Published in Travaux de logique 18, 103-129, 2007
which should be used for any reference to this work

Kit Fine on *The Limits of Abstraction*

Bob Hale

Kit Fine's book is a study of abstraction in a quite precise sense which derives from Frege. In his *Grundlagen*[1], Frege contemplates defining the concept of *number* by means of what has come to be called Hume's Principle – the principle that the number of *F*s is the same as the number of *G*s just in case there is a one-to-one correspondence between the *F*s and the *G*s. Frege's discussion is largely conducted in terms of another, similar but in some respects simpler, proposal – to define the concept of *direction* by laying down the principle (the Direction Equivalence) that the direction of a line *a* is the same as that of line *b* if and only if lines *a* and *b* are parallel. Such principles have come to be known as abstraction principles. More generally, abstraction principles are ones of the shape:

$$\forall\alpha\forall\beta\,(\S(\alpha) = \S(\beta) \leftrightarrow \alpha \approx \beta)$$

where $\approx$ is an equivalence relation on entities of the type over which α and β vary and, if the principle is acceptable, § is a function from entities of that type to objects. Since 'the direction of' is intended to stand for a function from *objects* of a certain sort (viz. straight lines) to other objects (their directions), the Direction Equivalence is a first-order, or in Fine's terminology *objectual*, abstraction; since 'the number of' is intended to stand for a function from *concepts* (or properties) to objects, Hume's Principle is a higher, or as Fine says *conceptual*, abstraction. As is well-known, Frege himself abandoned the idea of defining number implicitly or contextually by means of

[1] Frege (1884). The relevant sections are §§ 62-67.

Hume's Principle – adopting instead an explicit definition of the number of *F*s as the extension of the concept *concept equinumerous to the concept F* – because he thought that an adequate definition should settle the question whether Julius Caesar, for example, is or is not the number of some concept, but that the proposed implicit definition cannot do so. But interest in abstraction principles has revived in the last couple of decades, largely as a result of Crispin Wright's attempt[2] to show that in spite of the many difficulties confronting it – including the notorious Julius Caesar problem – Hume's Principle can after all serve as a foundation for arithmetic.

Although Kit Fine allows that a theory of abstraction can be developed in which some fundamental mathematical theories may be reconstructed, the overriding message of his book, as its title suggests, is largely negative. In particular, he is quite out of sympathy with the idea that abstraction principles (e.g. Hume's Principle) might play any significant rôle in providing philosophical foundations for mathematics – even for those parts of it which can be obtained within the formal theory of abstraction developed in parts III and IV of his book. I suspect that, given the assumptions underlying and shaping Fine's approach to the subject, this negative assessment is more or less inevitable. Here, I want to explore what I believe to be some of the more important of those assumptions, and suggest some reasons why we might not find them irresistable.

We can begin with a point of agreement: if abstraction principles are to play a foundational rôle, it must be possible to view suitable such principles as serving somehow to *explain* or *define* fundamental mathematical concepts, such as cardinal number and real number.

What, at bottom, we disagree about is whether they can be so viewed. Fine believes not, and most of parts I and II of *Limits* is devoted to arguing his case. Since he denies that abstraction principles can serve in a definitional capacity, he does not ask: Which abstraction principles are *acceptable as definitions*? Instead, his

[2] Wright (1983). The possibility of basing arithmetic on Hume's Principle was first noted, I believe, by Charles Parsons (1965). A good deal of the subsequent work in the neo-Fregean programme Wright initiated is collected in Hale & Wright (2001).

questions are: Under what conditions is an abstraction principle *true*? Which are the true abstraction principles?

I believe that Fine's negative assessment of the prospects for using abstraction principles in a definitional, and thence foundational, rôle has its source, ultimately, in his adoption of what could be called an *externalist*[3] standpoint with regard to abstraction. Centrally, this involves thinking that what objects there are is largely independent of abstraction. The acceptability of abstraction principles is then a matter of their being true of a certain domain – the universe of all objects whatever – of (largely) independently constituted composition and cardinality. Under a suitable assumption about the size of this all-inclusive domain of objects – that the cardinality of the domain is unsurpassable[4] – the acceptable abstraction principles can be characterised as those which are non-inflationary[5] and predominantly logical[6]. But abstraction principles themselves cannot be used to give us any assurance of the truth of such an assumption, which must be justified independently, if it can be justified at all.

Clearly if Fine's claim that abstraction principles can play no useful definitional rôle is correct, then such principles can enjoy no philosophical advantage and can bring no epistemological or ontological gains. Since that runs directly counter to what neo-Fregeans have claimed for them, I will concentrate on Fine's critique of abstraction principles as definitions. After I have explained where and why I differ from him on this issue, I will go on to venture some more

[3] Fine uses the term in a somewhat different sense – in which a characterization of a position (e.g. that of the uncompromising abstractionist) is externalist if it can only be regarded as correct, or even as intelligible, by one who does not hold the position. Note that rejecting externalism in my sense does not require adoption of uncompromising abstractionism – one can allow that there are abstract objects which cannot be seen as introduced by abstraction, so long as there aren't too many of them – so many that those which can be so viewed constitute only an insignificant proportion.

[4] A cardinal $\mathbf{c}$ is unsurpassable if $2^{\mathbf{d}} \leq \mathbf{c}$, where $\mathbf{d}$ is the number of cardinals less than $\mathbf{c}$. See Fine p. 7.

[5] That is, where $\approx$ is the abstraction's equivalence relation, the number of $\approx$-equivalence classes must not exceed the number of objects. See Fine p. 4.

[6] That is, the identity criterion given by the abstraction's equivalence relation 'involves only an 'exponentially small' number of objects in relation to the number of objects in the universe as a whole' See Fine p. 7, where the notion of exponential smallness is further explained.

speculative and tentative thoughts about the influence of the externalist perspective on his thinking about abstraction.

Fine contrasts what he calls 'orthodox' or 'classical' definitions with 'unorthodox' or 'non-classical' ones. I am not sure how he means to draw this distinction in general, but in the case of definitions which he takes to be intended to result in an assignment of objects to the defined terms, the idea seems to be that orthodox definitions presuppose the existence of the objects to be assigned as referents, whereas unorthodox definitions do not.

