Abstractionist class theory: Is there any such thing?

Michael Potter

Final draft

Timothy Smiley and I have published two papers jointly (2001; 2002). Their purpose was to point at problems for the abstractionist (or neo-Fregean) account of arithmetic. In the course of our work on these papers Timothy taught me a great deal about how philosophy should be written: he simply would not let me include a sentence whose meaning was not completely clear to him. The current article, which extends the exploration of difficulties with abstractionism from the case of numbers to that of classes, provides an opportunity for readers to judge to what extent I have taken this lesson to heart.

1 Abstractionism

An abstractionist, as I shall use the term here, is someone who attempts to ground an account of part of mathematics on an *abstraction principle*, i.e. an equivalence of the form

$$\Sigma(F) = \Sigma(G) \text{ iff } F \sim G,$$

where $\sim$ is some suitably chosen equivalence relation on properties which can be defined without using the term-forming operator Σ. It is standardly assumed by abstractionists that the background logic is second-order. This assumption is not critical to the formulation of the position: there is nothing essentially second-order about the *notion* of an abstraction principle. However, the assumption *is* critical to the non-triviality of abstractionism: in a purely first-order context, the consequences of abstraction principles turn out to be disappointingly anodyne. Much could be said about the ontological assumptions implicit in the abstractionist's adoption of second-order logic, but I shall not say it here.

What I want to mention instead is the *prima facie* appeal of abstractionism as a route to the grasp of a range of abstract entities such as natural numbers, real numbers, or classes. To understand this appeal we need to see what distinguishes abstractionism from the axiomatic method. Suppose that we wish to introduce a term-forming operator Σ. The axiomatic method, in bare outline, consists in grounding a theory about Σs by laying down as axioms one or more sentences involving Σ which are sufficient to have the sentences of the theory as logical

consequences. The difficulty, of course, is to explain the status of the axioms. If the account is not to amount to a version of formalism, we cannot simply choose the axioms at will, but need to be able to regard them as *true*, since then it will follow that their consequences are likewise true. But explaining this leads to familiar problems. If we treat the axioms as truths that we have come to recognize about a domain of abstract objects of which we already have a grasp, then we owe an epistemological debt: how did we acquire our knowledge of these truths? If, on the other hand, we treat the axioms as true by *stipulation*, then we have the difficulty of explaining what constrains us in our stipulations. We obviously cannot stipulate just *any* sentences to be true. We cannot, for instance, make inconsistent stipulations. Hilbert famously tried to persuade Frege that in mathematics consistency is the only constraint (Frege 1980, pp. 31–52), but many (including Frege himself) have not been convinced.

Abstractionism offers the hope, at least, of a route out of this impasse. If our axiom has the particular syntactic shape of an abstraction principle, then there may be a special way of grasping it as true which does not require us to appeal to a problematic contact with abstract objects but does not lapse into formalism either. This special way involves, according to the abstractionist, two steps: first, we grasp the content of the equivalence relation on the right hand side of the abstraction principle; then we reconfigure it so that it turns into the content of an equality relation between objects on the left hand side.

It is controversial whether this process works as advertised (or even whether it really exists). What is plain, at the very least, is that if there is such a process, then its mechanism is delicate, and a lot can go wrong. One of the main reasons for this delicacy is that it is unclear whether we should think of the objects to which the abstraction principle introduces us as 'new' or 'old'.

In one sense, they are new: the abstractionist's solution to the Julius Caesar problem consists in insisting that the kind of objects the Σs are is wholly revealed by the abstraction principle which introduces them. They do not have a history in some other guise. Thus, in particular, Julius Caesar, who manifestly does have a history, is not a natural number.

But in another sense, the abstractionist has to insist that the objects are old. They are old because they fall within the range of the quantifiers which occur in our specification of the equivalence relation on the right hand side of the abstraction principle. This is what is known as *impredicativity*. It is important for pragmatic reasons, since it is what gives abstraction principles their power. Predicative abstraction principles, which introduce us to altogether new entities not falling in the range of the quantifiers on the right hand side, are completely harmless or, if you prefer, useless: they cannot serve as a foundation for mathematics.

The delicacy of the situation thus consists in the need to explain how the Σs may be sufficiently new that the abstractionist does not owe a prior explanation for our grasp of them, yet sufficiently old that they already occur in the range of the quantifiers as understood by someone who grasps the right hand side of the abstraction principle. The position thus depends for its coherence on the conceptual gap between universal and existential quantification, on the one

hand, and infinite logical products and sums, on the other. As Frege put the point, 'If I utter a sentence with the grammatical subject "all men", I do *not* wish to say something about some Central African chief wholly unknown to me.' (1984, p. 227)

The delicacy just noted explains the restriction I imposed when I defined the abstractionist position, that the equivalence relation on the right hand side of an abstraction principle should be definable in terms not involving the operator Σ to which the principle is intended to introduce us. I shall call a principle *circular* if it does not obey this restriction, i.e. if the equivalence relation occurring on the right hand side is defined by using the term-forming operator introduced on the left hand side.

