevant sets. But that is, twice over, not the right kind of way to address the issue for the purposes of abstractionism. For one thing,

---

[11] The reader should note Boolos' ready assimilation of "the number belonging to F" to a definite description – of course, it *looks* like one. But the question whether it is one depends on whether it has the right *kind* of semantic complexity. The matter is important, and we will return to it below.

[12] Boolos (1997), p.306

abstractionism's epistemological objectives require that the credibility of abstraction principles be *self-standing*. They are not to (need to) be shored up by appeal to independent ontological commitments – and if the abstractionist harbours any ambition for a recovery of set-theory, especially not by appeal to a prior ontology of sets. However there is a deeper point. Abstraction principles purport to introduce *fundamental* means of reference to a range of objects, to which there is accordingly no presumption that we have any prior or independent means of reference. Our conception of the epistemological issues such principles raise, and our approach to those issues, need to be shaped by the assumption that we may have – indeed that there may be possible – no prior, independent way of conceiving of the objects in question other than as the values of the relevant function. So when Boolos asks, what reason do we have to think that there is any function of the kind an abstraction principle calls for, it is to skew the issues to think of the question as requiring to be addressed by the adduction of some kind of evidence for the existence of a function with the right properties that takes elements from the field of the abstractive relation as arguments and objects of some independently available and conceptualisable kind as values. If the best we can do, in order to assure ourselves of the existence of a relevant function or, relatedly, of the existence of a suitable range of objects to constitute its values, is to appeal to our independent ontological preconceptions – our ideas about the kinds of things we take to exist in any case – then our answer provides a kind of assurance which is both insufficient and unnecessary to address the germane concerns: insufficient, since independent ontological assurance precisely sheds no light on the real issue – viz. how we can have reason to believe in the existence of the function purportedly defined by an abstraction principle, and accordingly of the objects that constitute its range of values, when proper room is left for the abstraction to be fundamental and innovative; unnecessary since, if an abstraction can succeed when taken as fundamental and innovative, it doesn't need corroboration by an independent ontology.

5

Let us therefore refashion question (E) as follows:

> (E′) How do we know – what reason have we to think – that the transition, right to left, across the biconditional instances of abstraction principles is truth preserving, once it is allowed that the means of reference it introduces to the (putative) values of the (putatively) defined function may be *fundamental*, and that no antecedently available such means may exist?

An answer to (E′) in any particular case must disclose a kind of reason to believe in the existence of a suitable function which originates simply in resources provided by the abstraction principle itself, and independent of collateral ontological preconceptions. Those resources must pertain to what an abstraction can accomplish as an implicit definition of its definiendum – the new term forming operator. Allow, at least pro tem, that an abstraction principle, laid down as an implicit definition of its abstraction operator, may at least succeed in conferring on it a *sense*. So much is tacitly granted by Boolos when he writes in the passage quoted above:

> If there is such a function then it is quite reasonable to think that whichever function octothorpe denotes, it maps non-equinumerous concepts to different objects and equinumerous ones to the same object, and this moreover because of the meaning of octothorpe … But do we have any analytic guarantee that there is a function which works in the appropriate manner?

For it is, after all, by its stipulated role in the relevant version of Hume's principle that the meaning of octothorpe is fixed. So the question is: what, for functional expressions – one standard practice calls them *functors* – needs to be in place in order for possession of sense to justify ascription of reference?

For Frege, functors are to be conceived as an instance of the more general category of *incomplete* expressions: expressions whose 'saturation' by a singular term results in a further complex, object-denoting term. So let's ask in the first instance: is there something general to be said about what justifies the ascription of reference to an incomplete expression? And what, in particular, is the role played by sense? I am not, in posing this question, taking it as uncontroversial that incomplete expressions as a class should be credited with a potential for reference as well as sense. The ques-

tion is rather: for a theorist not already inclined – because of nominalist scruple or whatever reason – to deny reference to incomplete expressions across the board, what should justify the ascription of reference in any particular case?