Prominent among the *unorthodox* kinds of definition Fine discusses are what he terms 'definition by reconceptualization' and contextual or implicit definitions of the sort he calls 'creative'. Crispin Wright and I hold that Hume's Principle can be used to implicitly define the number operator, and claim that the idea of reconceptualization can play a key rôle in explaining how we may acquire warranted beliefs about abstract objects. Since we could scarcely maintain the latter claim, if we agreed that the use of Hume's Principle as a definition presupposes the existence of the objects which are to be the values of the number operation, we cannot view the proposed definition as orthodox (at least, not in Fine's sense). So it may seem that our disagreement with Fine must largely concern whether it can function as a definition by reconceptualization or as a creative implicit definition. As we shall soon see, matters are not quite this straightforward. But it is certainly true that our main[7] points of disagreement with him concern what we may as well agree to call unorthodox definition, and I shall concentrate on these.

[7] I have some disagreements with Fine about definition in general, and in particular with his claim that definitions always presuppose a domain of discourse. Although they are, in my view, important, I don't have space to go into them here.

Reconceptualization and implicit definition

(i) Reconceptualization

Fine discusses two 'non-classical' ('unorthodox') approaches to defining number by means of Hume's Principle. In the first of them, the leading idea is that the definition works by reconceptualization; in the second, that it involves an application of the context principle. Although both ideas derive from Frege, Fine is not concerned with faithful interpretation of Frege's thought – as he says, his interest is in 'whether the ideas themselves can be sustained'. His own view is that they cannot. While I do, and shall, disagree with him about that, I share the same primary focus of interest – in the ideas rather than Fregean exegesis. So I am especially interested in whether this broad agreement between us extends to agreement over what 'the ideas themselves' *are* which may or may not be sustainable. I regret that it seems to me not extend very far at all, with the unfortunate result that Fine largely fails to engage with the ideas in what I believe to be their best shape.

Fine treats reconceptualization and contextual definition as two alternative and entirely separate approaches. He does so because he takes 'definition by reconceptualization' (his term, not mine) to rest 'on the idea that new senses may emerge from the reanalysis of a given *sense*', whereas 'the context principle ... rests upon the idea that certain truths may be used to fix the *reference* of the terms that they contain. Thus whereas the first-mentioned approach rests on the adoption of an unorthodox mechanism for the *determination of sense*, the second rests on an unorthodox mechanism for the *determination of reference*' (p. 35). Definition of the number operator by reconceptualization is supposed to proceed in two stages – 'first, the formula on the left is taken to have the same sense as the formula on the right; second, this sense is subject to reanalysis in such a way as to provide a sense for the number operator' (p. 37). The formulae on left and right should be taken, Fine holds, to be open sentences – thus in the case of Hume's Principle, they will be 'the number of Fs = the number of Gs' and 'the Fs are equinumerous with the Gs', where F and G are concept

variables. The process, he suggests, can be seen as combining the methods of explicit and implicit definition. I think he means that the left hand formula is *explicitly* defined to have *the same sense* as the right hand formula, and the number operator is then *implicitly* defined as having the sense it must have, if the left and right hand formulae are to be synonymous. What he says is that 'The idea behind reconceptualization is to get the implicitly defined sense, not directly from the principle, but indirectly from the explicitly defined sense' (p. 38). It is anyway clear enough that, on the view he is discussing, reconceptualization involves treating the left and right hand formulae having the same sense and taking this to result, somehow, in a unique sense for the operator embedded in the left hand formula.

Fine does not deny that definition by reconceptualization – as *he* understands it – can ever succeed in determining a unique sense. He allows that examples such as one given in Frege's *Begriffsschrift* – whereby the one-place predicate 'commits suicide' is defined from the two-place predicate 'kills' by, in effect, identifying the argument-places – might be seen as successful applications of reconceptualization. But he denies that one can accept Hume's Principle, or any other abstraction principle, as a definition by reconceptualization, on the ground that doing so would force us to claim sameness of sense in cases where it is clearly incorrect to do so. Thus in his main example, we are to consider the statements:

(A) $\forall x(Fx \leftrightarrow Gx)$

(B) $\S xFx = \S xGx$

(C) $\forall x(x=\S xFx \leftrightarrow x=\S xGx)$

(D) $\S x(x=\S xFx) = \S x(x=\S xGx)$

where § is the class abstraction operator, and we are to prescind from the inconsistency of Frege's Basic Law V. Fine argues that:

> ... (B) should have the same sense as (A) and (D) the same sense as (C). Now presumably (C) should also have the same sense as (B). Indeed, it is hard to see how a criterion of synonymy could be so tolerant as to allow the synonymy of (B) to (A) (or of (D) to (C)) and yet not tolerant enough to allow the synonymy of (C) to (B). If this is

> right, it follows that (A) has the same sense as (D). But then if the sense of the operator given through reconceptualization is unique, the sense of §xFx will be the same as the sense of §$x(x$=§$xFx)$, which is presumably not so (p. 40 – with minor correction).

It is, he concludes, 'difficult to see what general principles governing identity of sense might permit its reconceptualization and yet prevent its proliferation'.

I think Fine is right to resist any attempt to assimilate definition by means of abstraction principles to the kind of example given in the *Begriffsschrift*, but my reason for so thinking is probably not one on which we agree. Frege's claim there is that, in many cases, one and the same sentence can be analysed into function and argument(s) in different ways. He emphasizes that this does not affect the 'conceptual content', which remains the same under different decompositions of the sentence into function and argument(s). In such cases, it is then very natural to think of a single complex sense, expressed by the whole sentence, as analysable into constituent senses in different ways. In particular, one can then plausibly view the concept of committing suicide as obtained by 'reanalysing' the complex sense expressed by, say, 'Cato killed Cato', by taking 'Cato' as a single argument replaceable at both occurrences. But it is essential to the plausibility of this view that we have a *single* (unambiguous) sentence analysable in different ways. This is precisely what we don't have, in the case in which Frege speaks of carving up the content in a new way – in the passage in *Grundlagen* which has inspired talk of reconceptualization. For precisely this reason, I do not think we can view Frege's discussion in *Begriffsschrift* as supplying the model for his later talk of carving up content in new way. I believe we should think of reconceptualization – not as a process of supplying a new analysis of some single given sense – but in a quite different way. I shall say more of this later. Fine, however, makes no distinction between reconceptualization and reanalysis, which he construes in a way that suits the *Begriffsschrift* example – i.e. as a process by which a single sense is subjected to different decompositions. It is then

hardly surprising that he finds that abstraction principles cannot be viewed as involving reconceptualization.