2 Abstractionist class theory

So much for abstractionism in general. What of the project of using it as a basis for the theory of classes? (Here and throughout, I call extensional entities introduced to us by means of an abstraction principle 'classes'. I want to leave open the possibility that the theory to which the abstraction principle is adjoined *already* speaks of extensional entities of a different kind: if it does, I shall call these other entities 'sets'.)

At first sight, the obvious abstraction principle for classes would be

(V) $\{x : Fx\} = \{x : Gx\}$ iff $\forall x Fx \equiv Gx$,

but of course neo-Fregeans cannot adopt it in this unrestricted form, since it is inconsistent. So instead they restrict it to the case where F and G are class-forming. In order to shoehorn this into the form of an abstraction principle, they usually make the 'class-of' operator everywhere defined by assuming that $\{x : Fx\}$ is some specially chosen object when F is not class-forming. So instead of the conditional form

If F and G are class-forming, then $\{x : Fx\} = \{x : Gx\}$ iff $\forall x Fx \equiv Gx$,

their abstraction principle takes the form

(V') $\{x : Fx\} = \{x : Gx\}$ iff either $\forall x Fx \equiv Gx$ or neither F nor G is class-forming,

which is of the right syntactic shape to be an abstraction principle and sends all non-class-forming properties to the same object.[1]

But of course this is no use on its own: abstractionists need an account of which properties are class-forming. One constraint, of course, is that our definition of 'class-forming' should make the abstraction principle consistent. It is worth spelling out what this amounts to. The point to note is that Russell's paradox can be thought of as a counting argument. Our abstraction principle

[1] This idea is due to Boolos (1989).

(V') asserts the existence of a one-to-one function from the class-forming concepts (individuated extensionally) into the objects. Such a function will exist if and only if the number of concepts (counted extensionally) is less than or equal to the number of all objects. So the requirement that we should avoid contradiction radically underdetermines our response to Russell's paradox: there is plainly a great number of possible definitions of 'class-forming' that obey the constraint just mentioned.

Stating the consistency requirement in the form of a constraint on how many class-forming concepts there are is slightly more than a mere change of notation, since it makes the point that the issue is global, not local: Russell's paradox does not isolate one particular concept as problematic. This is an obvious point, but much of the recent literature on the topic has been bedevilled by a curious failure to keep it in mind. One symptom of this is the tendency of abstractionists to use the term 'good' for what I have here called 'class-forming', and 'bad' for its complement. That is dangerously misleading: the counting argument does not show that there is anything especially bad about the concepts which fail to be class-forming. The proof of Russell's paradox provides us with a method, given any putative one-to-one function from concepts to objects, of finding a concept which does not have an image under this function. What it does not do is to define *one* concept which, independent of the function currently under consideration, is intrinsically unmappable. 'Bad', in abstractionist set theory at least, is a relative term.

3 The iterative conception

In an earlier article (2005), Peter Sullivan and I suggested that abstractionists cannot hope to rely on the iterative conception to motivate their choice of abstraction principle. Why not? I think it will be worthwhile to spend some time explaining this point, since there seems to be considerable confusion about it among the abstractionists themselves. Let us distinguish two strategies that an abstractionist might use to try to generate the iterative hierarchy. The first is to find an abstraction principle which generates a lot more classes than the well-founded ones, and then restrict to the well-founded ones by definition. The second is to tailor the abstraction principle to well-foundedness from the start. Consider these two strategies in turn.

The first strategy is hopeless as a response to the paradoxes. What we wanted was a conception of classes which we have reason to believe in. If we have reason to believe the abstraction principle, we thereby have reason to believe in the existence of all the classes to which it introduces us. If you choose to pick out some of those classes with a special name ('set', for instance) and devote your attention to studying them, that is your right. But this plainly does not amount to adopting the iterative conception of class.

When we made this point before (2005, p. 192), Hale and Wright responded by saying that it seems confused. The procedure of labelling the well-founded classes as sets, they said,

> would indeed provide an explanation of the non-existence of a universal set, for it would provide the means to show that the abstract it associated with the universal concept would not comply with the definition it offered of set, the motivation for which could be precisely that well-foundedness is required by the asymmetric dependency of sets on their members. (2008, p. 194)

But this is surely missing the point. The issue is how to explain the paradoxes. On the procedure under consideration, it would have to be the case that the abstraction principle for classes had *already* disposed of them. Whether the class to which the universal concept gives rise is well-founded or not is—as Hale and Wright elsewhere (p. 193n) seem to acknowledge—neither here nor there.

What, then, of the second strategy. This, let us recall, was to adopt an abstraction principle tailored to well-foundedness from the start, so that the only classes it gives rise to are the well-founded ones. This would neatly avoid the problem with the first strategy just noted. Unfortunately, however, it cannot be done: no non-circular abstraction principle can be expected to generate just the well-founded classes. The reason, in a nutshell, is that the definition of well-foundedness depends on the membership relation, which in turn is defined, according to the abstractionist methodology, in terms of the class-forming operator.[2]

One abstractionist who has examined this issue in some detail is Roy Cook. In a recent and technically inventive article, he has attempted to find an abstraction principle which generates only the well-founded classes. However, the abstraction principle Cook proposes for this purpose—what he calls *Newer V*—is circular in the sense mentioned earlier. That is to say, the definition of the equivalence relation on the right hand side of the principle makes reference to the term-forming operator which the principle is attempting to introduce.

