Notre Dame Journal of Formal Logic
Volume 44, Number 4, 2003

Book Review

Kit Fine. *The Limits of Abstraction*. Clarendon Press, Oxford, 2002. x + 204 pages

1 Introduction

Kit Fine's long article [13], introducing his distinctive take on neo-Fregeanism, has now been expanded into a short book of the same title. (For those familiar with the article version, the philosophical material from it appears as chapter 1 in the book, and the technical material as chapters 3 and 4. According to the book's preface, "The major change is the addition of a new part on the context principle." This addition constitutes chapter 2 of the book. There is also an index of technical terms, which would have been more useful if it had been arranged alphabetically.)

The present review of that book is divided into three parts of unequal length. The long introduction Section 2 surveys recent neo-Fregeanism. Then Section 3 summarizes Fine's technical contributions, which presumably are what is of primary interest for readers of the present journal. The brief conclusion Section 4 touches on more purely philosophical issues.

2 Neo-Fregeanism

The tale has often been told of Frege's heroic effort to ground mathematics in logic and of Russell's devastating discovery of a contradiction in the basic assumption underlying that effort. Nonetheless, I will begin with a very brief retelling of the tale. Frege scholars will recognize it as a simplification of the real history. No brief account can be anything but.

Frege's most enduring achievement was his first one, the creation of modern logic in his *Begriffsschrifft*. But his logical system was a grander affair than the logic one finds in present-day textbooks, and we must begin with a description of its key features. Underlying the logic is a logical grammar distinguishing various types of expressions. Fundamental are sentences (S) and "names" (N) as Frege calls them, though they included all singular terms, both proper names plus singular definite descriptions. An expression with k blanks in it, that if filled with expressions of

Printed July 7, 2004
2000 Mathematics Subject Classification: Primary, 03A05, 00A30
Keywords: neo-Fregeanism, abstraction, context principle

types $T_1, \ldots, T_k$ will produce an expression of type S may be said to be of type $T_1 \cdot \ldots \cdot T_k \to S$, and similarly for N. Corresponding to each grammatical type T of expression is an ontological type O of entity to which expressions of type T are taken to refer. The terminology that will be used here for the types of entity there will be occasion to mention below is indicated in the adjoining table, as is the distinctive style of variable that will be used for each type. However, the expression "concept" is often also used in a more general sense to cover not just one-place, first-level concepts, corresponding to type $N \to S$, but for the entities corresponding to any type of the form $T_1 \cdot \ldots \cdot T_k \to S$. Context should make plain whether one is speaking of concepts as opposed to relations and superconcepts, or concepts including relations and superconcepts.

Type	Example	Referent	Variables
S	Socrates is wise	truth value	
N	Socrates	object	$x, y, \ldots$
$N \to S$	... is wise	concept	$X, Y, \ldots$
$N^2 \to S$	... taught—	relation	$R, S, \ldots$
$(N \to S) \to S$	Someone ...	superconcept	$\mathbf{X}, \mathbf{Y}, \ldots$
$(N \to S)^2 \to S$	Whoever ..., —	superrelation	$\mathbf{R}, \mathbf{S}, \ldots$

Table 1 Partial list of Fregean grammatico-ontological categories.

When, for instance, an expression of type N refers to an object x, and an expression of type $N \to S$ to a concept X, and the sentence obtained by putting the former expression into the blanks of the latter expression is true, then the object is said to *fall under* the concept. If Socrates is wise, then Socrates falls under the concept of being wise. To indicate that x falls under X one writes $X(x)$. From atomic formulas of this kind, and their analogues for other types, together with atomic formulas of the kind $x = y$, more complex formulas can be built up using the logical operations of negation, conjunction, disjunction, and the like, as well as universal and existential quantification over entities of all types.

It is a background assumption, the axiom of *concept comprehension*, of Frege's logic that for every such formula $\Phi(x)$, expressing a condition that may or may not hold of an object x, there is a corresponding concept, one under which an object falls if and only if the condition does hold of it.

$$\exists X \forall x (X(x) \leftrightarrow \Phi(x)) \tag{1}$$

An analogous assumption is made for many-place and higher-level concepts. It is to be understood that there may be *parameters* in (1), meaning free variables other than the one displayed, and that what is really being taken as the axiom is the *universal closure* of what has been displayed. Thus if there are two object parameters u and v, and a concept parameter W, the axiom is really the following:

$$\forall u \forall v \forall W \exists X \forall x (X(x) \leftrightarrow \Phi(x, u, v, W)) \tag{2}$$

Similarly with other schemes below.

The notion for concepts that is analogous to identity among objects is called *coextensiveness*. It will be convenient to introduce an abbreviation for this notion, which is definable in terms of logical operations for which we already have notations:

$$X \equiv Y \leftrightarrow \forall x(X(x) \leftrightarrow Y(x)) \tag{3}$$

The notation may also be used for many-place and higher-level concepts.

Frege also introduces a proof-procedure. When one dispenses with all concept variables, one has first-order logic, and Frege's proof-procedure is equivalent to those in present day logic textbooks. When one dispenses with all but the first level of concepts and relations, one speaks of second-order logic, and when one dispenses with all but the first two levels of concepts and relations and superconcepts and superrelations, of third-order logic. So far as proof-procedures are concerned—and proof-theoretic or syntactic deducibility, not model-theoretic or semantic consequence, which had not been born or thought of in Frege's day, is all that was of concern to Frege, and all that will be of concern to us—so-called second- and third- and higher-order theories are simply first-order theories with several styles of variables, and the distinctive axiom of comprehension (1) (to which may be added an axiom of extensionality for concepts and an axiom of choice, details of which need not concern us here).