I think Fine is probably right, too, to be sceptical about the possibility of framing general criteria for identity of sense which would allow us to treat the left and right sides of Hume's Principle, for example, as having the same sense, without running into the problem of proliferation. So if it were the case that using Hume's Principle (or any other abstraction) as a definition required us to hold that its left hand side re-expresses the same sense as its right – as it would if it was intended to function as a definition by reconceptualization in Fine's sense – this would constitute a serious problem for us [neo-Fregeans]. But it seems to me – now, anyway[8] – that it doesn't. Neither the claim that Hume's Principle can function as an implicit definition of the number operator, nor the claim that, when the number operator is so defined, statements of numerical identity can be taken as reconceptualizing the same states of affairs as corresponding statements about one-one correlation between concepts – neither claim need involve taking its left and right sides as having the same sense.

In contrast with Fine, I do not view ideas about reconceptualization and the context principle as belonging to two entirely distinct approaches to definition. I do not – and never did – regard reconceptualization as a *type of definition* or mechanism (unorthodox or otherwise) for determining senses. When we use an abstraction principle to fix or explain the sense of the embedded abstraction operator, what we are giving is a kind of implicit definition[9]. In general, an implicit definition proceeds by stipulating the truth of a suitable sentence incorporating the definiendum, with the intention

[8] Some of my own earlier published work on this issue (especially Hale 1997) suggests that there needs to be a relation of sameness of (weak) sense between the left and right sides of an abstraction principle. Various authors besides Fine, including Michael Potter and Timothy Smiley, and William Stirton, have pointed out serious difficulties in the way of characterizing a suitable relation of sameness of sense. In the text below I explain why I think any such requirement is misguided.

[9] Thus for us, implicit or contextual definition is primarily a means of fixing *sense*, not – as Fine supposes – a mechanism for determining *reference*. So we are in sharp disagreement both with his view of reconceptualization and with his view of contextual definition.

that it is to mean precisely what it needs to mean – given the already fixed meanings of the expressions composing the sentential context in which the definiendum is embedded – for that sentence to express a truth. It is thus an implicitly definitional stipulation of the truth of a sentence which is the 'mechanism' by which, if all goes well, the sense of the definiendum is determined. That the LHS of an abstraction principle (more precisely, instances of the LHS) may then be regarded as reconceptualizing the states of affairs describable by (corresponding instances of) its RHS is *not* the *means* whereby the sense of the abstraction operator is fixed – it is rather a *result* of the operator's being defined in this particular way.

Let me say a little by way of explanation and defence of these last two claims, starting with my claim about implicit definition. In an *explicit* definition, there clearly *is* a claim or stipulation[10] about identity of sense – what is claimed or stipulated is that the definiendum has (or is to have) the same sense as the definiens. It is, in general, obvious that implicit definition by contrast involves *no* claim or stipulation about identity of sense – the definiendum is normally a single word or phrase, whereas the definiens is a complete sentence containing it, and there is no suggestion that part of the sentence has the same sense as the whole! The idea rather is – at least on the account of the matter I favour – that by embedding the definiendum in a suitable sentence, otherwise comprising only expressions whose meanings are already fixed, we can, by stipulating that the sentence is to be true, confer a meaning on the definiendum; it is to mean just what it needs to mean – no more, no less – for the definiens sentence to be true. Of course, there are important constraints on the enterprise – conditions for its success – but these need not concern us just now. The case is no different when the vehicle of implicit definition is a biconditional or perhaps a generalized biconditional, as in the case of an abstraction principle. There is still no claim or stipulation of sameness of sense. In particular, it is not being claimed or stipulated that the left and right sides of

[10] A claim if the definition aims to capture pre-existing usage, a stipulation otherwise.

(instances of) the biconditional are synoynmous. The stipulation is only that the (generalized) biconditional is to be true. If the definition is successful, the sense of the definiendum is fixed as what it needs to be, for that (generalized) biconditional to be *true*, and hence for (instances of) its left and right sides to have the *same truth-value* – it is *not* fixed as what it needs to be for them to have the *same sense*[11].

Turning now to the second claim, about reconceptualization, there are three points I want to make.

(i) The first simply expands on, and I hope explains, my claim that reconceptualization is not the *means* by which sense – the sense of the number operator, say – is determined, but is rather the *result* of its being determined in a quite different way. If, like Fine, we understand the claim that one sentence – e.g. 'the number of knives = the number of forks' – reconceptualizes what is expressed by another – e.g. 'the knives correspond one-one with the forks' – as requiring the two sentences to be synonymous, then my contrast between reconceptualization as a means of fixing sense and reconceptualization as a result of sense being fixed in some other way will be beside the point, as far as his objection (the problem of proliferation) is concerned. It wouldn't matter whether the claim of synonymy is the means whereby the sense of the number operator, say, is fixed or rather a result of its being fixed in some other way – for it is the claim itself that is objectionable. But as I have explained, there is no need to view the process of implicitly defining the number operator by means of Hume's Principle as involving any claim or stipulation about the synonymy of instances

[11] A possible source of confusion on this point lies in the fact – on which Fine remarks – that an explicit definition can normally be recast as a (generalized) biconditional. For example, instead of giving the definition " 'vixen' means 'female fox' ", we may define: $\forall x$ (x is a vixen $\leftrightarrow$ x is a female fox). Fine suggests that for this reason we can always treat explicit definition as a special case of implicit definition. In one way, this may be right – one can cast an explicit definition in the form of a stipulation of the truth of a sentence expressing a condition involving the definiendum – but in another, more important way, I think it is wrong or at least very misleading, because it obscures a fundamental difference between definitions of the two types, only one of which presents two expressions as having the same sense. Even if it is agreed that explicit definitions can be given as generalized biconditionals, in which the left and right sides have the same sense, it would of course be a gross error to infer that *any* definition which proceeds by stipulating the truth of a biconditional presents its left and right sides as synonymous. I am not, of course, suggesting that Fine makes any such inference.

of its left and right sides. And once that is accepted, there is every point in my contrast.