Cook is aware of this circularity, but he is not as troubled by it as I think he should be. 'On the face of it,' he says, the objection that *Newer V* is circular 'does not seem overly compelling'.

> As is well known, for the implicit definitions codified in abstraction principles such as *Hume's Principle* and *New V* to do the work intended, the quantifiers on the right-hand side of the biconditional must range over all objects, including the abstracts being introduced (and defined) on the left. Once this is accepted, there seems little reason not to explicitly refer to extensions in our definition of the identity conditions for extensions, since we are already forced to quantify over them in such a definition. (Cook 2008*b*, p. 430)

But this is to collapse precisely the conceptual gap I mentioned earlier which gives impredicative abstractionism its hope, however slender, of success. It has been one of the themes of all my writings on the neo-Fregean project to suggest,

[2]See Jané and Uzquiano (2004) for a detailed study of the non-well-founded models of non-circular abstraction principles.

against Hale and Wright, that the understanding we need to have of a domain if we are to quantify over it precludes us from having the freedom they claim in introducing objects that fall within it. But I readily concede that the point is by no means obvious. Hale and Wright, on their part, have responded in articles of their own to some of the points I have raised. What is agreed on all sides is that the matter is by no means trivial. If Cook were right, however, the dispute would be still-born: it would be quite plain that there is no explanatory gap for the abstractionist to exploit.

Perhaps Cook sees some merit in this point, since he takes the trouble later in the same article to give an alternative formulation—he calls it *Newer V**—which attempts to avoid circularity. However, the attempt fails: *Newer V** makes use of a relation whose definition involves the term-forming operator it attempts to introduce. A circle with two steps in it rather than one is still a circle. And as far as I can see, any attempt to repair Cook's proposal is bound to founder because of the cumulative nature of the iterative hierarchy he aims to generate.

4 On conceptions

I have already talked a couple of times of the 'iterative conception' of classes. I want shortly to focus on another conception, known as the limitation-of-size conception, and ask what reason an abstractionist might have to adopt it. But before I do, I should perhaps say a little about what asking this question amounts to. What is it for an abstractionist to 'adopt a conception'?

To ask this is to return to the heart of the delicacy in the abstractionist's position that I mentioned earlier. For one might well suppose that to adopt a conception of a range of objects it is necessary to have the kind of problematic epistemological access to the objects whose avoidance was one of the motivations for the neo-Fregean programme in the first place. If so, the abstractionist programme can have the advantages claimed for it only on condition that it can be prosecuted without any appeal to such a conception.

But remember that the appeal of abstractionism is based on the idea that the abstraction principle is a route to the grasp of a concept. And not all abstraction principles *could* be such a route: in the case of an inconsistent principle such as (V) there is no concept to grasp. And even if a principle is logically consistent, it may conflict with other beliefs we already have or might wish to adopt: the 'parity principle', for instance, is logically consistent on its own but entails that there are only finitely many objects (Boolos 1998, pp. 214–5). The notion of a conception is intended to help us with this difficulty. It would be irresponsible simply to pick our abstraction principles blind, so instead we select those that conform to a conception we can articulate of a range of objects of a certain kind.

This is not a way of speaking that all abstractionists are especially comfortable with: to them, it evidently sounds suspiciously close to the Gödelian platonist's appeal to intuitions. In response to an earlier attempt by me and Sullivan to press this understanding of the matter, for example, Hale and Wright

(2008) said that our 'conception of the need for a kind of explanation of the paradoxes', which we had 'reproach[ed] abstractionism for failing to supply', seemed

> to be wholly driven by, and dependent upon, an extreme realism about the nature of sets—a realism extreme enough to make sense of the project of trying to diagnose the respects in which naive comprehension goes astray as, so to speak, a thesis of the natural history of sets. When they call for explanation, what they seek is to understand the principles of set existence ... in a realist spirit which seeks to winkle out the divine truth for which Law V fumbles and to arrive at what it is about the essential nature of sets which underwrites that truth, whatever it is.

'Abstractionism,' Hale and Wright went on to insist,

> wants nothing to do with the kind of realism that creates this demand for 'explanation'. Indeed it is the conviction that any such position is epistemologically hopeless, coupled with receptiveness to a 'face-value' construal of mathematical theory, that provides the single most influential motive for the abstractionist project in the first place. It is important that philosophers interested in these issues register the point that Potter's and Sullivan's insistence on the putative explanatory shortcomings of abstractionism in this area rests on a demand which, absent a full-blown platonist conception of the subject matter, should simply be rejected as misconceived. (p. 195)

I remain unconvinced, however, that the demand for explanations that Sullivan and I attempted to articulate in our earlier articles (see also Potter and Sullivan 1997) presupposes realism in relation to the subject matter in question, far less that this realism need be thought of as 'extreme'. I do not imagine that the explanation for the paradoxes which I am seeking is a 'divine truth', or even that it is unique. It seems quite plausible that there might be several different explanations, each leading to a somewhat different theory. And talk of 'conceptions' is certainly not the prerogative of the realist: Michael Dummett's articulations of his anti-realist philosophy of mathematics are littered with such talk.