To this background logical apparatus Frege adds, for the purpose of developing mathematics, the axiom of *extension existence*, to the effect that to any concept X there corresponds an object $\ddagger X$, called the *extension* of X, in such a way that the extensions of two concepts are identical if and only if the concepts themselves are coextensive:

$$\ddagger X = \ddagger Y \leftrightarrow X \equiv Y \tag{4}$$

(The operator $\ddagger$ is of type (S $\rightarrow$ N) $\rightarrow$ N.) An analogous assumption is made for many-place and higher-level concepts.

The extensions of one-place concepts were called "classes" by Frege. The connection that holds between one object x and another object y when the latter is the extension of a concept under which the former falls, amounts in this terminology to membership of object x in class y, and we may introduce the usual epsilon-notation as an abbreviation:

$$x \in y \leftrightarrow \exists Y(y = \ddagger Y \;\&\; Y(x)) \tag{5}$$

Most writers after Frege have substituted the Cantorian terminology of "set" and "element" for that of "class" and "member". Some have used the term "class" in a sense close to Frege's "concept". To avoid ambiguities, I will avoid the term "class" altogether (except as an inseparable part of the set phrase "equivalence class") and continue to speak of objects and "concepts", and to call those objects that are extensions of concepts "sets". Fortunately there will be no occasion to mention the extensions of relations, and there will be no need for a label for them. (Frege had available in German *two* common enough but quite different words, the Teutonic *Beziehung* and the Latinate *Relation*, and was able to use one for many-place concepts and the other for their extensions.)

Note that (1) and (4) (and the definition (5)) together imply the axiom of *set-comprehension*, according to which for any condition there exists a unique corresponding set, one of which an object is an element if and only if the condition holds of it:

$$\exists! y \forall x(x \in y \leftrightarrow \Phi(x)) \tag{6}$$

This set y is usually denoted by the term $\{x : \Phi(x)\}$. An alternative formulation (closer to that of Frege himself) would simply allow the introduction of such terms into the official language, and corresponding to (6) would assume the following:

$$x \in \{x : \Phi(x)\} \leftrightarrow \Phi(x). \tag{7}$$

A superrelation $\mathbf{R}$ is called an *equivalence* if it is reflexive, symmetric, and transitive, so that we have

$$\begin{aligned}&\forall X \mathbf{R}(X, X) \ \& \\ &\forall X \forall Y (\mathbf{R}(X, Y) \rightarrow \mathbf{R}(Y, X)) \ \& \\ &\forall X \forall Y \forall Z (\mathbf{R}(X, Y) \ \& \ \mathbf{R}(Y, Z) \rightarrow \mathbf{R}(X, Z)).\end{aligned} \tag{8}$$

In Frege's system one may define for any such $\mathbf{R}$ and any concept X the *equivalence class*—even those who say "set" in ordinary contexts say "class" in this special context—$\S_{\mathbf{R}} X$ of X with respect to $\mathbf{R}$, as follows:

$$\S_{\mathbf{R}} X = \{Y : \mathbf{R}(X, Y)\}. \tag{9}$$

One then has the following:

$$\S_{\mathbf{R}} X = \S_{\mathbf{R}} Y \leftrightarrow \mathbf{R}(X, Y). \tag{10}$$

Note that $\ddagger X$ is the same as $\S_{\equiv} X$.

One could make similar definitions one level down, with equivalence relations on objects, but it is equivalence superrelations on concepts that do the most work in Frege's system. For purposes of developing arithmetic, Frege first defines an equivalence, here to be written $\approx$, of *equinumerosity*, as follows:

$$\begin{aligned}X \approx Y \leftrightarrow \exists R (\forall x (X(x) \rightarrow \exists ! y (Y(y) \ \& \ R(x, y))) \\ \& \ \forall y (Y(y) \ \rightarrow \ \exists ! x (X(x) \ \& \ R(x, y))))\end{aligned} \tag{11}$$

Frege then defines the *number* of a concept X to be equivalence class of X with respect to this equivalence. Writing $\# X$ for $\S_{\approx} X$, we then have as a special case of (10) the following, which has come to be called *Hume's Principle* or HP:

$$\# X = \# Y \leftrightarrow X \approx Y \tag{12}$$

In the remainder of the development of arithmetic Frege so to speak "forgets" how numbers were defined, and uses *only* (12). (That Frege in fact uses only (12), without further recourse to (4), in developing arithmetic was first noted explicitly by Charles Parsons [22]. That Frege *self-consciously* uses only (12) without further recourse to (4) has been suggested by Richard Heck. In view of this latter suggestion, the verb "forgets" should not to be taken literally.)

In general, when one assumes (10) independently of or while forgetting the definition (9), one avoids the term "equivalence class" and calls $\S_{\mathbf{R}} X$ the *abstract* of X with respect to $\mathbf{R}$. Thus Frege develops arithmetic on the basis of the sole assumption that numbers are abstracts of concepts with respect to equinumerosity. On this basis Frege defines the specific numbers zero, one, two, and so on, as well as the general notions of natural number and the successor, and proves the so-called Peano postulates for these notions. One can then go on to define addition and multiplication with the usual recursion equations, to prove the usual associative and commutative and distributive laws, to define exponentiation, to prove the usual laws of exponents, and so on. Frege does not "remember" the general notions of extension and equivalence class until he wants to go beyond arithmetic to analysis and has to introduce real numbers, which again are introduced as certain equivalence classes or abstracts.