Let's try the case of simple predicates. Take it that in order to assign a sense to a predicate, it suffices to associate it with a sufficiently determinate satisfaction-condition: to fix under what circumstances it may truly, or falsely, be applied to an item in some appropriate assigned range. And take it that the question whether it has a reference amounts to whether we have thereby succeeded in associating it with a genuine property. Then there is a contrast between two broad ways of taking the question. On one way of taking it, the relevant notion of genuine property is akin to that in play when we conceive it as a non-trivial question whether any pair of things which both exemplify a certain set of surface qualities – think, for example, of a list of the reference-fixers for 'gold' given in a way independent of any understanding of that term or an equivalent – have *a property in common*. When the question is so conceived, the answer may be unobvious and negative: there may be 'fool's' instances of a putative natural kind, or there may even just be no common kind underlying even normal cases of presentation of the qualities in question. Theorists who think of *all* properties in this way – sometimes termed "sparse" theorists – will recognise a gap between a predicate's being in good standing – its association with well-understood, feasible satisfaction conditions – and its hitting off a *real worldly property*. However this conception stands in contrast with that of the more "abundant" theorist, for whom the good standing, in that sense, of a predicate is *already* trivially sufficient to ensure the existence of an associated property, a (perhaps complex) *way of being* which the predicate serves to express.[13] For a theorist of the latter spirit, predicate sense will suffice, more or less,[14] for predicate reference. The sparse theorist, by contrast, will view the relationship as very much akin to that which obtains in the case of complex singular terms: the sense of – the satisfaction condition of – a predicate will aim at an underlying property fit to underwrite in some appropriate manner the capacity of an object to meet that satisfaction

[13] The terminology of abundant and sparse properties originates in Lewis (1986). The general distinction is in Armstrong (1979). See also Bealer (1982) and Swoyer (1996) For a useful overview see Mellor and Oliver (1997).

[14] "More or less" because the abundant theorist may still want to deny reference to certain significant predicates – for instance, those associated with inconsistent satisfaction conditions, or which embed empty terms ("That car is *my dog's favourite colour*").

condition, and the predicate will have reference only insofar as there is indeed such a property provided by the world. Whether that is so will then depend in turn on one's metaphysics of worldly properties.[15]

It is clear enough that the two conceptions of property need not be in competition: it is perfectly coherent to work with both simultaneously. What do compete, however, are the two associated views of predicate reference since no-one inclined to admit both conceptions of property is going to wish to maintain, presumably, that in the case when a predicate is associated with properties of both kinds, it somehow divides its reference over them both, or something of the sort. The natural compatibilising view will be, rather, that it is for the abundant properties to play the role of *bedeutungen* in semantic theory, and the sparse ones to address certain metaphysical concerns.[16]

For predicates at least, then, there is *a* good conception of reference such that to confer a sense is, more or less, to confer a reference. Do these ideas suggest a way of responding to Boolos's question, and thence to question (E′), for the target case: the functors introduced on the left hand side of instances of abstraction principles? Well, there are evident disanalogies. Any predicate associated with a (sufficiently) determinate satisfaction condition is, ceteris paribus, assured of reference to an abundant property. But it seems there should be room for a would-be functor to have sufficient sense to be associated with a determinate condition on any function that is to qualify as presented by it and yet fail to present one. Setting aside any issue about the existence of a range of suitable *arguments* for the purported function in question – as we may in the case of abstraction principles – there are two ways this can happen. One is if a relation can meet the condition in question and yet not be functional – not *unique*, i.e. many- (or one-) one. And the other is precisely if there are no objects suitable to constitute *values* for the purported function in question in the first place. The sense assigned to a putative functor may precisely carry sufficient information to enable us to show that the associated relation is not many-one (nor one-one) or that it fails to correlate the intended range of arguments with anything at all. Functors generally may have sense yet fail to present any function – so fail to have reference – if these conditions, of uniqueness and existence, are not met.

---

[15] For example, versions of both Aristotelian and Platonic conceptions of property are consistent with sparseness. For discussion of varieties of sparseness see Schaffer (2004).