(ii) The second concerns the *object* of reconceptualization. As we have seen, Fine takes the objects of reconceptualization to be *senses*. By contrast, I think of conceptualizing as bringing *what we talk and think about* under concepts. Senses are not – or not usually, anyway – what we talk and think about, but are rather the means by which we talk and think about other things. Accordingly, I think of reconceptualization as bringing what we are already accustomed to talk and think about in one way under (new) concepts in a new way. Thus what gets reconceptualized, when we introduce the concept of direction by means of the relevant abstraction principle, is a certain type of states of affairs – the state of affairs consisting in line *a* being parallel to line *b* is reconceptualized or redescribed in terms of identity, or thought of instead as consisting in the direction of line *a* being identical to the direction of line *b*. And precisely because reconceptualization involves the introduction and deployment of new concepts, there can be no question of the sentences which articulate the reconception have the same sense as those which conceptualize things in the old way.

(iii) The third concerns (what can be said about) the *conditions* under which it is appropriate to think of one sentence as reconceptualizing what is conceptualized, in a different way, by another.

In my view, there is no need for a *separate* account of the conditions under which one sentence can be regarded as reconceptualizing the type of state of affairs depicted by another – i.e. in addition to an account of the constraints on implicit definition in general. Whenever the vehicle of implicit definition is a biconditional (or universal closure of such), then – provided the definition conforms to the appropriate general constraints – one can regard the left hand side as reconceptualizing the state of affairs depicted by the right hand side.

It is a nice question whether this condition is ever satisfied in any cases other than ones which are covered by the obvious generalization of what Frege says in *Grundlagen* §64. When Frege himself presents his example, what he claims is that we carve up the content of the

symbol for parallelism in a new way – 'we replace the symbol // by the more generic symbol =, distributing the special content of the former between *a* and *b*'. What is especially interesting here is that it is the content of the *symbol* for parallelism which Frege thinks of as being carved up in a new way – i.e. as opposed to the content of the whole sentence. I now think this may be more important than I formerly supposed[12], and that it is primarily at the sub-sentential level of relational expressions, rather than with complete sentences, that recarving or reconceptualization should be viewed as taking place. Clearly the salient feature of Frege's example is that we have an equivalence relation between entities of a certain sort. When *x* and *y* stand in a certain equivalence relation, they are the same in a certain respect. If sameness in that respect is of interest or importance to us, we may introduce a word (an abstract noun) for the respect in question – we can then say that *x* and *y* are the same in length, shape, volume, etc., or that they have the same length, etc. The state of affairs consisting in an equivalence relation's holding between *x* and *y* is redescribed – or reconceived – as an identity. So a sufficient condition for a kind of state of affairs depicted by sentences of a certain type, S, to be reconceptualizable using sentences of a different type, R, is that R-sentences assert identities, S-sentences assert that some other equivalence to hold, and R-identities are true just in case suitably related S-equivalences hold. Perhaps there are other sufficient conditions, but perhaps there aren't – I don't see that there has to be a more general story about when one can think of the same 'content' as 'carved' in a new way.

(ii) Implicit definition

As we've seen, Fine contrasts *traditional* implicit definitions with *creative* ones, as he calls them. Where what are defined are terms for objects, traditional implicit definitions, he holds (p. 56), presuppose the existence (and uniqueness) of objects for the newly defined terms to refer to, and we are therefore obliged, in giving such a definition, to

[12] in, for example, Hale (1997).

justify these existential presuppositions. So one can get no epistemological advantage from implicit definitions traditionally understood. By contrast, 'creative' implicit definitions are supposed not to presuppose the existence of the relevant objects:

> The purpose of the definition may indeed be to assign objects to the terms. But these objects are not selected from a previously given domain. Rather the objects are introduced into the discourse simultaneously with their assignment to the terms. (*ibid.*)

One might now suppose – as I think Fine does – that a neo-Fregean must take abstraction principles to function as creative implicit definitions, rather than orthodox or traditional ones. But this is not so. It is of course true that the neo-Fregean cannot agree that abstraction principles presuppose the existence of objects to be the referents of terms formed with the abstraction operator, and so cannot regard them as *traditional* implicit definitions *as Fine characterizes them* – i.e. as carrying such a presupposition. But it would be a gross error to infer from this that the neo-Fregean must take abstraction principles to be creative definitions in any clear or significant sense.

It is in fact quite unclear just what creative definitions, as Fine describes them, are supposed to be. One might think that, if they are to avoid presupposing the existence of the objects which are to be assigned to the defined terms, these objects must be supposed to be somehow produced by the definition. But Fine denies this: the relevant objects are not to be thought of as literally brought into existence by the act of definition – rather, he tells us, the objects are to be viewed as existing 'independently of our practice of referring to them; but it will help to create or constitute our reference to those items. The practice will be constitutive of our reference to the objects, if not of the objects themselves'. I find this rather obscure. For one thing, the claim that the objects 'exist independently of our practice of referring to them' does not seem to square with the earlier claim that the objects are 'not selected from a previously given domain' but 'introduced into the discourse simultaneously with their assignment to

the terms'.[13] Nor is it clear how there can be such a thing as 'our reference' [to certain objects] unless the objects in question exist, in which case our practice can 'constitute our reference' [to them] only if either (i) the existence of the relevant objects *is* after all presupposed or (ii) the practice *does* – contrary to what Fine says – constitute the objects themselves.

However, the most important point for present purposes is that there seems to be ample space for a third view about how implicit definitions by abstraction work, and the conditions for their proper use, distinct from both the traditional and creative views as Fine presents them.