Unfortunately, just as Frege was putting the finishing touches on his *magnum opus*, expounding all these developments, Russell, who had independently rediscovered some of Frege's ideas, discovered the famous Russell paradox, that (7) leads to a contradiction when applied to the condition "x is a set $\;\&\; x \notin x$".

Looked at from a Cantorian point of view, the problem with Frege's system appears to be this. Suppose there are κ objects and write $\exp(\kappa)$ for 2^κ. Then (1) seems to tell us there will be $\exp(\kappa)$ concepts, one for each set of objects (corresponding to the condition of being an element of that set), while (4) seems to tell us that there will be at least as many objects as concepts (since ‡ gives a one-to-one function from the latter to the former). Together, they tell us that $\exp(\kappa) \leq \kappa$. But this is impossible, since Cantor has proved $\kappa < \exp(\kappa)$ for all κ. Analyzing Cantor's proof as it applies to this situation, one is led to consider the Russell set, and this is in fact how Russell discovered his paradox. Or so he tells us in an autobiographical sketch ([27], p. 44).

The dream ever since the discovery of Russell's paradox has been to find some comparatively simple and natural modification of Frege's inconsistent assumptions that would on the one hand render his system consistent, but on the other hand still leave it adequate for the development of mathematics. Pursuit of this dream is the project of *neo-Fregeanism* in the broadest sense of that label.

Some are likely to object that the development of axiomatic set theory has rendered Frege's work of historical interest only. Set theorists now work, it will be said, with a well-understood, standard conception of set, the *iterative* conception, which motivates a well-understood, standard system of axioms, **ZFC**, and suggests with varying degrees of compellingness certain further axioms, called *large cardinal axioms*. This being so, it will be said, tinkering with repairs to Frege's antique system is pointless. (This hypothetical objection bears more than a passing resemblance to remarks in the eminent set theorist D. A. Martin's notorious review [20] of a book on set theory by the eminent philosopher W. V. Quine. See also the latter's reply [25].)

In response to this objection it must be conceded that a neo-Fregean project will be of little interest unless it achieves a reconstruction of a very substantial part of classical mathematics. But few philosophers who have read philosophical accounts of the iterative conception of set, or accounts of attempts to motivate large cardinal axioms, can seriously believe that the iterative conception provides a single, unified principle clearly sufficient to develop all the axioms of **ZFC**, let alone large cardinals. (For a skeptical look at the iterative conception, see [4].)

We know from Gödel's second incompleteness theorem that we cannot develop classical arithmetic, let alone analysis or higher set theory, on a basis that is self-evidently consistent, let alone self-evidently true. Still, it is not obvious that we cannot hope to do better in the way of providing heuristic motivations for our fundamental assumptions than has been done by set theorists in general or large cardinal theorists in particular. While some of us would not be, many philosophers might be willing to sacrifice higher set theory if a neo-Fregean system adequate for basic arithmetic and analysis could be developed that seemed philosophically better-founded.

In considering the issue how substantial a part of classical mathematics a given neo-Fregean project can succeed in reconstructing, it is useful to have in mind a scale of more and more substantial parts established by mathematical logicians. Some key points on such a scale are illustrated in the adjoining table.

The table is perhaps best read from the bottom up. A precise specification of the weak systems of arithmetic in the table will not be material for our purposes. Suffice

Theory	Name	Domain(s)
$\mathbf{ZF}^2$	Morse-Kelly set theory	Pure sets of all ranks, concepts
ZF	Zermelo-Frankel set theory	Pure sets of all ranks
Z	Zermelo set theory	Pure sets of all ranks
$\mathbf{P}^\omega$	Simple theory of types	Individuals, sets of, sets of sets of, etc.
$\mathbf{P}^3$	3rd order arithmetic	Numbers, sets of, sets of sets of
$\mathbf{P}^2$	2nd order arithmetic	Numbers, sets thereof
$\mathbf{P}^1$	Peano arithmetic	Numbers
$\mathbf{Q}_\omega$	Grzegorczyk arithmetic	Numbers
$\mathbf{Q}_2$	Kalmar arithmetic	Numbers
$\mathbf{Q}_1$	Wilkie arithmetic	Numbers

Table 2 Partial scale of strength of fragments of classical mathematics.

it to say that in $\mathbf{Q_1}$, called in the literature $\mathbf{I\Sigma_0}$, or sometimes **PFA** for *polynomial functional arithmetic*, one can speak of zero, successor, addition, and multiplication, and prove basic laws for these. But the scheme of mathematical induction,

$$(\Phi(0)\ \&\ \forall x(\Phi(x) \to \Phi(Sx))) \to \forall x\Phi(x),$$

which is essential for proving many more serious number-theoretic theorems, is only available for a restricted kind of formula, in which all quantifiers are bounded, so that one can say "there is a $u \leq t \ldots$" but not "there is a $u \ldots$" in a condition about which one wants to do a proof by induction. $\mathbf{Q_2}$, called in the literature $\mathbf{I\Sigma_0}$(exp), or sometimes **EFA** for *exponential functional arithmetic*, adds exponentiation. $\mathbf{Q_3}$, called in the literature $\mathbf{I\Sigma_0}$(superexp) adds superexponentiation, and so on. Continuing in this way $\mathbf{Q}_\omega$ has all the operations in the series that begins with addition, multiplication, exponentiation, and so on, a sequence in which each operation is obtained by iterating the preceding one. This has the effect of making available all primitive recursive functions; the system is a variant version of the system called *Skolem arithmetic* or **PRA** for *primitive recursive arithmetic*.

