What is wrong with abstraction?*

Michael Potter†
Peter Sullivan‡

Abstract

We correct a misunderstanding by Hale and Wright of an objection we raised earlier to their abstractionist programme for rehabilitating logicism in the foundations of mathematics.

1 The issue

Consider an abstraction principle of the form

$$\Sigma(F) = \Sigma(G) \equiv F \sim G,$$

where 'Σ' is a term-forming operator and $\sim$ is an equivalence relation on concepts in whose definition 'Σ' does not occur. There are three ways of taking this:

1. Predicative;
2. Free;
3. Impredicative.

Taken in the predicative manner the abstraction principle tells us the identity conditions for a new kind of object. No presumption is made that these new objects fall in the range of the quantifiers antecedently understood. In technical terms what this amounts to is that we adopt a many-sorted logic, so that the quantifiers on the right hand side of the equivalence range over objects of one sort while the objects referred to on the left hand side are of another sort. Predicative abstraction principles are entirely harmless: they cannot lead to contradictions of the sort that stymied Frege, but equally they cannot deliver substantive mathematical theories such as classical arithmetic.[1]

*We are grateful to audiences in Cambridge and St Andrews for their comments on earlier versions of this paper.

†Fitzwilliam College, Cambridge, CB3 0DG, and Department of Philosophy, University of Stirling, Stirling, FK9 4LA, United Kingdom.

‡Department of Philosophy, University of Stirling, Stirling, FK9 4LA, United Kingdom.

[1]See Potter (2004, appendix C) for further discussion of this point.

The second way to understand an abstraction principle is in the context of a single-sorted free logic. If we understand it in this manner, we do suppose that any objects referred to by the terms on the left hand side of the equivalence are of the sort that fall within the range of the quantifiers on the right hand side, but we do not commit ourselves to the assumption that there *are* any such things. We leave it open, that is to say, whether the terms on the left hand side refer to anything. Whatever commitment the abstraction principle involves is then purely conditional: *if* there are such things as Σs, it says, these are their identity conditions. Once again, an abstraction principle understood in this manner is harmless, but in a rather different way from before. In order for it to be of much use, it will have to be afforced with an additional assumption that, at least in certain cases, the terms on its left hand side do refer to objects. At that point what we have will be a restricted case of our third, impredicative sort of interpretation.

This third way of interpreting the abstraction principle involves assuming that the terms on the left hand side do indeed refer and that what they refer to are objects falling within the range of the quantifiers on the right hand side. This third way of interpreting the principle holds out the prospect of delivering as consequences significant mathematical theories such as arithmetic, analysis or set theory; but at the same time it is philosophically problematic in ways that the first two are not.

Of course, whatever the history of the philosophy of mathematics shows, it shows at least that arithmetic, analysis and set theory are problematic, so to the extent that they can be shown to be logical consequences of impredicatively understood abstraction principles we might expect those abstraction principles to be correspondingly problematic. What we want to discuss now, though, is whether basing these mathematical theories on an abstraction principle has any special philosophical virtues that alternative treatments do not possess. After all, each of the mathematical theories in question has been given an axiomatic foundation that does not involve an abstraction principle. Arithmetic, for instance, can be based on the Peano-Dedekind axioms; analysis can be based on the assumption that the real numbers form a complete ordered field. Those who have canvassed the idea of basing these mathematical theories on impredicative abstraction principles — most notably Hale and Wright (2001) — have claimed that the structural pattern exemplified by these principles gives them special properties which deliver satisfying solutions to at least some of the problems in the philosophy of mathematics. What is at issue here, then, is whether impredicative abstraction principles are indeed special in this way. We can, for instance, argue about whether taking Hume's principle as a foundation for arithmetic has advantages over the Peano-Dedekind axiomatization as a foundation of arithmetic such as elegance or economy. But does it, as Hale and Wright claim, have a further advantage of a wholly different kind that arises from the structural pattern that it shares with other abstraction principles?

In an earlier article (Potter and Sullivan 1997) we attempted to articulate one way of thinking about the problems that are involved in regarding an impredicative abstraction principle as providing a route to an understanding of a

range of objects.

> 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. (It does not share everything. A commonly accepted thought about natural kind terms is that the conception that guides their use may be in significant respects false to their reference without undermining that reference. The thought that is relevant here is only that the guiding conception of a thing will not contain the whole truth about it, not that it will perhaps not be wholly true.) (pp. 145–6)

Hale and Wright responded to this point in their article 'To bury Caesar', but seem to have misunderstood it. They represent what we were arguing as being

> not that numbers may have an additional nature, undisclosed by Hume's principle, in the same way that MPs have an additional nature undisclosed by the principle *Members of Parliament*.[2] *That* concern ... is effectively precluded by the admission that Hume's Principle is a necessary truth. (2001, p. 395n.)