On this third view – the view Wright and I have tried to articulate – we do *not*, in laying down an abstraction principle as an implicit definition, *presuppose the existence of objects* to which the new singular terms are to refer. The case is entirely different from that in which we propose to introduce terms for objects of some already recognised kind – e.g. we propose to introduce a term to denote a particular number, or heavenly body, or person. Before we can properly define: 'Let t be the largest prime p such that p+2 is also prime', we have to show that there is a prime meeting the stated condition. Strictly, before we can define 'Vulcan is the planet, orbiting the Sun, whose presence is the primary cause of such and such irregularities in the orbit of Mercury', we have to show that there is such a planet – or at least provide good, if not decisive, reasons to believe that there is one. And so in other such cases. But while such a requirement is entirely appropriate in these cases, where what is being introduced is a term purporting reference to an object of some already recognised kind, it can scarcely be meaningfully applied in cases of the sort which concern us – which are precisely distinguished by the fact that we are in the business of simultaneously defining a new

[13] Of course, the two claims don't directly contradict one another – saying that the objects are introduced into the discourse *could* be construed so as to be consistent with the idea that they already exist, independently of their introduction to the discourse, waiting in the wings as it were. But then the claim that they are introduced into the discourse simultaneously with their assignment to the terms would be a flat tautology. And what has become of the contrast with being selected from a given domain?

range of singular terms and setting up a corresponding sortal concept under which their referents (if they have referents) will fall.

Nor, of course, do we somehow *create such objects* in or by giving the definition. What is stipulated, when an abstraction principle is advanced as an implicit definition, is *not* the *existence* of certain objects – referents for the terms featured on its left hand side – but the *truth* of (indefinitely many) biconditionals co-ordinating identity-statements linking such terms with statements involving the relevant equivalence relation over the underlying domain. The truth of any given one of those identity statements, and hence the existence of objects to which its ingredient terms refer, is not stipulated, but follows only given the truth of the co-ordinated statement to the effect that the abstraction's equivalence relation holds among the relevant objects (or concepts, in case of a higher-order abstraction such as Hume's Principle). And the truth of that latter statement will be always a matter of independently constituted fact (about parallelism of certain lines, or one-one correspondence between certain concepts, etc.). What is brought into existence by the stipulation – if anything is – is not objects, but a certain sortal *concept*[14]. What objects, if any, fall under it is – as I've said – entirely dependent upon the truth of instances of the abstractive biconditional's right hand side.

What we do create – if all goes well, at least – is not (new) objects but a new *concept*. Thus our view of implicit definition contrasts sharply with Fine's. As we have seen, he takes implicit definition to be 'an unorthodox mechanism for determination of *reference*', whereas for us, it is a means of fixing *sense*. Even when, as in the case of implicit definition by abstraction, the definition serves to introduce a new range of singular terms, it does so by fixing the conditions for those terms to have reference – it fixes their sense, not their reference. What reference – if any – the new terms possess depends, as I have

[14] Here and in the next paragraph, I use 'concept' as roughly equivalent to 'sense' or 'meaning', not in Frege's post-1892 sense in which a concept is the reference of a grammatical predicate (i.e. function from objects to truth-values or, roughly, property). By contrast, the concepts in the range of the second-order quantifiers in a conceptual abstraction are not senses or meanings but properties. It is a further question whether these properties should be individuated extensionally or intensionally.

explained, on matters entirely independent of the definition. In so far as that involves an application of the context principle, what is appealed to is simply the idea that it suffices for an expression to refer to an object that there be suitable true statements in which it functions as a singular term.

None of this means that anything goes, or that there aren't constraints to which proper implicit definition ought to conform. We can create, or introduce, a concept only when there is logical or conceptual room for one. To put it another way, any concept we can form must – already and independently of any choice or decision of ours – be a *possible* concept. Hence the need for the various constraints we have proposed – such as consistency, harmony, conservativeness. Now, if there is an issue here about presuppositions, it concerns – not the requirement that we justify, prior to giving an implicit definition of a type of term, the belief that there exist objects for terms of that type to refer to – rather whether we are obliged to show (in advance, as it were) that there is a possible concept, i.e. that our proposed definition conforms to such constraints. Fine is very firmly of the opinion that we are thus obliged:

> ... it is highly implausible that we might altogether avoid the epistemic cost involved in adopting Hume's Law as a classical implicit definition of number by adopting one of the less orthodox forms of definition in its place. For whatever the mechanism by which the sense or reference of the defined terms is secured, we surely require assurance that the form of words by which the definition is given can be asserted without danger of contradiction. (*Limits*, 35).

And in a connected footnote, he adds:

> It is for this reason that I do not think we can simply pass from the stipulation of Hume's Law as a definition to its assertion as a truth, without the need for further justification. Hale and Wright ... recognize the need for the definition to be consistent if the transition is to be safe, but fail to acknowledge that the definition-monger is himself under an obligation to show that the transition is safe.

Since, on our view, we implicitly define the number operator by stipulating the *truth* of Hume's Principle, there can be no question of a

transition from *stipulating* Hume's Principle *as a definition* to *asserting* it *as a truth*, and so no question of justifying such a transition. But there remains, evidently, a substantial issue over which we and Fine disagree. A proper discussion of this issue requires a paper to itself – here I can only briefly and somewhat dogmatically indicate how I see the matter.

A belief or set of beliefs cannot be true, and so cannot amount knowledge, unless it is consistent. Must we know or be able to show – if we are to be entitled to the belief (or set of beliefs) – that it is consistent? Not only does this not follow – it does not seem that it can be a perfectly general requirement, since it would be viciously regressive[15]. If that is right, then we can be entitled to hold beliefs without prior or independent entitlement to believe them consistent. I claim that something similar holds with regard to stipulations involved in implicit definition. If an expression e is to be implicitly defined by stipulation of the truth of a sentence S containing e, but otherwise composed of expressions whose meanings are already fixed, S must be consistent, i.e. it must be possible for S to express a truth. But it does

[15] For simplicity, take the case of a single belief, that p. If it is a universal requirement of entitlement to belief, that one should know or be able to show that the belief is consistent, then I must know, and so be entitled to believe, that my belief that p is consistent. But then I must know, and so be entitled to believe, that my belief that p is consistent is itself consistent, and so on. It may be claimed that since the belief that p is consistent is weaker than the belief that p, the regress is not vicious. I don't think this is correct. It is true, of course, that the claim that p is consistent is weaker than the claim that p (or that p is true), from which it follows. But it is still a significant claim. Clearly what is important is consistency in the semantic sense – possible truth – rather than just syntactic or proof-theoretic consistency, which is at most a necessary condition for semantic consistency. Of course, if one takes the relevant notion of possibility to be governed by the S4 principle, the claim that the claim that p is consistent is itself consistent (i.e. that it is possibly true that it is possibly true that p) is logically equivalent to the claim that p is consistent (i.e. that it is possibly true that p). But it would be a mistake to think, for this reason, that there is no regress – that we grind to a halt after one step. In the case with which we are concerned, I have to be entitled to believe that it is possibly true that p *independently* of any entitlement to believe that p – i.e. we must have some reason, other than my belief that p, to believe that it is possibly true that p. Let this be q. Then by the universal requirement, I must be entitled to believe that q is consistent, i.e. that it is possibly true that q. Now, either this entitlement to believe that it is possibly true that q derives from my prior entitlement to believe that q – in which case the latter entitlement does not depend on the former, and the alleged universal requirement is not universal after all – or it rests upon some other belief, r, to which I must be independently entitled, so that the regress continues.