$\mathbf{P^1}$, often called **PA**, is the system of formal arithmetic most often met with in textbooks of intermediate logic. It adds an *unrestricted* axiom of scheme of induction, which renders the inclusion of symbols and defining equations for exponentiation and other primitive recursive functions beyond addition and multiplication redundant. $\mathbf{P^2}$, often called *analysis*, is usually understood today as dealing with natural numbers and sets thereof rather than concepts applicable thereto, so that the notation $x \in X$ rather than $X(x)$ is used, though so far as what can be *proved* is concerned, the difference in notation and terminology makes no difference. $\mathbf{P^3}$ adds, on what is today the conventional understanding, one more layer of sets, or on a Fregean understanding, superconcepts. With the expansion of the language, there are more instances of the induction scheme, and appropriately applied these make the introduction of symbols and defining equations for addition and multiplication redundant. $\mathbf{P^4}$ would add another level, and so on.

Continuing in this way one arrives at $\mathbf{P}^\omega$ (which is, by the way, the system for which Gödel originally proved his incompleteness theorems). This in effect amounts to a variant version of the theory **PM** of types as developed by Russell and Whitehead in Principia Mathematica and simplified by Ramsey. Again today this is understood as involving sets of various levels, though on Russell's original understanding it involved instead "propositional functions". It is customary in type theory to drop the assumption that the objects at the bottom level are numbers specifically and to speak and think of them instead as simply individuals of some unspecified kind. One does not really need the Peano postulates about zero and successor, but only the assumption that there are infinitely many individuals (which can be expressed by saying there is a function that, like the successor function, is one-to-one, and there is an object that, like zero, is not in the range of this function). Type theory is just slightly weaker than Zermelo's axiom system **Z** for set theory, which in turn is significantly weaker than Frankel's amendment thereto, **ZF**, which is required for the development of Cantor's theory of transfinite ordinals and cardinals. $\mathbf{ZF}^2$, usually called **MK**, is slightly stronger. It is weaker than the theory obtained by adding to **ZF** even the weakest of the large cardinal axioms set theorists have investigated, but there will be no need to enter into the realm of large cardinals here.

A theory T_1 is interpretable in a theory T_2 when there is a translation, preserving logical structure, from the language L_1 of T_1 into the language L_2 of T_2, which translates every axiom and hence every theorem of T_1 into an axiom or theorem of T_2. Interpretability is easily seen to hold whenever T_1 is a subtheory of T_2 (using the identity function as a translation). Interpretability is also easily seen to be transitive (given that T_0 is interpretable in T_1 and T_1 is intepretable in T_2, composing the translation functions gives a translation function showing T_0 is interpretable in T_2). Instead of saying "interpretable" we may say "of lesser or equal interpretability strength", and then define "of strictly lesser interpretability strength" and "of equal interpretability strength" in the obvious way. In this sense each theory in the table is of strictly greater interpretability strength than the theories listed below it. Generally one proves interpretability by exhibiting a specific translation. For reasons connected with Gödel's second incompleteness theorem, to prove noninterpretability of T_2 in T_1 it generally suffices to show that the consistency of T_1 can be proved in T_2, and this is generally the method used.

It is important to note that a theory may be interpretable in a proper subtheory of itself. The best-known case occurs near the top of the table, in connection with the work of Gödel and Cohen on the consistency and independence of the axiom of choice (AC), the continuum hypothesis (CH), the generalized continuum hypothesis (GCH), and the axiom of constructibility (V $=$ L). Each of the theories **ZF**, **ZFC** $=$ **ZF** $+$ AC, **ZFC** $+$ CH, **ZFC** $+$ GCH, **ZFC** $+$ V $=$ L is a proper subtheory of the next: AC is not provable in **ZF**, CH is not provable in **ZFC**, GCH is not provable in **ZFC** $+$ CH, and V $=$ L is not provable in **ZF** $+$ GCH. Yet **ZFC** $+$ V $=$ L is interpretable in **ZF** (by relativizing all quantifiers to sets for which a certain condition holds, called "constructibility").

In the middle of the table, the vague statements that the union $\mathbf{Q}_\omega$ of $\mathbf{Q}_1$, $\mathbf{Q}_2$, $\mathbf{Q}_3, \ldots$ is a "variant version" of **PRA**, and that the union $\mathbf{P}^\omega$ of $\mathbf{P}^1$, $\mathbf{P}^2$, $\mathbf{P}^3, \ldots$ is a "variant version" of **PM**, would be put more precisely as statements about interpretability. At the bottom of the table, by a theorem of Wilkie, $\mathbf{Q}_1$ is interpretable in an ostensibly much weaker theory, *Robinson arithmetic* or **Q**, in which one cannot even prove the associative, commutative, and distributive laws. Indeed, a significant amount of arithmetic, algebra, and analysis can already be carried out, using suitable interpretations, at this level of interpretability strength. (For Wilkie's theorem, and more generally for precise definitions of and information about weak theories of arithmetic, with further references to the literature, the standard reference is [15]. A significant work too late to be mentioned there is [11].)

The theories on the list constitute targets a neo-Fregean may shoot for, in the sense that the goal is to develop a neo-Fregean theory in which theories as high up the list as possible can be interpreted. But generally speaking, the bulk of pure and applied mathematics can be squeezed in $\mathbf{P}^2$ and fit more comfortably into $\mathbf{P}^3$. What lies beyond is higher set theory. Higher set theory does have implications about, for instance, real numbers and comparatively simple sets thereof, but these are generally of interest only to specialists in certain areas not quite at the core of mainstream mathematics. (For an early example of a fairly down-to-earth result requiring higher set theory, see [14]. The same author has for three decades now been producing any number of additional examples.) And to return to the objection noted earlier, philosophical gains may compensate for mathematical losses, so many philosophers might be content with a weaker theory than the strongest on the list, if it is better-founded.