[16] Cf. Schaffer op. cit.

Ok. So the question is whether a significant doubt is possible about whether they are met in the case of the functors introduced by (the best) abstractions. Might uniqueness be open to reasonable doubt in such a case? Here is a consideration that strongly suggests not. In order to entertain such a doubt, one needs to associate the relevant functor – 'S' – with an underlying *relation* and then to think of 'S(*a*)' as purporting to denote what is the only object so related to *a*. Uniqueness fails just when there more than one such object. But is there in general any conception of such a relation somehow conveyed as part of the sense that is attached to an abstraction operator by its implicit definition via the relevant abstraction principle? Take the case of Hume's principle and the associated cardinality operator, glossed as "the number of". In order to raise a meaningful doubt about uniqueness, we need to identify an associated relation such that the sense of "the number of Fs" may be conceived of as grasped *compositionally*, via grasping this relation plus the presumption of uniqueness incorporated in the definite article. The issue of uniqueness will be the issue of the many-oneness of this relation, – something which might ideally admit of proof. It is very doubtful however whether there is any good reason to think of the sense assigned to the cardinality operator by Hume's principle as compositional in this particular way[17]. And if not – if the operator is best conceived

[17] The issue is not uncontroversial. MacFarlane (forthcoming), like Boolos above, canvasses the view that numerical terms having the surface form 'the number of *F*s' are Russellian definite descriptions, presumed constructed using an underlying relational expression '*x* numbers the *F*s' – so that a sentential context '*A*(the number of *F*s)', with the definite description having wide scope, gets paraphrased as '$\exists!x$(*x* numbers the *F*s $\wedge$ *Ax*)'. On this view, at least as MacFarlane presents it, numerical terms are not genuine singular terms at all but a kind of quantifier. One could still enquire whether the postulated numbering relation is functional – i.e. whether, for any *F*, there always exists a unique *x* which numbers *F*. This would now be a substantial question, both as regards existence and uniqueness. This is not the place for detailed criticism of MacFarlane's proposal (for a response, see Hale & Wright (forthcoming)). But it is worth briefly separating some issues. One, obviously, is whether MacFarlane's proposal is viable at all. If Hume's Principle works as an implicit definition in the way we propose, it defines a certain functor – the number operator – *directly*. There simply is no underlying relational expression, from whose sense that of the functor is composed. One can of course define a relational expression, '*x* numbers *F*', to mean '$x = Ny{:}Fy$' – but this relational expression is evidently compositionally *posterior* to the number operator. The question, for the viability of MacFarlane's proposal, must therefore be whether '*x* numbers the *F*s' can be defined independently, without presupposing prior understanding of numerical terms. It is certainly not obvious that it can. But even if it can, the more important issue for present purposes is not whether one could introduce the number operator on the basis of an underlying relation, but whether one can, as we

as semantically atomic – then there is no scope for a significant doubt about uniqueness of reference, since there is no associated condition which more than one item might satisfy.[18]

It is, on the other hand, by no means as evident that there is no room for a significant doubt about existence.[19] The abstraction operator refers (to a function) only if the singular terms it enables us to form refer (to objects.) What reason is there to think that (any of) these terms so refer?

To fix ideas, think of the routine ways in which one might satisfy oneself that *any* singular term refers. Suppose, for instance, you take it into

---

contend, introduce it as semantically atomic – if so, then there is, for the reasons noted in the main text, no scope for a significant doubt about *uniqueness* of reference for terms formed by its means.