If we *were* to abjure the demand for explanation, as Hale and Wright recommend, what else would there be to help us choose which abstraction principles to adopt? One strategy abstractionists might use is to abandon completely the search for principled underpinnings for our theory and justify it instead by its consequences.

> One might take the view that there is then only a pragmatic issue—that of finding a way of marginalising the exceptions in a way that somehow conserves the spirit of the original incoherent form of comprehension for sets and leaves scope for a powerful theory with as

> much as possible of the mathematical interest that drove the original. Abstractionism can certainly accommodate this project: the strategy will be to try to refashion the concept of set, exactly as encapsulated in suitable descendants of Law V, to underwrite a theory of suitable strength. There may be a number of ways of getting interesting results, none with any but a pragmatic claim to be superior to alternatives. (Hale and Wright 2008, p. 194)

Readers familiar with the history of the subject will of course recognize the strategy Hale and Wright are suggesting here. When Russell realized that he could not base mathematics on the ramified theory of types alone, but had to assume the axiom of reducibility to get what he wanted, he came up with the regressive method as a justification (see Russell 1973*b*). The underlying theory was to be chosen so that it had as consequences those mathematical propositions he already believed to be true and did not have as consequences any mathematical propositions he already believed to be false. More recently, Maddy (1988, p. 485) has used the term 'one step back from disaster' as a label for something similar.

Perhaps wisely, Russell himself never published his paper on the regressive method. I shall not here itemize in detail all the difficulties which the method faces when it is used to ground a logicist account of the foundations of mathematics. What I do want to note is the distance between the pragmatic stance, which Hale and Wright say the abstractionist can accommodate, and the ambitions of abstractionism as originally proposed. To see this, it will be useful to make use of the distinction drawn by Shapiro (2008) between the external and the internal perspective.

From the external perspective we can devise criteria which abstraction principles ought to satisfy. It has been suggested, for instance, that acceptable principles need to be *stable*, i.e. satisfiable in domains of all sufficiently large cardinalities.[3] But demonstrating that a particular abstraction principle *is* stable always seems to presuppose a background of standard set theory. This is a general problem for the strategy of hoping that when we have to choose which abstraction principles to adopt, technical results will do the triage for us. The proof that a particular principle satisfies the stability condition will typically, since it has to be approached model-theoretically, make use of a rich set theory in the metalanguage. What we were aiming for all along was an account of our grasp of the notion of set (or class), so appealing to a rich set theory in the metalanguage cannot do the trick.

So we fall back instead on the internal perspective. But from this standpoint it is hard to see why, if we take seriously the pragmatic stance Hale and Wright propose, the fact that our axiom has the syntactic form of an abstraction principle should any longer be of special significance. The abstractionist cannot, any more than anyone else, reasonably hope to dodge the demand for a conception to which their choice of foundational principle is answerable.

[3]See Weir (2008) for a technical discussion of stability.

The following quotation summarizes the point nicely. 'A major objective of the philosophy of set theory is ... to find a conception of set which gives us some reason for thinking that a corresponding set theory would be consistent.' The surprise is that this quotation is from ... the same paper by Hale and Wright (p. 193). Why, then, if they are willing to grant the need for a conception that guides our choice of principles, do they elsewhere see our discussion of such conceptions as evidence of extreme platonism?

The point Sullivan and I had been making in the article to which Hale and Wright were responding was that classes are, according to the iterative conception, an example of a range of abstract entities whose nature is not exhausted by their criterion of identity. (We cannot, in other words, deduce the iterative conception merely from the information that classes are extensional entities.) Our intention in making this point was to bolster the argument we had given in our previous article (1997) against what we there called the 'Lockean model' of reference. On the Lockean model, the job of a singular term is to point in a certain direction: it is then up to the world to decide whether there is anything there for the term to pick out. This model, whatever its virtues as an account of our reference to mountains and stars, makes little sense, in our view, when applied to mathematics. Hale and Wright's response to the Julius Caesar problem showed, we suggested, that they had not distanced themselves from this model sufficiently to leave room for them to be able to say of numbers—what they plainly would not say about mountains and stars—that they are *no more than* what our conception requires them to be. In our later paper, Sullivan and I used the iterative conception as an example to show that there may be abstract objects which are more than what our identity criterion requires them to be, and hence that our conception of them outruns the identity criterion, a possibility for which abstractionism leaves no room.