What would constitute being philosophically "better-founded" is an issue that may be postponed until some actual neo-Fregean theories have been produced for evaluation. But surely the requirement that whatever restriction is imposed on Frege's inconsistent system should be a "simple and natural" one is a crucial constraint if a neo-Fregean project is to have much philosophical significance. Needless to say, "simple and natural" is not a precisely-defined technical term like "consistent" or "sufficient to interpret T", where T is any of the theories on our list. Whether or not the constraint of simplicity and naturalness has been achieved in a given case will be a matter more of philosophical opinion than of mathematical fact. But the constraint is no less important for all that.

So much for neo-Fregean targets. Let us now consider strategies for reaching them. To begin with, note that set comprehension (7), which leads to paradox, derives from two assumptions, concept comprehension (1) and the existence of extensions (4), which latter is one instance of the existence of abstractions (10). There are accordingly multiple approaches one might take to repairing Frege's system: (A) restrict (1) and keep (4) unrestricted; (B) restrict (4) and keep (1) unrestricted; (B′) restrict (4) and restrict (1) also; or (C) drop (4) altogether, while keeping some other instance(s) of (12). Actual specimens of each of these strategies are listed in the adjoining table, along with an indication of how strong a fragment of classical mathematics becomes interpretable on each approach, before resorting to ad hoc additional assumptions. In all cases, the background logic is second-order.

The strategy of restricting comprehension was first investigated by Richard Heck, though it is impossible to do justice to his approach within the simple second-order framework adopted here. Insofar as the approach of restricting comprehension *can* be represented in our framework, the most obvious restriction to impose would be to assume (1) only for formulas $\Phi(x)$ that may contain quantifications involving object variables but none containing concept variables. This gives what are called the *simple predicative* concepts. The Russell paradox is blocked, because when one unpacks the definition (5) of $\in$, the condition of x being a set that is not an element of itself involves an ineliminable quantification over concepts. If one takes the original concept variables to range over simple predicative concepts, and adds a new style of concept variable, assuming for these (1) for formulas that may contain quantifications involving object *or* simple predicative concept variables, but not these new

Kind of Restriction	Author	Strength
None	Frege	$\rightarrow\leftarrow$
Restrict Comprehension	Heck et al.	$\mathbf{Q}_2$
Restrict Extensions: Zig Zag		
	Quine	$\rightarrow\leftarrow$
	Quine-Wang	?
Restrict Extensions: Limitation of Size		
unrestricted comprehension	von Neumann-Boolos	$\mathbf{P}^2$
restricted comprehension	von Neumann	$\mathbf{P}$
Drop extension, Keep some Abstracts		
equinumerosity	Wright et al.	$\mathbf{P}^2$
non-inflating, very nearly invariant	Fine	$\mathbf{P}^3$

Table 3 Partial chart of post- and neo-Fregean strategies.

variables, one gets the *second-degree* predicative concepts. Continuing in this way, introducing more styles of variables and more degrees of predicative concepts, produces the *ramified predicative* hierarchy of concepts. The full, unrestricted, original version of (1) is then called *impredicative* by contrast.

With just simple predicative comprehension and extension existence, one has of course extensionality, and can at least prove the existence and uniqueness of the null set and of the result of adding any given element to any given set. This is already enough to give the interpretability of $\mathbf{Q}$ by an old result of Wanda Szmielew and Alfred Tarski, and therefore of $\mathbf{Q}_1$ by Wilkie's theorem. (The Szmielew-Tarski result is stated without proof in the same work where the system $\mathbf{Q}$ is introduced [30]. A proof is given in [9]. An alternate proof, showing that extensionality is not needed, is given in [21]. With ramified predicative comprehension, one can get $\mathbf{Q}_1$ without relying on the difficult proof of Wilkie, by a method due to A. P. Hazen, and one gets $\mathbf{Q}_2$ as well. See [8].) And with this weak arithmetic one also gets the modicum of algebra and analysis that is interpretable therein. But that seems to be as far as one can go in terms of our list of fragments of classical mathematics, since the next item on the list, $\mathbf{Q}_\omega$ is strong enough to prove the consistency of the theory. (The observation is mentioned in [8], and as there stated was first made to the reviewer by Saul Kripke.)

The work of Heck is set in a more elaborate framework, as is that of Fernando Ferreira and Kai Wehmeier, but the results obtained remain well below $\mathbf{P}^2$, since the consistency of the theories involved can be proved there. For Heck [18] the question of interest is as much historical as logical: Could Frege have avoided contradiction by predicativity assumptions, while still carrying out his derivation of arithemtic? For this reason Heck employs a framework as close as possible to that of Frege himself, apart from the predicativity restriction, and this means that his framework in effect allows terms $\{x : \Phi(x)\}$ and objects denoted by them even if $\Phi(x)$ is impredicative, though it does not allow the assumption of (7) about those objects unless $\Phi(x)$ is predicative. Heck proves the interpretability of $\mathbf{Q}$ in the simple predicative theory of this kind *by methods close to Frege's own derivation of arithmetic from* (12) except at one crucial point, the proof that the successor function on natural numbers is total. The presence of the terms $\{x : \Phi(x)\}$ makes proving consistency harder, but Heck also proves the consistency of the corresponding ramified predicative theory, building on work of Terence Parsons [24].