not follow that, to be entitled to make such a stipulation, we must be able to prove this, or have some prior and independent entitlement to believe it. And in a similar way, it seems to me, a completely general requirement that any means of conferring sense on an expression must be known or shown not to lead to inconsistency, before we can be entitled to avail ourselves of it, would be equally unreasonable, because again regressive. Roughly, discharging the alleged obligation in any given case would require using some expressions whose meanings are taken as already settled – but if our entitlement to take *their* meanings as already settled depended upon a prior entitlement to believe that the means whereby that was, apparently, accomplished does not lead to inconsistency, then we would be saddled with a further obligation of the same kind, and so on. If that is right, then as in the case of beliefs, we can be entitled to proceed – in this case, employ means of fixing meanings – without prior or independent entitlement to believe in their consistency.

This line of argument does not settle a further important question. This is whether, in the case of entitlement to believe, we should accept that an entitlement to believe in consistency is *normally*, but not invariably, required – so that *absence* of an obligation to justify belief in consistency is the special case that needs to be explained or justified; or whether, rather, an entitlement to believe in consistency is *not* normally required, so that it is the *presence* of such an obligation which is the special case needing explanation or justification. And there will be a similar question regarding definitions and explanations of meaning more generally. But even on the more exacting view – that absence of an obligation to justify belief in consistency is the special case – given that we don't always have to justify belief in the consistency of our means of fixing meanings, it is plausible that when it comes to introducing/explaining the fundamental terms of a theory, we're entitled to adopt definitions without providing proofs of their consistency. If we do so, we have to acknowledge the possibility that our proposed definition will miscarry – will, perhaps in conjunction with other things we accept and are unwilling to give up, lead to

inconsistency[16]. Our not having a proof of consistency, and the (consequent) possibility that our definition will turn out bad because inconsistent, non-conservative, or otherwise defective doesn't necessarily mean we can't have knowledge – even *a priori* knowledge – based on the definition. Perhaps it would rule that out, if one supposed that certainty or infallibility is required, at least for *a priori* knowledge. But it seems to me that we should accept no such requirement, any more than we should do so in the case of empirical or *a posteriori* knowledge. What makes knowledge *a priori*, on the view I favour, is not guaranteed absence of all possibility of error, but absence of the possibility of a certain kind of error – absence of the possibility of empirical refutation.

Abstraction and inflation

I have suggested that Fine's thinking about abstraction is shaped by a certain kind of externalism, which I roughly characterized as the idea that what objects there are is – at least to a very large extent – independent of abstraction, so that the acceptability of abstraction principles is then to be viewed as a matter of their being true of a certain domain – the universe of all objects whatever – of independently constituted composition and cardinality. In this final section, I would like to air some doubts about this idea. These can most conveniently be raised in connection with one of Fine's necessary conditions for the truth of an abstraction principle. An abstraction principle always provides an identity-criterion, given by means of an equivalence relation, for the abstracts it 'generates'. The necessary condition I want to discuss is that:

> ... the identity-criterion should not be inflationary, the number of equivalence classes must not outstrip the number of objects. There must, that is to say, be a one-one correspondence between all the

[16] As indicated above, there are other constraints to which acceptable definitions should conform, besides consistency, including a form of conservativeness and more generally, harmony. I would extend my claims about consistency to these other constraints.

> equivalence classes, or their representatives, one the one hand, and some or all of the objects, on the other hand. (*Limits*, 4).

When Fine says that the number of equivalence classes (and so the number of abstracts) must not exceed the number of objects, what he means is that the number of equivalence classes (abstracts) must not exceed the number of all objects whatever. It is not difficult to see why one might take this to be a necessary condition for the truth of an abstraction principle. Let us suppose we are dealing with a conceptual abstraction – since it is only in this case that there is any danger of inflation. Then our abstraction operation will be a function from concepts to objects – that is, the domain of the function will comprise certain concepts[17], and its values will be objects. Which concepts are in the domain of the function will depend – in Fine's view, anyway, as we have seen – on an underlying domain of objects. They will be the concepts 'over' this underlying domain of objects. Now, if the underlying domain of objects includes all the objects there are, then assuming our abstraction principle is true, the abstracts it generates must be included in the underlying object domain. So there cannot be more of them than there are objects altogether – necessarily, the number of abstracts is no larger than the number of (all) objects. Thus if an abstraction principle is such that were it true, the number of its abstracts would exceed the number of objects, it can't be true. How could an abstraction principle fail in this way? The obvious answer, it may seem, is simple: suppose the domain of all objects whatever has a certainly cardinality, k. Then the number of concepts on this domain will far exceed this number. Taking concepts as individuated extensionally, there will be 2^k of them. So if our abstraction principle yields a distinct abstract for each concept – as happens if its equivalence relation is co-extensiveness – or even a distinct abstract for most concepts, there will be too many abstracts.

All this may seem quite straightforward. Let me try to explain why I don't think it is. On the face of it, the non-inflationariness condition assumes that it makes good sense to speak of the totality of all objects

[17] Concepts here in the sense of properties, rather than senses or meanings.

whatever, that there is such a thing, and that it has a number. I think this assumption – which one might think of as characteristically externalist[18] assumption – is very much open to question.

On the face of it, it makes good sense to think of one concept F as having as many instances as, or fewer instances than, another concept G only if F and G are both sortal concepts – that is, roughly, concepts with which are associated both criteria of application and criteria of identity. Thus on the widely accepted assumption that *brown* is a merely adjectival, non-sortal, concept, it makes no sense to speak of the number of brown objects, or of there being as many brown objects as there are Fs, for any bona fide sortal F. My worry about talk of the number of objects stems, initially, from a doubt on this score. *Object* isn't – or at least, isn't obviously – a sortal concept. It doesn't seem to make much sense – except in special cases, where some relevant sort can be gleaned from the context[19] – to ask how many objects there are.