In opposing the Lockean model, Sullivan and I had rejected the idea that the principle by which mathematical terms are introduced 'stands exactly on a level with' other mathematical truths **?**. We had insisted that 'we must *reject* the idea that there might be facts about numbers which are, from the perspective of the concept we have of them, quite contingent.' (p. 149, my emphasis) What we need to recognize, if we wish to free ourselves entirely from the Lockean model, is that in mathematics, in contrast to the empirical case, there is no distance between the existence of the objects and the coherence of our conception of them. This is what Frege realized when he said, 'In arithmetic we are not concerned with objects which we come to know as something alien from without through the medium of the senses, but with objects given directly to our reason and, as its nearest kin, utterly transparent to it.' (1953, p. 115)

How much of this do Hale and Wright grant? Elsewhere in their article (p. 202), they insist that the abstractionist makes 'no attempt to create *objects* or stipulate their existence—what is created, if all goes well, is not objects, but *grasp of a concept*'. This is no doubt a sensible attitude to take to empirical concepts, but if we repudiate the Lockean model, it can scarcely be the right thing to say about mathematics, since it insists on the distance between concept and object which only the Lockean model makes sense of. Acknowledging the

need for a justifying conception is not, as Hale and Wright suggest, a symptom of extreme platonism. In fact, matters stand the other way about: to reject the need for such a conception, as they do on p. 195, only makes sense on the assumption of extreme platonism.

5 Abstraction and limitation of size

Now that I have said something about why abstractionists ought to choose their abstraction principles to conform to some conception of the objects these principles purport to introduce, and also indicated why the iterative conception cannot provide them with what they need, let us focus on the limitationist idea that what determines whether a concept is class-forming is how many objects fall under it. For this is, indeed, the leading conception among abstractionists.

It will be helpful to distinguish three versions of the limitation-of-size principle of increasing strength. We write V for the universal concept which holds of every object. We write $F \preccurlyeq G$ to mean that there is a one-to-one correspondence between the Fs and some of the Gs, and $F \sim G$ to mean that there is a one-to-one correspondence between the Fs and all of the Gs.

(Size) If $F \sim G$, then F is class-forming iff G is.

(Lim) If $F \preccurlyeq G$ and G is class-forming then F is class-forming.

(All) F is class-forming iff $F \not\sim \mathrm{V}$.

(We shall consider another principle of a somewhat different kind later.)

Limitation of size has one great advantage over the iterative conception for the abstractionist, namely that the constraint on which concepts are class-forming can be formulated without referring to the membership relation. It is therefore possible to formulate non-circular abstraction principles that correspond to various versions of the limitation of size idea. The one that has been most extensively investigated is known in the literature as (New V):

(New V) $\{x : Fx\} = \{x : Gx\}$ iff $\forall x\, Fx \equiv Gx$ or the Fs and the Gs are equinumerous with everything.

This evidently amounts in effect to adopting (All). But other principles can be formulated that enunciate various strengths of (Lim). E.g. the principles

(Finite) $\{x : Fx\} = \{x : Gx\}$ iff $\forall x\, Fx \equiv Gx$ or neither the Fs nor the Gs are finite in number

(Countable) $\{x : Fx\} = \{x : Gx\}$ iff $\forall x\, Fx \equiv Gx$ or the Fs and the Gs are uncountable in number

(Set) $\{x : Fx\} = \{x : Gx\}$ iff $\forall x\, Fx \equiv Gx$ or neither F nor G is equinumerous with a set.

capture, respectively, the ideas that all classes are finite, that all classes are countable, and that every class is equinumerous with some set. The first two of these principles can easily be formulated in pure second-order logic; the last is most naturally expressed by first adjoining the axioms of first-order ZF to the background theory *before* we contemplate what abstraction principles we want to add. If it is solely pragmatic considerations that drive us, then (Set) is *prima facie* attractive, since it is ample for embedding classical mathematics as normally practised—excluding set theory itself, of course. The attraction diminishes, though, when we realize that we can make sense of (Set) only if we have a functioning theory of sets in the background, in which case we were presumably already in a position to embed classical mathematics, without the aid of any abstraction principles at all. A variant which avoids this difficulty is

(Access) $\{x : Fx\} = \{x : Gx\}$ iff $\forall x\, Fx \equiv Gx$ or F and G have subconcepts equinumerous with a strongly inaccessible cardinal.

This, it turns out, can be expressed in pure second-order logic.

Does the abstractionist have any reason to think that any of the versions of the limitationist conception which I have just listed is true? (Or, more longwindedly, is there reason to think that any of these abstraction principles provides us with a possible route to a grasp of some range of objects?) Let us consider (Size) first. It is the weakest of the three principles, and it undoubtedly has a certain plausibility if one has a conception of sets as barely extensional entities, i.e. as entities that have no other structure than is required in order to satisfy the axiom of extensionality. This fits well with abstractionism, since it is the abstractionist's contention that objects obtained by means of an abstraction principle have no more to them than is delivered by that principle.

On the other hand, (Size) is not really an expression of the notion of limitation of size at all: it might be better described as the 'Only size matters' principle. It is in any case too weak on its own to give us a useful theory of classes. Indeed, it does not on its own give us the existence of any classes whatever. Plainly, we need more (at least if we are not to give up the aspiration to have a theory of classes that is strong enough to play a role as foundation for the rest of mathematics).

The second limitation of size principle (Lim) still does not give us any classes on its own: it defers the question of how big is too big. Despite its relative weakness, however, it is much harder to motivate than the first principle. One strategy for doing so ('divide and rule') would be to note that (Lim) is equivalent to the conjunction of (Size) with the following:

(Sep) Every subconcept of a class-forming concept is class-forming

(The notion of a subconcept, as I use the term here, is extensional: so F is a subconcept of G just in case every F is a G.)