Ferreira and Wehmeier [12] answer a question left open by Heck, proving the consistency of the related theory that allows (1) provided Φ involves only a single initial universal quantification over concepts, and is assumed equivalent to some Ψ that involves only a single initial existential quantification over concepts. Wehmeier [33] points out and remarks negatively on the peculiarity of such systems that, while purportedly purely logical, they can prove the existence of *individuals*, or objects that are not extensions of concepts.

Turning now to restrictions on (4) and thence on (7), Russell early noted two possible strategies: *limitation of size* approaches, which allow the existence of $\{x : \Phi(x)\}$ only if there are *not too many* objects x for which the condition $\Phi(x)$ holds, and *zigzag* approaches, which allow the existence of $\{x : \Phi(x)\}$ only if the condition $\Phi(x)$ itself is not too complicated. For Russell's efforts, eventually abandoned, to pursue the zigzag approach, see [32] and the editorial commentaries in [29].

After Russell himself, the most prominent writer to follow a zigzag approach was Quine. The particular restriction Quine wished to place on $\Phi(x)$ he called *stratification*, and the resulting system has been called **ML**. Like other set theorists who allow anything like Fregean "concepts", Quine calls them "classes". Unlike other set theorists who allow "classes" in addition to sets, Quine does not have two distinct styles of variables for sets and classes, and does not distinguish between a set and the class whose members are the elements of that set: sets are simply some among the classes, namely, those capable of being elements of others. Allowance being made for both these differences, it may be said that in *both* editions Quine allows impredicative comprehension in the sense that any condition determines a "class"; but there is a restriction on which kinds of conditions are assumed to determine sets. The difference between editions is that a more severe restriction is imposed in the second than in the first. The details need not concern us, since J. B. Rosser derived a paradox in **ML** (1st edition). A repair, suggested by Hao Wang, produces a system **ML** (2nd edition), where the Rosser derivation is blocked. However, the consistency even of this system remains in doubt. (See the preface to the second edition in [26].)

George Boolos has evaluated what can be accomplished on the limitation of size approach, if one understands "not too many" to mean "not as many as there are objects altogether". Let U be the universal concept, the concept under which all objects fall, which is to say the concept of self-identity, given by the condition $x = x$. Then more formally the proposal is to allow (4) on the hypothesis that one does *not* have $X \approx U$. This particular understanding of limitation of size is often called the *maximal principle*. Boolos himself speaks of the principle of *small extensions*. The system that results is strong enough to interpret $\mathbf{P}^2$, as he observes in [4].

One does not get more because, unfortunately, the assumption that a concept has an extension unless the objects falling under it are *as many as there are objects* does not imply the conclusion that a concept has an extension unless the objects falling under it are *very, very many* without the further assumption that there are very, very many objects. To get the usual infinity axiom (asserting the existence of not just infinitely many finite sets but of an infinite set) and power axiom (asserting the existence of the set of all subsets of any given set) of **ZF**, one has to keep going back, so to speak, to the Court of Limitation of Size, to get new rulings about how many are too many: that infinitely many are not too many, that as many as there are subsets of any given set is not too many. With infinity axiom as an ad hoc additional assumption, one gets $\mathbf{P}^3$; with power as a further such assumption, one gets $\mathbf{ZF}^2$.

The idea of limitation of size is a borrowing from Cantor, who distinguished the merely transfinite consistent multiplicities, which form sets, from the absolutely infinite inconsistent multiplicities, which do not. And as Boolos notes, the maximal principle is a borrowing from von Neumann's attempt to reduce Cantor's inchoate theory to rigorous axiomatic form. Von Neumann assumed only simple predicative comprehension, with the result that in place of $\mathbf{P}^2$, or $\mathbf{P}^3$ with infinity, or $\mathbf{ZF}^2$ with power as well, his approach yields only $\mathbf{P}$, $\mathbf{P}^2$, and $\mathbf{ZF}$.

I have not yet mentioned the approach that has attracted the most attention in recent years, and to which some writers would restrict the label "neo-Fregean", namely, the *abstractionist* strategy of rejecting (4) altogether, and basing the derivation of mathematics on (12) and/or some further or other special cases of (10). The chief proponent of this approach has been Crispin Wright, whose book [34] was the work that ushered in the neo-Fregean era. For the observation of Parsons about Frege's using only (12), without further recourse to (4), in developing arithmetic only became widely known when a sketch of the development of arithmetic in second -order logic plus (12) was presented by Wright, who conjectured the consistency of the system just mentioned, and claimed substantial philosophical advantages for basing arithmetic upon it. Unfortunately, a joint paper of Wright and Neil Tennant, which would have contained a more rigorous formal development, never appeared, while Tennant's own rigorous version was set in a context not calculated to attract a broad readership among those interested in neo-Fregeanism, that of a book-length work on intuitionistic relevance logic [31]. Meanwhile, that the system proposed by Wright was indeed consistent as conjectured was noted independently by several writers [7], [17], [19]. But it was George Boolos [5] who worked out that second-order logic with (12) is of the same interpretability strength as $\mathbf{P}^2$. Collaborators with Wright, notably Bob Hale, have proposed ad hoc additional or alternative assumptions that can boost this to $\mathbf{P}^3$ or beyond. The joint collection of Hale and Wright [16] includes both Hale's technical work on neo-Fregean analysis, as well as much philosophical discussion by both authors.