But as a representation of our intentions this is wrong. A central (and several times repeated) contention of our paper was that to accept a certain principle as true of a range of objects is not to admit that this principle embodies a fully adequate or exhaustive conception of those objects. Nor will this follow if the principle is accepted as a necessary truth about them. For present purposes we are happy to grant that Hume's Principle is a necessary truth. We denied, and continue to deny, that granting this *in itself entails* that numbers cannot have an additional nature undisclosed by Hume's Principle in the same sort of way that MPs have an additional nature undisclosed by the apparently similar principle *Members of Parliament*.

The source of the misunderstanding between us thus seems to be that Hale and Wright have a conception of abstract objects according to which there can be no more to their nature than follows from their identity conditions, whereas we do not. Our purpose in this note, therefore, is to try to make it plausible that abstract objects might have such a nature.

At this point, though, there is an expository difficulty. Much of the discussion in the literature, including the paper of ours to which Hale and Wright responded, has focussed on the project of using Hume's Principle as a route

[2] *Members of Parliament* is the principle that two people have the same MP if and only if they live in the same parliamentary constituency. We introduced this principle in (1997) as an example of an abstraction principle which tells us something, but plainly not everything, about the objects (MPs) it refers to.

to the natural numbers. What we wish to claim about this case is only that, for all Hale and Wright have said, numbers may have a nature additional to that implied by their satisfaction of Hume's Principle. We do not claim that numbers in fact have any such nature. Our point was only that the contention, for a given range of mathematical objects, that they have no such additional nature will require an argument specific to the case in question, and that Hale and Wright have neither supplied such an argument nor seen the need for it.

Given all this, the case of natural numbers is an unpromising one to use dialectically to persuade anyone of the possibility to which we are trying to draw attention. For this reason we shall switch examples and discuss instead the case of set theory, which exemplifies much better the point at issue.

2 The example

Consider, then, the following abstraction principle:
($\mathrm{Ext}_{\mathrm{Good}}$)
$\mathrm{Class}_{\mathrm{Good}}(F) = \mathrm{Class}_{\mathrm{Good}}(G) \equiv ((\mathrm{Good(F)} \vee \mathrm{Good(G)}) \supset (\forall x)(Fx \equiv Gx))$,

where 'Good' is some suitable second-order predicate. Which predicates count as suitable here is controversial. At the very least, in order that the abstraction principle should not be inconsistent, we need the number κ of objects in the domain to be at least as great as the number (counted extensionally) of good concepts.[3] Since the number of concepts in total (counted extensionally) is 2^κ and Cantor's theorem tells us that $\kappa < 2^\kappa$, we are constrained to place some restriction on what concepts we should regard as good; we cannot count *all* concepts as good without falling into contradiction, as Frege found out to his cost.

One restriction which has commonly been proposed is to take 'Good' to mean 'Small', where a concept is said to be small if it is not equinumerous with the universal concept (the concept under which all objects fall). But of course this notion of goodness is not in itself privileged over any other. If we are to settle on the abstraction principle $\mathrm{Ext}_{\mathrm{Small}}$, more plainly need to be said to explain why.

The motivating idea here is known as 'limitation of size'. The thought is that some things have an extension just in case there are not too many of them. But *why* should there being a lot of them prevent some things from forming a class? As one of us has remarked elsewhere (Potter 2004, p. 228), the literature of set theory is curiously free of arguments for believing that this might be true. The principle was first mentioned in print by Russell (1906), but the justification he offers for it is little more than bluster. And Von Neumann (1925), although he appeals to the distinction between classes that are "too big" and ones that are not, makes no attempt whatever to motivate the distinction.

[3]To count concepts extensionally means to count them as different just in case they differ in respect of the objects that fall under them.

The alternative to limitation of size that has found favour in the literature is of course the iterative conception of set. According to this conception sets are extensional, just as the abstraction principle decrees, but are in addition to be thought of as composed or made up from their members. One of us has written at length (Potter 2004) about the difficulty involved in explaining what this amounts to. The conclusion there drawn is that it is best thought of as a primitive metaphysical relation of dependence or presupposition that holds between a set and its members. The significance of this relation of dependence or presupposition is that it allows us not merely to remove the set-theoretic contradictions but to motivate their removal: if a set depends on its members, then a set of all sets would, if it existed, depend on itself; and that, we might think, is obviously impossible.