Since we have already noted that the abstractionist has a plausible argument for (Size), we are free to focus on whether we can justify (Sep). This principle is certainly hallowed by the familiarity of long usage, being what is generally

known as the axiom of separation. But familiarity is not in itself sufficient to make something true, and the abstractionist owes us an argument for believing (Sep). It is hard to see what this argument might be. At any rate, the most obvious argument for (Sep) is that the subconcepts of a class-forming concept occur earlier in the class-forming process than it and hence are class-forming. But this argument makes direct appeal to the *iterative* conception, which, as we have seen, is not available to the abstractionist.

Suppose, however, that we could somehow motivate (Lim). Note that in the presence of (Lim) it is evidently inconsistent to assume that there is a universal class (else the Russell class would exist too). So, on pain of inconsistency, certainly no more classes could exist than the ones given by (All). So to motivate acceptance of (All), it seems that we need to invoke a maximality principle: whatever classes can exist (as far as consistency considerations are concerned) do exist. But it is far from clear that there is any intelligible reason the neo-Fregean could invoke for accepting that principle here. Why accept (All) rather than, say (Finite) or (Countable)? It is far from clear that the modality the neo-Fregean invokes here ('what can exist as far as consistency considerations are concerned') is really intelligible in this context, since the objects whose existence is in question are, as far as logic is concerned, necessarily existent. In this respect the neo-Fregean seems to be in a worse position than the iterativist, since the iterativist's conception of the metaphysical nature of sets is richer than the neo-Fregean's, and hence might more plausibly be one to which a non-logical modality would be applicable.

To sum up, the position seems to be that (Size) has some prospects for an abstractionist motivation, but that (Lim), and *a fortiori* (All), do not, unless it can be explained why one might think that (Sep) is true.

6 Abstraction and the definite conception

I want to consider now a rather different principle for identifying the class-forming concepts, one that has a *prima facie* motivation as an explanation for the paradoxes of classes. This is the idea that the class-forming concepts are just those that are not indefinitely extensible. Let us call this (for want of a better title) the *definite* conception of set:

(Def) A concept F is class-forming iff it is not indefinitely extensible.

The definite conception goes back to Russell (1906), who uses the term 'self-reproductive process' for what I am here calling indefinitely extensible concepts. 'It is natural to suppose,' Russell says (p. 152), 'that the terms generated by such a process do not form a class.'

Some abstractionists (e.g. Shapiro and Wright 2006) have recently suggested basing neo-Fregean logicism on the definite conception. Are they right to do so? Does it, that is to say, motivate a suitable abstraction principle for classes? It is easy to see why one might think that it does. For, as Russell noted, all the paradoxes may be treated as demonstrations that certain concepts are

indefinitely extensible. Nothing more tempting, then, than to think that the resolution of the paradoxes is to ban indefinitely extensible concepts from having extensions.

Tempting it may be, but in fact it is thoroughly confused. To see why, we need to consider a little more closely what is meant by indefinite extensibility. In one place Wright (2008, p. 25) calls this notion 'tantalising'. Why? What is wrong with defining it as follows?

> A concept F is *class-indefinitely-extensible* if there is a function f such that for every class A of Fs, $f(A)$ is an F not belonging to A.

The answer is that, although there is nothing wrong with this on its own, it will not do for the abstractionist's purposes, because it makes use of the notion of class (which is why I have labelled it with the prefix 'class'). If we were to adopt the principle of class abstraction with 'class-forming' interpreted to mean non-class-indefinitely-extensible as just defined, we would have fallen foul of just the circularity problem that we discussed earlier in connection with the iterative conception. So if our characterization of indefinite extensibility is to do the work the abstractionist needs it to do, it must be formulated independently of the notion of class. The trouble is, though, that if it *is* formulated independently, our reason to be suspicious of the extensions of indefinitely extensible concepts disappears.

More generally, the point here is that the notion of indefinite extensibility is relative to a notion of totality. Suppose we have a notion of totality, call it Π-totality. Then (following Shapiro and Wright 2006) we can say that a concept F is *indefinitely extensible relative to* Π if there is a function f such that for every Π-totality A of Fs, $f(A)$ is an F not belonging to A. But the abstractionist is then faced with a dilemma: if the notion of totality he appeals to is the notion of class he is trying to access, he violates the circularity constraint; if it is not, then he is left unable to explain what is wrong with thinking that the indefinitely extensible concepts are class-forming.

To make the point vivid, consider as an example the case where Π is the iterative concept of set (which is, by a familiar argument from Russell's paradox, indefinitely extensible relative to itself). Our prior theory will then presumably contain an axiomatization of this conception, such as ZF or ZFC. If we then add to this theory a new notion of class, introduced by means of an abstraction principle, what reason do we have to refuse to form a class of all sets? Not worries about consistency, certainly. But what other problem with this class is there supposed to be? To think that there is something problematic about the class of all sets (where *class* and *set* are distinct notions) is surely to misunderstand the paradoxes completely.