3 The General Theory of Abstraction

One way to approach the work under review is through consideration of the most pressing objection to abstractionism, the *bad company* objection. The objection is that no cogent reason has been given for assuming (12), when it is known to be surrounded by other abstraction principles, other instances of (10), that are untenable because they lead to contradiction.

A first thought is that one should simply say that only consistent abstraction principles are to be admitted. There are however, two insuperable difficulties with this thought. The first difficulty is that there is no effective procedure for determining whether a given abstraction principle is consistent. This was shown by Heck, as follows. For any Θ, the condition $\Theta \vee X \equiv Y$ defines an equivalence that will agree with the universal equivalence (making all concepts are equivalent) if Θ holds, but will agree with $\equiv$ (making coextensive concepts equivalent) if Θ fails. Hence the corresponding abstraction principle

$$\spadesuit X = \spadesuit Y \leftrightarrow (\Theta \vee X \equiv Y)$$

will be consistent if and only if Θ is, and there is no effective procedure for determining *that* by Church's theorem.

The second difficulty is that two abstraction principles may be each separately consistent, and yet assuming them both jointly may be inconsistent. This was shown by Boolos. A simplified version of his example goes as follows. Call X and Y *almost* coextensive, written $X \equiv_0 Y$, if for all but finitely many objects x, x falls under the one if and only if it falls under the other. Then, like (12), the abstraction principle

$$\heartsuit X = \heartsuit_Y \leftrightarrow X \equiv_0 Y \tag{13}$$

is consistent. But whereas (12) has models with an infinite domain of objects (and in particular one whose domain is $0, 1, 2, \ldots, \aleph_0$), and none with a finite domain (since on a domain with n objects it requires $n + 1$ numbers, 0 through n), by contrast (13) has finite models (where $\equiv_0$ coincides with the universal equivalence) but no infinite ones (where it requires too many abstracts). So while each of (12) and (13) separately has a model, their conjunction has none. (This is a simplification of Boolos's original example.)

A second thought is that, since what goes wrong in the various cases of inconsistency so far is a problem *inflation*, where we have to introduce too many abstracts because an equivalence has too many equivalence classes, we should require that an equivalence first be proved to be noninflationary, to have not-too-many equivalence classes, before we allow ourselves to assume the existence of abstracts for it. Now the assumption that Φ is noninflationary can be expressed as follows:

$$\exists R \forall X \exists x \forall Y (\forall y (Y(y) \leftrightarrow R(x, y)) \rightarrow \Phi(X, Y)) \tag{14}$$

This says that there is a relation R such that for every concept X there is an object x such that X is equivalent to the "section" of R at x, which is to say, the concept Y under which an object falls if and only if that object is R-related to x. This implies that there can be no more equivalence classes of concepts than there are objects, and the converse implication also holds (assuming the axiom of choice). But there is a serious difficulty with requiring (14) to be proved before the existence of abstracts for the equivalence defined by Φ is allowed to be assumed. The difficulty is that it will make it impossible to assume (12) and thus introduce numbers, until it has been somehow proved that there are infinitely many objects, whereas Frege's idea was, and the neo-Fregean idea has always been, to get infinitely many objects from (12).

Even apart from the difficulty noted so far, there is the further difficulty that one cannot assume, even for equivalence relations on *objects* as opposed to equivalence superrelations on *concepts*, and even for those that have only *two* equivalence classes, that abstracts always exist. For the inconsistent naïve theory of sets (7) can be interpreted in a theory that allows unlimited object-abstraction, by identifying $\{x : \Phi(x)\}$ and its complement with the two abstracts for the equivalence given by

$$(\Phi(x) \ \& \ \Phi(y)) \vee ({\sim}\Phi(x) \ \& \ {\sim}\Phi(y)) \tag{15}$$

And a parallel difficulty arises for concept-abstraction, identifying $\ddagger X$ and its complement with the two abstracts for the equivalence given by

$$(X \equiv Z \ \& \ Y \equiv Z) \vee ({\sim}X \equiv Z \ \& \ {\sim}Y \equiv Z) \tag{16}$$

Or at least this is so provided it is assumed that, when dealing with more than one equivalence, the abstracts with respect to distinct equivalences are distinct, at least if the corresponding equivalence classes are. (In the somewhat garbled allusion to the problem in my own review [7] of Wright, I neglected to state this important proviso.) For in set theory, there is a standard way of providing abstracts for equivalences on *objects*, attributed to Dana Scott, that can always be used, namely, taking the abstract of x to be the *truncated* equivalence class of x, the set of all y equivalent to x that are of lowest possible rank. This will, however, in some cases assign x identical abstracts with respect to distinct equivalences, even though the equivalence classes are distinct, since distinct equivalence classes may have identical truncations. But assuming distinctness of abstracts for distinct equivalences, at least when the corresponding equivalence classes are different, one does get an inconsistency, because there are simply too many equivalences, even considering ones with no more than two classes. This problem Fine calls *hyperinflation*.

The main difference between the Scottish School—Wright and associates, working at the Arché Institute at the University of St. Andrews—and the author of the work under review is that the former are content to proceed piecemeal, adding specific abstraction principles one by one, while the latter wishes to develop a *general* theory that will, so to speak, admit all admissible abstraction principles at once. On both approaches there is an awareness of bad company objections, but for the Scottish School it is dealt with by special pleading on behalf of whatever abstraction principle(s) they wish to admit, while Fine's work is essentially devoted to examining the only obvious simple and natural restriction that can prevent hyperinflation. Thus the limits of Fine's approach are likely to be the limits of any form of neo-Fregean abstraction that takes the bad company objection fully seriously.