It might be – and I believe often is – thought that there is no difficulty here, since we can easily supply a suitable concept, i.e. one that is both sortal and applies to each and every object whatever. The usual candidate here is self-identity – that is, we can just take the number of all objects whatever to be Nx: $x = x$, the number of self-identicals. This particular proposal is problematic, because the status of *self-identical* as a genuine sortal is open to question. Certainly it cannot be a genuine sort if – as I think – the relation of identity is itself sortally dependent, in the sense that an identity statement $x = y$ can be true only if there is some sortal concept F such that x is the

[18] Although I did not have Hilary Putnam particularly in mind when I introduced the term 'externalism' in this context, externalism as I conceive it has obvious affinities with what he calls 'external' or 'metaphysical realism'. Recall in particular that it is characteristic of Putnam's metaphysical realist to think of 'the world' as having 'a fixed totality of mind-independent objects'.

[19] For example, we can ask how many objects were removed from the accused pockets, and be given the answer: 'Five – a box of matches, three coins and a handkerchief'. It isn't easy to say precisely what is required here for something to count as one object – but it is clear that in normal circumstance, undetached parts of the handkerchief, or the images of HM the Queen impressed on the coins, would not count as distinct objects.

same F as y (briefly $x =_F y$).[20] An independent argument due to Wright seems to show that *self-identical* is not sortal. It seems undeniable that if F is any sortal concept, then so will be its restriction by any other concept G, irrespective of whether G is sortal or merely adjectival. For example, given that *horse* is sortal, *brown horse*, for example, must likewise be sortal, even though *brown* (or *brown thing)* is itself no sortal. But now if *self-identical* were sortal, *brown self-identical* would likewise have to be so. But since every object is necessarily self-identical, *brown self-identical* is equivalent to *brown* simpliciter – necessarily an object is brown and self-identical just in case it is brown. Since *brown* is <u>not</u> sortal, neither is *brown self-indentical* nor, therefore, *self-identical*. If this is right, then the seemingly good question: *How many self-identicals are there?* has no determinate answer, and 'Nx: $x = x$' has no determinate reference. There is no universal number. There is also space, I think, for a further doubt, about whether the contexts 'There are just as many Fs as Gs' and 'There are fewer Fs than Gs' are well-defined, or have determine truth-conditions, when one or both of F and G is non-sortal, and hence whether *self-identical* can be a suitable filler for F or G in those contexts.

An objector might concede that a concept F must be sortal for the *how many* question and talk of the number of Fs to be in good order, and agree that *self-identical* is therefore, as it stands, unsuitable, but argue that we can get around this and re-instate the number of all objects, by defining it slightly differently. First note that if F is sortal, then so is *self-identical F* (i.e. the concept for which the predicate 'x is the same F as x' – briefly '$x =_F x$' – stands). Of course, one can't get around Wright's difficulty just by picking some particular sortal concept F and using *self-identical F* in place of *self-identical*. More precisely, *self-identical F* will – though sortal – fail to apply to every object unless F itself does so; but if F itself is a universal sortal, then the detour through self-identity is a waste of time. We may, however,

[20] For a very clear explanation and defence of the thesis of the Sortal Dependence (as opposed to the Sortal Relativity) of Identity, see Wiggins (1980). For Wright's argument, see 'Is Hume's Principle Analytic?' in Hale & Wright (2001, 307-34). The argument is given on p. 315.

form the complex predicates '$\forall F\, x =_F y$' and '$\exists F\, x =_F y$'. And from these in turn we may form '$\forall F\, x =_F x$' and '$\exists F\, x =_F x$'. Presumably the first of these last two is true of no object whatever, and it would seem that every object whatever must satisfy the second. And this – or so it might be supposed – gives us a way out: just define the number of all objects as Nx: $\exists F\, x =_F x$.[21]

Does that settle the matter? I don't think so. A concept F's being sortal is a *necessary* condition for the how many question to be in good order and for the corrresponding term 'Nx: Fx' to have determinate reference. But it is arguably *not sufficient.* Indeed, it is fairly obviously insufficient, if there are – as there certainly seem to be – concepts which are sortal but indefinitely extensible in Dummett's sense (however precisely one thinks that difficult notion is best to be explicated[22]). The concepts of *ordinal number*, *cardinal number* and *set* all seem to be in this case. And, since the ordinals, cardinals and sets are among the objects that there are, it is plausible that any universal sortal concept must likewise be indefinitely extensible[23]. But in any case, there is a particular reason to doubt that 'Nx: Vx' can have a determinate reference. For – given that our proposed definition of the universal concept V involves quantification over (sortal) concepts – it could do so only if it were already determinate what sortal concepts there are. It can scarcely be that there is a determinate answer to the question: *How many objects are there?* – where this is construed as: *For how many x do we have* $\exists F\, x =_F x$*?* – unless there is

[21] Since $x =_F x$ iff Fx, one could just as well use '$\exists F\, Fx$' in place of '$\exists F\, x =_F x$'. Of course, this way out is good only if the concept $\exists F\, x =_F x$ is itself a genuine sortal concept. The mere fact that '$\exists F\, x =_F x$' is true of every object is certainly not enough to make it a sortal predicate – any more than the fact that 'x has mass' is true of every physical object is enough to make it a sortal predicate of physical objects. For *being, for some F, the same F as itself* to be a genuine sortal, there needs to be a criterion of identity for the objects falling under it. But so, it seems, there is. Let us abbreviate our predicate '$\exists F\, x =_F x$' by 'Vx'. Suppose b and c both satisfy 'Vx'. What condition is both necessary and sufficient for b and c to be one and the same V? Well, the obvious answer is that b and c are one just in case for some single F, $b =_F c$. 'V' has thus both a criterion of application – Vx iff for some F, $x =_F x$ – and a criterion of identity – $x =_V y$ iff for some F, $x =_F y$.

[22] For a rough explanation, see the below, six paragraphs on.