It would be foolish to deny that the iterative conception of set based on a primitive dependency relation has its difficulties. Nevertheless, it does provide a satisfying explanation for and resolution of the set-theoretic paradoxes, which the limitation of size conception so far has not supplied. Our purpose here is not to argue for one rather than the other, but rather to note that Hale and Wright's conception of abstract objects as having no essence not determined by their identity conditions simply rules out our thinking of sets as having any such internal structure as the iterative conception seems to require. They cannot hope to find a notion of goodness whose corresponding abstraction principle neatly delivers only the well-founded classes, because any characterization of what it is for a concept to have a well-founded extension will have to use the 'Class' operator in its definition. They therefore have little choice but to settle for limitation of size, not because it answers to their conception of what classes are, but simply because it can be fitted into the abstractionist framework. This seems to us to be a case probably of backing the wrong horse, and certainly of backing it for the wrong reason.

The abstractionists' procedure of basing the theory of classes on $\mathrm{Ext}_{\mathrm{Small}}$ does not, of course, prevent them from constructing a theory of well-founded classes within this framework. It is easy to define in second-order logic with a 'Class' operator what it is for a class to be well-founded and, having done so, to show that the standard theory holds for such classes (see Boolos 1989). But this is not to adopt the iterative conception, nor, crucially, to make it available as an explanation for the non-existence of a universal class. The abstractionist may indeed claim that well-foundededness is an interesting property for a class to have, and hence that the well-founded hierarchy is especially worthy of study. But what was wanted was an explanation of, and solution to, the paradoxes of set theory: averting your eyes does not count as a solution.

3 The moral

There are two uses of abstraction principles that are clearly unobjectionable. If, first, we view Hume's principle (or Basic Law V, for that matter) predicatively, it is harmless and plausible, but quite useless as a route to substantial mathematical theories. If, alternatively, we regard it as a truth about objects of

which we already have a conception, then it can in appropriate cases be used, along with second order logic, as the axiomatic base for a development of part of mathematics. But the epistemological crux will then be our explanation for its *truth*: this explanation will have to appeal to features of the objects in question. In the case of arithmetic, for instance, we need to explain how we come to know that Hume's principle is true about numbers, conceived as antecedently grasped abstract objects. And there is no obvious reason to think that this will be any easier than the corresponding task for the Dedekind-Peano axioms.

Whatever virtue Hale and Wright's project possesses seems to us to stem from an attempt to draw on what is most plausible in both these uses of abstraction principles. Their thought is that we can, by means of an abstraction principle, advance to a grasp of objects of which we did *not* already have a clear conception. There is a sense in which the objects to which the terms on the left hand side of the principle refer are new: they are, one might say, new to *us*. (This is why they think that we owe no epistemological debt at this point.) But in another sense the objects are not new: they already fall within the range of the quantifiers on the right hand side of the principle. How is that? Hale and Wright think that they owe no more explanation for this than that these quantifiers are genuinely universal. If they quantify over absolutely everything, they will automatically include in their domain anything of which we may come to a conception.

But this is highly questionable. If we quantify over everything that our current conceptual scheme allows for, then we are in a perfectly good sense (namely, the only sense our current conceptual scheme allows for) quantifying over absolutely everything. Can we, though, meaningfully quantify over everything we may in the future find reason to include in our world view? It would be wrong, because inconsistent, to deny that we can: we just did so in the previous sentence. But this negative point is hardly enough to justify what has emerged as the central presupposition of Hale and Wright's approach: that there are undeniably *some* things we can say now about things undreamt of in our philosophy is surely too thin a basis for assuming that we have a grasp of full second-order logic as applying to them.

References

Boolos, G. (1989), 'Iteration again', *PhilTopics*, 17: 5–21

Hale, B. and Wright, C. (2001), *The Reason's Proper Study*, Oxford: Clarendon Press

Potter, M. (2004), *Set Theory and its Philosophy: A critical introduction*, Oxford University Press

Potter, M. and Sullivan, P. (1997), 'Hale on Caesar', *Phil. Math.*, 5: 135–52

Russell, B. (1906), 'On some difficulties in the theory of transfinite numbers and order types', *Proc. Lond. Math. Soc.*, 4: 29–53 (repr. in Russell 1973, pp. 135–64)

——— (1973), *Essays in Analysis*, Allen and Unwin

Von Neumann, J. (1925), 'Eine Axiomatisierung der Mengenlehre', *J. reine angew. Math.,* 154: 219–240