[18] Lest there be any misunderstanding, this concern needs sharply distinguishing from the concern about uniqueness raised by Harold Hodes in Hodes (1984). Hodes' concern is based on the fact that one can, consistently with the truth of Hume's Principle, permute the references of terms formed by means of the number operator, provided one makes compensating adjustments elsewhere (e.g. to the extension of the <-relation). Thus besides the 'standard numberer' which takes empty concepts to 0 as value, singly-instantiated concepts to 1, doubly-instantiated concepts to 2, and so on, there are many non-standard numberers – e.g. one which coincides with the standard numberer except in its values for empty and singly-instantiated concepts (1 and 0, respectively), compensating with a non-standard <-l relation which coincides with standard < except that we have 1 < 0. Hodes grants, at least for the sake of argument, that the number operator, as introduced by Hume's Principle, will denote a function – the trouble, he thinks, is that there is no unique, privileged such function that it can succeed in defining; rather, there are infinitely many such functions, between which it is powerless to discriminate. The problem is not that it is open whether "the number of" succeeds in picking out any operation whose values are, as required by functionality, unique but that it is unsettled whether it succeeds in picking out any unique such operation. This kind of doubt is not at issue in the text, and demands a quite different response. The crux is whether Hodes succeeds, as he claims, in demonstrating that a special, distinctively recalcitrant type of indeterminacy afflicts numerical terms as introduced by Hume's Principle – i.e. that we have something worse that the kind of permutational indeterminacy that can be engineered for expressions of any type, and is not confined to those purporting reference to abstracta. See Hale (1987), pp.220–224 for some further discussion.

[19] To be sure, one kind of doubt about existence *is* pre-empted by the same point. There can be no doubt whether certain items stand in a relevant underlying relation to anything if there is no relevant underlying relation – if there is no prior relation *R* such that 'the □ of *A*" is constrained to stand, if for anything, then for the unique *B* such that *R*(*A*, *B*). But those anxious about the existential consequences of abstraction principles will probably not be quickly persuaded that any proper doubt about existence here has to assume this pattern.

your head to try to show that "Bin Laden" is the name of a real man, rather than, say, the focal point of an elaborate fiction, promulgated by the CIA. There are various courses of action you might undertake to try to settle the matter, at least to your own satisfaction. But ultimately, what you need to do is gather evidence which is arguably sufficient for the truth of an *identity statement*, 'q = Bin Laden', for some 'q' whose reference to a real man is not in question. In this, 'q' might be a compendious definite description of the words and actions ("the man who said and did all of these things: …") of an unquestioned real man; or it might be a token demonstrative for the robed, bearded figure standing before you at the entrance to a cave in the Tora-Bora mountain range and revealed only after many days blind-folded travelling on the back of a donkey. The point generally is that verification of the existence of a referent for a term N is verification of a statement of the form: $(\exists x)(x = N)$. And the premium method for doing that is to verify an identity, $q = N$, where the existence of a referent for 'q' is not in doubt.

But this model exactly presupposes, of course, that the term in question is not *fundamental*. What about the case when N is a term purporting to stand for a new kind of object for which it is understood that no anterior means of reference need exist in the language – so that it is a given that there need be no suitable 'q'? In these circumstances verifying that N refers cannot be a matter of verifying that it co-refers with any expression, even a demonstrative, whose reference is not in doubt. So what can it be?

The only possible answer appears to be that such a feat of verification must consist in verifying – if not an identity statement linking the term in question with another whose reference is assured – then *some* form or forms of statement embedding the term in question whose truth *requires* that it refer: a statement, or range of statements, in which the term in question occupies a *reference-demanding* position. Such will be afforded by provision of the means to verify some form of *atomic* statement configuring such terms. Identity contexts are one kind of atomic statement. So abstraction itself – as a characterisation of putatively canonical grounds for the verification of such identity contexts – supplies a paradigm means, indeed an example it seems of the only foreseeable broad kind of means, for accomplishing the assurance required. If it is not acceptable, what would be acceptable instead?