7 From the definite conception to ordinal limitation

In this section I want to consider an argument, hinted at by Russell (1906) and filled out by Shapiro and Wright (2006), which purports to show that (Def) entails:

(Ord) F is class-forming iff Ord $\not\preccurlyeq F$,

where 'Ord' refers to the concept *ordinal*. The significance of this argument is that (Ord) may be thought of as a kind of limitation-of-size principle. So the argument, if valid, does provide us after all with a route from the definite conception to a kind of limitation-of-size theory.

The argument in question is the following.

- Suppose first that Ord $\preccurlyeq F$. Then F has a subconcept G which is equinumerous with the ordinals. Since the concept Ord is indefinitely extensible, so is G. And since G is indefinitely extensible, so is F.

- Suppose conversely that F is an indefinitely extensible concept. So there is a function f which takes each class A of Fs to an object $f(A)$ which does not belong to A but satisfies F. Suppose that we define a function g on the ordinals by transfinite recursion as follows:

  $$g(\alpha) = f(\{g(\beta) : \beta < \alpha\}).$$

  Then it is easy to see by transfinite induction that all the $g(\alpha)$ are Fs. So g is a one-to-one function from the ordinals to the Fs. Thus Ord $\preccurlyeq F$.

Notice, however, that this proof makes use of two further principles. The first is that if G is a subconcept of F and G is indefinitely extensible, then F is indefinitely extensible. The second is that at each stage of the recursion the $g(\beta)$ for $\beta < \alpha$ form a class. Let us consider these two principles in turn.

What reason do we have to think, first, that every superconcept of an indefinitely extensible concept is indefinitely extensible? Or, equivalently, what reason do we have to think that if a concept is not indefinitely extensible, none of its subconcepts is indefinitely extensible? That latter principle is an analogue of (Lim) and suffers from a similar lack of obvious motivation. Indeed we might wonder whether the concept *class* itself provides a counterexample. For while various subconcepts picking out certain kinds of class might be indefinitely extensible, it is not as obvious as many authors seem to think that the concept *class* is indefinitely extensible relative to itself. There are two kinds of proof that have been offered of this. The first is as follows.

> For each class A of classes let $f(A) = A$. Now $A \notin A$, and so $f(A)$ is a class not belonging to A, as required.

But this proof makes use of the principle that $A \notin A$, i.e. that classes do not belong to themselves. Why believe that? The iterative conception offers a reason—no class can depend on itself—but in the present context this notion of dependency is not available. Quine's NF allows self-membered classes in abundance.

The other standard proof of the indefinite extensibility of the concept *class* goes as follows.

For each class A of classes, let $f(A) = \{x \in A : x \notin x\}$. If $f(A) \in A$, then $f(A) \in f(A) \Leftrightarrow f(A) \notin f(A)$, which is absurd; so $f(A) \notin A$, i.e. $f(A)$ is a class not belonging to A, as required.

This proof makes no use of the irreflexivity of membership, but it does make use of separation in order to guarantee that $f(A)$ is a class. Once again, as we noted earlier, this is obvious on the iterative conception, but it is far from clear why it should hold for the non-iterativist.

The second principle the proof implicitly assumes is that the objects $g(\beta)$ for $\beta < \alpha$ form a class. Why should they? Of course, it follows at once from the axiom of replacement that they do; but as this axiom is generally presented as a form of limitation of size principle,[4] appealing to it here simply takes us round in a circle.

8 Abstractionism and ordinal limitation

In the last section we considered an argument intended to show that the definite conception entails (Ord). This principle can be thought of as another variant of the limitation of size notion that we considered earlier. Indeed it is not hard to show that it is intermediate in strength between the strong principle (All) and the weaker (Lim).

> Suppose that (Ord) holds but that (Lim) does not. There are therefore concepts F and G such that $F \preccurlyeq G$ and G is class-forming but F is not. So Ord $\preccurlyeq F$ by (Ord). So Ord $\preccurlyeq G$ by transitivity. Hence G is not class-forming. Contradiction. Hence (Ord) entails (Lim).
>
> Suppose that (All) holds but (Ord) does not. There are two possibilities.
>
> 1. There is a non-class-forming concept F such that Ord$\not\preccurlyeq F$. But $F \sim V$ by (All), and so Ord$\not\preccurlyeq$V, which is absurd.
> 2. There is a class-forming concept F such that Ord$\preccurlyeq F$. Then $F \not\sim V$, so Ord$\not\sim V$, and hence Ord is class-forming. From this we can generate the Burali-Forti paradox, provided that the notion of ordinal in play here is one for which any class of ordinals has an ordinal. (For more on this assumption, see below.)
>
> Thus (All) entails (Ord).

The interest of (Ord) lies in the fact that, as we have seen, there is an argument intended to show that it follows from the definite conception. This argument is problematic, as we have seen, but let us put those concerns to one side. If we *could* establish (Ord) on the basis of the definite conception, would that help the abstractionist? On the face of things, it would. The difficulty with the definite conception for the abstractionist, let us recall, was that it comes

[4] See, for instance, Boolos (1971).

in two varieties: on the first variety, where 'definite' is interpreted relative to the notion of class we are trying to define, it is circular; on the second, where it is interpreted relative to some other notion (such as that of set), it is wholly unmotivated. Our new principle (Ord), on the other hand, offers the hope that it might lead us to a non-circular abstraction principle for classes in which the class-forming concepts are precisely those into which the ordinals cannot be injected.