[23] This will be so, if we can assume that if F is indefinitely extensible and $\forall x(Fx \rightarrow Gx)$, then G is likewise indefinitely extensible.

a determinate answer to the question: *What sortal concepts are there?* It is at least not obvious that there can be a determinate answer to *that* question.

Someone might protest: "There is no difficulty over that. For any given domain of objects, the corresponding domain of concepts is fixed. For each and every way of dividing the domain of objects, there is a concept, and those are all the concepts. If the domain of objects comprises k objects, there are thus 2^k concepts." But we have been here before – to assume a domain of objects 'given' is simply to fail to engage with the problem, or to assume it somehow solved. We cannot both assume a given domain of objects as a means of fixing the range of the quantifier 'For some *F*', and at the same time use that quantifier to define the supposedly sortal predicate '*V*' (i.e. '*x* is an object').

If a domain of objects is already somehow fixed as comprising k objects, then it is, of course, quite right that there are 2^k concepts on the domain – at least provided that concepts are individuated extensionally. But whilst there is no general objection to treating concepts extensionally – as, in effect, determined simply by what objects fall under them – it is questionable whether they are appropriately so treated in the present context, for at least two reasons.

The first, more general, reason is that it makes sense to think of concepts extensionally only if it is already determinate what objects belong to the domain on which they are to be thought of as defined. That condition may well be met in a particular case – it will be met if, for example, we are considering the domain comprising exactly the natural numbers. But it clearly cannot be assumed met in the present case.

Secondly, and more specifically, the whole point of insisting that an identity-statement $x = y$ has to be understood as true if and only if x and y are one and the same F, for some appropriate sortal F, is lost, if the covering sortal F is thought of as determined purely extensionally. The point is – at least in part – that objects cannot be individuated save as instances of some sortal concept or other, so that unless some appropriate sortal is specified or understood from the context, it is simply not determinate what is being asserted, when it is said that

$x = y$. If objects could be individuated simply as objects, there would be no justification for insisting that '$x = y$' must be understood so that it can be true only if $x =_F y$ for some specific sortal F, such as *horse*, *person*, *number* or the like – any identity-statement x = y could be understood as claiming simply that x is the same *object* as y.

If what I've said is right, the universal concept *self-identical under F, for some sortal F* exhibits something akin to the property of indefinite extensibility. I'm not sure that it *is* indefinitely extensible in the usual sense, which requires, for a concept G to be indefinitely extensible, that given any definite collection of Gs, there is an object satisfying the requirements for being G which cannot, on pain of contradiction, be one of that collection. I shall say that the universal concept is *sortally indeterminate*, leaving open the question of the relations between sortal indeterminacy and indefinite extensibility. Sortal indeterminacy differs from – and is in a way more radical than – that more familiar kind of indeterminacy we usually call vagueness. A concept is vague – at least on some widely accepted views of vagueness – when there are objects of a sort to which the concept applies, in regard to which it is indeterminate whether the concept applies or not. Some sortal concepts are vague (e.g. *novel*, *large bird*) and naturally to these no definite number can be assigned – but at least in some cases we may still be able to assign lower and/or upper bounds on their number, and so claim truly or falsely that there are more/fewer of their instances than there are Fs, for some other (possibly also vague) sortal (e.g. there are more large birds than there are small opera houses). But the sortal indeterminacy of the concept of object captured by $\exists F\, x =_F x$ is evidently not a matter of their being *objects* – or entities of any kind – with regard to which it is indeterminate whether or not *they* fall under the concept. The indeterminacy results from its not being determinate what sortal concepts there are. Of course, there may be a determinate collection of sortal concepts expressed, or expressible, in some given language. And maybe any sortal concept has to be expressible in some possible language. But is there a determinate collection of all sortal concepts? This seems to me doubtful. And for this reason, it seems to me

doubtful that one can – as one can with vague concepts – sensibly talk of there being more concepts than objects.

If these doubts are well-founded, it seems to me that one cannot satisfactorily formulate a necessary condition for the truth, or acceptability, of an abstraction principle in terms of its generating no more abstracts than there are objects altogether. Whether there is a way of reformulating Fine's requirement which captures the basic idea that good abstractions should be non-inflationary is a good question, but I do not know the answer to it. It seems clear that there is no evident reason to ban abstractions which inflate on definite domains of determinate cardinality. Fine himself is, as I understand it, happy enough with abstractions which, like Hume's Principle, inflate on domains up to a certain cardinal size and stabilize thereafter. And I cannot myself see any reason to object to abstractions which inflate on all cardinally definite domains – such abstractions would generate a universe of abstracts which, like the universe of sets as standardly conceived, is not closed off, i.e. has itself no cardinal size. I am not sure that it is possible to characterize an objectionable kind of inflationariness without recourse to the idea that it is the kind of inflationariness that leads to contradiction. But if that's the best one can do, the requirement seems to me to boil down to the requirement of consistency, which we must have anyway.[24]

[24] I am very grateful to Kit Fine for extensive and very helpful discussion of our points of disagreement. I should also like to thank the organisers and other participants in the Neuchâtel Logicism conference for stimulating discussion of a somewhat shorter version of this material.

References

FINE K. (2002). *The Limits of Abstraction*. Oxford: Clarendon.

FREGE G. (1879). *Begriffsschrift*. Halle: L. Nebert. [Engl. trans. in van Heijenoort J. (ed). *From Frege to Gödel: A Source Book in Mathematical Logic, 1879-1931*. Harvard University Press. 1967].

FREGE G. (1884). *Die Grundlagen der Arithmetik*. Breslau: W. Kœbner. [Reprinted with Engl. trans. by Austin J. L. *The Foundations of Arithmetic*. Oxford: Blackwell. 1950].

HALE B. (1997). *Grundlagen* §64. *Proceedings of the Aristotelian Society* XCVII. 243-261.

HALE B. & WRIGHT C. (2001). *The Reason's Proper Study*. Oxford University Press.

PARSONS C. (1965). Frege's Theory of Number. In Black M. (ed). *Philosophy in America*. Ithaca: Cornell University Press. 180-203.

WIGGINS D. (1980). *Sameness and Substance*. Oxford: Blackwell.

WRIGHT C. (1983). *Frege's Conception of Numbers as Objects*. Aberdeen University Press.