There is a parallel here with material world scepticism. Imagine a situation in which we have only one means of reference to material objects – demonstratives, say, perhaps qualified by a sortal predicate: "that

man", "this tree", and so on (*material demonstratives*). And suppose we are challenged to produce a reason to think that *any* uses of such expressions succeed in referring. Once again, any such reason would have to be reason to think that certain statements – "that man is running", "that tree is tall" – embedding material demonstratives in reference-demanding ways are, in their context of use, true. And that in turn will demand a conception of what justifies taking such a statement to be true. Such a conception, so says the Sceptic, will be that of the occurrence of a certain pattern of experience – a pattern which might be fully described in terms of appearances, without commitment to entities of the kind in question. Since the evidence may be so described, independent assurance is wanted that successful referential use of the relevant expressions is possible in the actual world – a fortiori that there are middle sized physical objects out there to be referred to at all – before we may justifiably take such evidence to establish the truth of the appropriate type of statements.

Responses to this kind of scepticism about material objects are of course various. They include denying the 'neutralist' (Lockean) conception of experience it exploits, and allowing that conception but denying that any need is thereby entailed for *independent* corroboration of a material world ontology before experience can carry the evidential significance customarily accorded it. Abstractionism, in so far as it reads an ontology of abstracta into the commitments of the right-hand-sides of abstractions, stands comparison with the former (direct realist!) line. But the question I would press on the anxious metaphysician is this: if one is not content to acquiesce in a sceptical view of the referential aspirations of material demonstratives, how is it relevantly different with the terms introduced by abstraction?

6

The friend of abstraction may take some satisfaction in the dialectical situation just suggested. But it is actually very much *not* where I want – or promised – to end up. If the best that can be done with an obdurate doubt about the truth-preservingness of the transitions right to left across the instances of an abstraction is to make good an analogy with the relation between experience and material world claims as viewed by a Sceptic, then I have precisely not made good on what I characterised as of the essence of abstraction: the contention of the conceptual sufficiency of the truth of the

right-hand sides for the truth of the left. The whole point was to be that there *is* no metaphysical hostage in the transition, no need for an 'assist' from the World, and therefore no scope for doubt, even Sceptical doubt, that the requisite assistance is to hand. The best response to (E′), therefore, cannot rest upon a comparison between doubt about the inference, right to left, across an instance of an abstraction principle and scepticism about the reality of ordinary material objects. Rather, it has to be to make out a perspective from which abstraction actually involves nothing akin to the element of epistemological *risk* which scepticism finds in our purported cognitive commerce with the external world.

Let's step back. To ask, with Boolos, how we know that there is any function – hence, any objects to constitute its range of values – that behave as an abstraction principle demands is, in effect, to view the principle as proposed in a spirit of *reference fixing*: as 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 objects to satisfy. This is the conception of the matter articulated in the following passage:

> What did Locke realise about 'gold'? Effectively, that there is an element of blind pointing in our use of such a term, so that our aim outstrips our vision. Our conception fixes what (if anything) we are pointing at but cannot settle its nature: that is a matter of what's out there. One image of the way [Hume's Principle] is to secure a reference for its terms shares a great deal with this picture.[20]

On this conception, we 'point blindly', using the sortal concept and terms explained by an abstraction principle, in the hope of *hitting off* reference to a range of entities qualified to play the role that the principle defines, and it is accordingly readily intelligible how the process might fail – it goes with the 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 – let alone that it is secured to entities of which the principle states a necessary truth. Rather, this is something which needs to be verified as a by-product of our, so to say, *finding* a range of objects 'out there' to which the conception embodied in the principle is (necessarily) faithful. And of course if that is to be possible, the ob-

[20] Sullivan and Potter [1997], pp.145–146, quoted in Potter & Sullivan (2005).

jects in question must first be given to us under some *other* mode of presentation.

It is pointless to deny that it is *possible* to regard abstraction principles in this fashion. One can always ask, with respect to any particular domain of objects, whether there are any that are so related to the elements of the abstractive domain that identity and distinctness among them is tracked by the obtaining, or non-obtaining, of the relevant equivalence relation on pairs from that domain. It may be that in a particular case, the answer is not only affirmative but necessarily so – and in that case, the abstraction principle too will state a necessary truth, even when understood in the reference-fixing spirit. But this spirit – necessary for an 'anxious metaphysical' stance – is simply in flat tension with the abstractionist conception of the matter; indeed, it is to view abstraction principles in a manner inconsistent with their capacity to serve the process of abstraction itself. Properly viewed, the very stipulative equivalence of the two sides of an instance of an abstraction principle is enough to ensure both that it is not to be seen as proposed as part of a project of reference-fixing and that there *is no significant risk of reference failure*.