But a moment's thought shows that this hope is illusory: all that has happened is the ambiguity in the notion of definiteness has been transferred to the notion of an ordinal. There are, just as before, two possibilities. The first is that it is built into our notion of ordinal that any class of ordinals has an ordinal: call that the class-notion. If that is our notion, then specifying the notion involves the notion of class, and hence we cannot use it in a non-circular abstraction principle. The second possibility is that our notion of ordinal can be specified independently of the notion of class, and can be used in a non-circular abstraction principle. But now there is no doubt that the ordinals form a class, in which case they are an immediate counterexample to (Ord).

9 Conclusion

In this article we have considered three possible motivations for an abstractionist account of the theory of classes, iterative, limitationist and definite. The first and last of these fail because all the ways of articulating them that have any prospect of explaining the paradoxes mention the membership relation, and are therefore illegitimate because circular. This leaves limitation of size as the abstractionist's only hope. Of the limitation principles we have discussed, (Ord) stands out as having a different character from the others because it is either circular (if the notion of ordinal in question involved is one that implicates classes) or wholly unmotivated (if not). Let us call the other limitation of size principles *pure*. In pragmatic terms, some of them are certainly capable of delivering non-circular abstraction principles strong enough to embed classical mathematics, and some—such as (Set) or (Access)—can be shown (if we assume a reasonably strong set theory in the metalanguage) to obey an appropriate conservativeness condition. However pragmatically successful these principles may be, though, they have so far stubbornly resisted all attempts to make them seem at all plausible. It is very hard to see why an abstractionist might think that the simple fact of there being a lot of Fs should prevent F from being class-forming.

Our discussion of the failings of the iterative and definite conceptions now provides a clue to the reason for this resistance. If all we want is to *avoid* the paradoxes, then the solution is simple: we define 'class-forming' in such a way that there are no more class-forming concepts than there are objects. Various pure limitation-of-size principles are available that achieve this. But this is too weak a constraint: with a little ingenuity we could plainly formulate all sorts of other, non-limitationist principles that meet it too. What we want is a

motivation for our chosen abstraction principle that simultaneously *explains* the paradoxes about classes. We want the consistency of the principle not to seem like a brute fact about it, explained only by a consistency proof that appeals to an even stronger background set theory in the metalanguage. But if we want to explain the class paradoxes, and not just avoid them, it seems inevitable that we shall have to use the membership relation in our explanation, which is precisely what the abstractionist, on pain of circularity, cannot do.[5]

References

Boolos, G. (1971), 'The iterative conception of set', *J. Phil.,* 68: 215–231

——— (1989), 'Iteration again', *Phil. Topics,* 17: 5–21

——— (1998), *Logic, Logic, and Logic*, Cambridge, MA: Harvard University Press

Cook, R. T. (2008*a*), *The Arché Papers on the Mathematics of Abstraction*, London: Springer

——— (2008*b*), 'Iteration one more time', in *The Arché Papers on the Mathematics of Abstraction* (Cook 2008a), pp. 421–54

Frege, G. (1953), *The Foundations of Arithmetic*, Oxford: Blackwell

——— (1980), *Philosophical and Mathematical Correspondence*, Oxford: Blackwell

——— (1984), *Collected Papers*, Oxford: Blackwell

Hale, B. and Wright, C. (2008), 'Abstraction and additional nature', *Phil. Math. (III),* 16(2): 182–208

Jané, I. and Uzquiano, G. (2004), 'Well- and non-well-founded Fregean extensions', *J. Phil. Logic,* 33(5): 437–65

Maddy, P. (1988), 'Believing the axioms I', *J. Symb. Logic,* 53: 481–511

Potter, M. and Smiley, T. (2001), 'Abstraction by recarving', *Proc. Arist. Soc.,* pp. 327–38

——— (2002), 'Recarving content: Hale's final proposal', *Proc. Arist. Soc.,* 102: 351–354

[5]I am grateful to Alex Oliver, Luca Incurvati, Roy Cook, Peter Smith and Peter Sullivan for comments on earlier drafts of this article.

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

——— (2005), 'What is wrong with abstraction?', *Phil. Math. (III),* 13(2): 187–93

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

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

——— (1973*b*), 'The regressive method of discovering the premises of mathematics', in *Essays in Analysis* (Russell 1973*a*), pp. 272–83 (written in 1907)

Shapiro, S. (2008), *Frege meets Dedekind: A Neo-Logicist treatment of real analysis*, [n. pub.], pp. 219–52

Shapiro, S. and Wright, C. (2006), 'All things indefinitely extensible', in *Absolute Generality*, Oxford University Press, pp. 255–304

Weir, A. (2008), 'Neo-Fregeanism: An embarrassment of riches', in *The Arché Papers on the Mathematics of Abstraction* (Cook 2008*a*), pp. 383–420

Wright, C. (2008), 'Is Hume's principle analytic?', in *The Arché Papers on the Mathematics of Abstraction* (Cook 2008*a*), pp. 17–43