How can there be no such risk? In order to understand this, we need to be mindful again of the distinction between sparse and abundant properties and the role it can play in the semantics of predicates. For in general terms, the abstractionist metaphysics of abstract objects, and of reference to them – sometimes called *minimalism* – stands to the conception of the matter that underwrites the reference-fixing model as an abundant conception of properties stands to a sparse one. The analogy admittedly needs some care. On the most generous version of 'abundance' theory, there is for predicates, as remarked, no gap between sense and reference: the association of a predicate with a sense – a determinate satisfaction-condition, even if a necessarily unsatisfiable one – is enough to ensure the existence of a property – a way of being – to play the role of the reference of the predicate. It is not, by contrast, part of the minimalist view of the reference of singular terms introduced by abstraction to conceive of reference as bestowed purely by sense. But nor, according to the minimalist view, is reference secured by the abstraction's merely serving to introduce a conception of a kind of object whose exemplification requires a form of worldly co-operation going beyond anything that can be assured by the laying down of an abstraction principle which is in good standing by normal criteria – and so in particular features a *bona fide* equivalence relation. Anyone should agree that a justification for regarding a singular term as having ob-

jectual reference is provided just as soon as one has justification for regarding as true certain atomic statements in which it functions as a singular term. According to the abundant – "neo-Fregean" – metaphysics of objects and singular reference, such a justification is provided by the very manner in which sense is bestowed upon abstract singular terms, which immediately ties the truth conditions of self-identities featuring such terms to the reflexivity of the relevant relation. As with the abundant conception of properties, there is no additional gap to cross which requires "hitting off" something on the other side by virtue of its fit with relevant specified conditions, as the property of being composed of the element with atomic number 79 is hit off (or so let's suppose) by the combination of conditions that control our unsophisticated use of 'gold'. But nor is it the case that reference is bestowed by the possession of sense alone. The latter view, for singular terms, is Meinongianism. The abstractionist view agrees with the reference-fixing conception that it takes, over and above the possession of sense, the truth of relevant contexts to ensure reference. But it diverges from the reference-fixing conception in what it holds has to be accomplished before those contexts may justifiably be taken as true, and in how straightforward it views the accomplishment as being.

Can we make this clearer? On the abundant view of properties, predicate sense suffices for reference. But it is not the abstractionist view of singular terms that sense suffices for reference – the view is that the truth of atomic contexts suffices for reference. However everyone agrees with that. The controversial point is what it takes to be in position reasonably to take such contexts to be true. The point of analogy with the abundant view is that this is not, by minimalism, conceived as a matter of hitting off, Locke-style, some 'further' range of objects. We can perfect the analogy if we consider not simple abundance but the view that results from a marriage of abundance with Aristotelianism. Now the possession of sense by a predicate no longer suffices, more or less, for reference. There is the additional requirement that the predicate be true of something, and hence that some atomic statement in which it occurs predicatively is true. That is a precise analogue of the requirement on singular terms that some atomic statement in which they occur referentially be true. And abstractionist minimalism with respect to objects and singular reference is the exact counterpart of Aristotelian abundance with respect to properties and predicate reference. The Lockean conception, by contrast, is to be compared to the position of the 'sparse' opponent of the abundant Aristotelian who construes the relevant range of predicates as purporting reference to sparse

properties. On that view there is scope for a doubt whether a relevant predication is true, even when the subject meets the working satisfaction-conditions assigned to the predicate – for there may be no genuine property associated with meeting those conditions. Likewise on the Lockean view, there is scope for a doubt whether an abstract-identity is true even though the appropriate equivalence relation holds between the relevant elements in its field – for there may be no, as it were, 'sparse' – metaphysical Worldly – objects suitable to serve as the referents of the relevant abstract terms. The abstractionist conception of the truth of the right-hand sides of instances of good abstractions as conceptually sufficient for the truth of the left-hand sides precisely takes the terms in question out of the market for 'hitting off' reference to things whose metaphysical nature is broadly comparable to that of sparse properties, and assigns to them instead a referential role relevantly comparable to that of predicates as viewed by the abundant Aristotelian.

Let me begin to draw things together. Aside from the earlier, rather obvious remarks about the requirement of the truth of the corresponding Ramsey sentences, we have been rather neglecting question (M):

> What does the world have to be like in order for (the best examples of) abstraction to work?

What, in the light of the foregoing discussion, should now be said in answer? First, for each equivalence relation which is to underpin an abstraction – for all we have said, indeed, for *every* equivalence relation – there has to be an associated function taking each of the elements which are equivalent under the relation to a common object and no two inequivalent elements to the same such object. Second, the existence of such a function will of course require the existence of a properly behaved range of values. The anxious metaphysician and the abstractionist can agree thus far. Their disagreement concerns what it takes for that to be so. The anxious metaphysician thinks of the issue on the analogy of the existence of a sparse property: just as a predicate's being semantically well-behaved and even featuring in true atomic predications is no assurance that it refers to one of the real properties characteristic of the divisions in the metaphysical World, so the fact that the terms introduced by an abstraction behave as singular terms should and feature in what, if the abstraction is accepted, are well understood and often verified contexts, is no assurance that they refer to any of the real objects in the metaphysical World. One who subscribes

to this way of thinking then has to take a decision about whether they refer at all, with the minimalist conception of objects and singular reference on offer to play a role in a positive answer counterpart to that of abundant Aristotelian conceptions of property and predication. If the offer is spurned, the metaphysician will have to deny that abstractions can ever be simply stipulatively true. For the abstractionist, by contrast, there is no well-conceived objection to the unqualified stipulation of (the best) abstractions – if it seems otherwise, it is only because one is trying to combine their stipulative character with a reference-fixing conception of them – and the abundance of the entities thus recognised is simply the objectual counterpart of the abundance of abundant properties.

These remarks are not a *defence* of minimalism but merely a reminder – since it seems that one may be needed – of the kind of background thinking about objects and ontological commitment which undergirds the abstractionist view. Perhaps this background thinking constitutes a 'metaontology'. If so, then there is much more to say about the spirit of this metaontology – especially about the sense, if any, in which it is happily described as 'platonist'. But if it is accepted, the answer to question (M) could not be simpler: a world in which abstraction works is a world in which there are equivalence relations with non-empty fields.

REFERENCES

David Armstrong (1979) *A Theory of Universals*, Cambridge: Cambridge University Press.

George Bealer (1982) *Quality and Concept*, Oxford: Oxford University Press.

Paul Benacerraf (1973) 'Mathematical Truth' *Journal of Philosophy* 70, pp.661–680.

George Boolos (1990) 'The Standard of Equality of Numbers' in George Boolos ed. *Meaning, and Method: Essays in Honor of Hilary Putnam* Cambridge: Cambridge University Press 1990, pp.261–277; reprinted in Boolos (1998), pp.202–219.

George Boolos (1997) 'Is Hume's Principle Analytic?' in Richard G. Heck, Jr., ed. *Logic, Language, and Thought* Oxford: Oxford University Press 1997; pages references are to the reprint in Boolos (1998), pp.301–314.

George Boolos (1998) *Logic, Logic, and Logic* Cambridge, Mass.: Harvard University Press, 1998.

Ross Cameron (2007) 'Truthmakers and Ontological Commitment: Or How to Deal with Complex Objects and Mathematical Ontology without Getting into Trouble' http://www.personal.leeds.ac.uk/%7Ephlrpc/research.htm.

Matti Eklund (2006) 'Neo-Fregean Ontology' *Philosophical Perspectives* 20, pp.95–121.

Bob Hale (1987) *Abstract Objects* Oxford: Basil Blackwell
Bob Hale (1994) 'Is Platonism Epistemologically Bankrupt?' *Philosophical Review*, Vol.103, 2, pp.299–325.
Bob Hale (2000) 'Reals by Abstraction' *Philosophia Mathematica* (3) Vol.8, pp.100–123, reprinted in Hale & Wright (2001).
Bob Hale (forthcoming) 'Neo-Fregeanism and Quantifier-Variance: Chairman's Remarks' (forthcoming in *Proceeedings of the Aristotelian Society*).
Bob Hale & Crispin Wright (1994) 'A reductio ad surdum? Field on the contingency of mathematical objects' *Mind* Vol.103, No.410, pp.169–184.
Bob Hale & Crispin Wright (2001) *The Reason's Proper Study: Essays towards a Neo-Fregean Philosophy of Mathematics* Oxford: Oxford University Press.
Bob Hale & Crispin Wright (2002) 'Benacerraf's Dilemma Revisited' *European Journal of Philosophy* Vol.10, No.1, pp.101–129.
Bob Hale & Crispin Wright (2009) "The *Metaontology* of Abstraction" in David Chalmers, David Manley and Ryan Wasserman, eds., *Metametaphysics: New Essays on the Foundations of Ontology*, Oxford: Oxford University Press
Bob Hale & Crispin Wright (forthcoming) "Focus Restored", reply to MacFarlane (forthcoming) in Øystein Linnebo, ed. Special Issue of *Synthese* on the Bad Company Problem.
Katherine Hawley (2007) 'Neo-Fregeanism and Quantifier Variance' *Aristotelian Society Supplementary* Volume LXXXI, pp.233–249.
Eli Hirsch (2002) 'Quantifier Variance and Realism' *Philosophical Issues* 12, *Realism and Relativism*, pp.51–73.
Harold Hodes (1984) 'Logicism and the Ontological Commitments of Arithmetic', *Journal of Philosophy* Vol.81, pp.123–149.
David Lewis (1986) *On the Plurality of Worlds*, Oxford: Basil Blackwell.
John MacFarlane (forthcoming) 'Double Vision: Two Questions about the Neo-Fregean Programme' in Øystein Linnebo, ed. Special Issue of *Synthese* on the Bad Company Problem.
D.H. Mellor & Alex Oliver (1997) 'Introduction', in *Properties*, Oxford: Oxford University Press, pp.1–33.
Michael Potter & Peter Sullivan (2005) 'What Is Wrong With Abstraction?', *Philosophia Mathematica* 13, pp.187–193 .
Chris Swoyer (1996) 'Theories of Properties: From Plenitude to Paucity', *Philosophical Perspectives* 10, pp.243–264 .
Jonathan Schaffer (2004) 'Two Conceptions of Sparse Properties', *Pacific Philosophical Quarterly* 85, pp.92–102.
Stewart Shapiro (2000) 'Frege Meets Dedekind: A Neo-logicist Treatment of Real Analysis' *Notre Dame Journal of Formal Logic* Vol.41, pp.335–364.
Theodore Sider (2007) 'Neo-Fregeanism and Quantifier Variance' *Aristotelian Society Supplementary Volume* LXXXI, pp.201–232.
Peter Sullivan and Michael Potter (1997) 'Hale on Caesar', *Philosophia Mathematica* 5, pp.135–152.
Alan Weir (2003) 'Neo-Fregeanism: An Embarrassment of Riches' *Notre Dame Journal of Formal Logic*, 44, pp.13–48.

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

Crispin Wright (1990) 'Field and Fregean Platonism' in A.D. Irvine, ed. *Physicalism in Mathematics*. Dordrecht/Boston/London: Kluwer, reprinted as Essay 6 in Hale & Wright (2001).

Crispin Wright (1999) 'Is Hume's Principle Analytic?' *Notre Dame Journal of Formal Logic*, 40, pp.6–30; reprinted in Hale & Wright (2001).