The key notion for Fine is *invariance*, which is defined as follows. A *permutation* of a domain of objects is a one-to-one function π from that domain onto itself. Such a permutation induces a permutation, by abuse of language also denoted π, on *concepts* (and on *relations*) applying to objects in the domain: x falls under πX if and only πx falls under X (respectively, x and y fall under πR if and only if πx and πy fall under R). A concept X (respectively, relation R) is *invariant* if X is coextensive with πX (respectively, R is coextensive with πR) for all permutations π. A little thought shows that there are only two invariant concepts (the universal and null concepts), and four invariant (two-place) relations (the universal, identity, distinctness, and null relations). In general the number of invariant n-place relations for $n \geq 1$ is $\exp(\exp(n-1)+1)$.

Things will turn out to be much more interesting one level up, so let me set down the relevant definitions. A superconcept $\mathbf{X}$ is invariant if πX falls under $\mathbf{X}$ whenever X does, for all concepts X and permutations π of objects. (Note that it is only those permutations on concepts that are induced by permutations on objects that are being considered. If one allowed arbitrary permutations of concepts, the situation would be the same as it was one level down.) A superrelation $\mathbf{R}$ is invariant if πX and πY fall under $\mathbf{R}$ whenever X and Y do, for all concepts X and Y and all permutations π of objects. A superrelation $\mathbf{R}$ is *doubly* invariant if πX and ρY fall under $\mathbf{R}$ whenever X and Y do, for any concepts X and Y and any permutations π and ρ of objects. These last two notions are of most interest when $\mathbf{R}$ is an equivalence. (Double invariance for an equivalence amounts to the ostensibly weaker requirement that X is always $\mathbf{R}$-equivalent to πX, for any concept X and permutation π.)

In general, a concept, relation, superconcept, or superrelation definable in purely logical terms (without parameters) will be invariant. At the second or super level there will be more invariants than these, but in abstract logic and model theory, invariants are regarded as "logical" in a generalized sense. It is this fact that would make an invariance restriction on abstraction arguably a principled rather than an ad hoc one. (However, even if one grants that logic must not distinguish among objects in general, so that permuting them would make no difference, it is not obvious that an exception should not be made for *logical* objects, if one assumes there are such things.)

We can now introduce a preliminary version of Fine's general theory of abstraction. The theory, when being considered in itself, is naturally formulated as a third-order theory, with axioms being universal statements of the form $\forall\mathbf{R}(-\mathbf{R}-)$. But for purposes of comparison with other approaches, it may be reformulated in a second-order version, replacing such a single axiom by an axiom scheme, according to which for every formula $\Psi(X, Y)$ it is an axiom that $-\Psi(X, Y)-$. The third-order version involves a new predicate which may be written $@(\mathbf{R}, X, u)$ and is intended to mean "u is the abstract of X with respect to $\mathbf{R}$".

A slightly redundant list of axioms may be given as follows. To avoid trivialities, we assume there are at least two objects. Then the first substantive axiom says that if *any* concept has an abstract with respect to $\mathbf{R}$, then *every* concept has an abstract with respect to $\mathbf{R}$.

$$\exists X \exists u @(\mathbf{R}, X, u) \rightarrow \forall X \exists u @(\mathbf{R}, X, u) \tag{17}$$

The next axiom says that the abstracts of two concepts with respect to the same superrelation are identical if and only if the concepts fall under that superrelation.

$$@(\mathbf{R}, X, u) \ \& \ @(\mathbf{R}, Y, v) \rightarrow (u = v \leftrightarrow \mathbf{R}(X, Y)) \tag{18}$$

A little thought shows that this implies that abstracts exist only for equivalences, and that when they exist they are unique. When it exists, the unique abstract u of X with respect to the equivalence $\mathbf{R}$ may be denoted $\S_{\mathbf{R}}X$, and we will then have (10) for any $\mathbf{R}$ for which abstracts exist. The next axiom says that abstracts of an object for different equivalences are distinct, at least unless the corresponding equivalence classes coincide:

$$\S_{\mathbf{R}}X = \S_{\mathbf{S}}X \rightarrow \forall Z(\mathbf{R}(X, Z) \leftrightarrow \mathbf{S}(X, Z)) \tag{19}$$

This might be strengthened in either of two incompatible ways:

$$\S_{\mathbf{R}}X = \S_{\mathbf{S}}X \leftrightarrow \forall Z(\mathbf{S}(X, Z) \leftrightarrow \mathbf{S}(X, Z)) \tag{19a}$$

$$\S_{\mathbf{R}}X = \S_{\mathbf{S}}X \rightarrow \mathbf{R} \equiv \mathbf{S} \tag{19b}$$

The final axiom asserts the existence of abstracts for all doubly invariant, at-most-two-class equivalences.

$$\begin{aligned} &\mathbf{R} \text{ is an equivalence } \& \\ &\mathbf{R} \text{ is doubly invariant } \& \\ &\mathbf{R} \text{ has at most two equivalence classes} \rightarrow \\ &\forall X \exists u @(\mathbf{R}, X, u) \end{aligned} \tag{20}$$

Later we will consider strengthenings of the theory by weakening the second and third hypotheses—the *invariance* and *noninflating* conditions—of (19). But first let us see that even with the present version *third-order arithmetic is interpretable in the general theory of abstraction*.

Proof Let us first note an equivalent version of the general theory of abstraction in the form under consideration. Given any invariant superconcept $\mathbf{X}$, it is easily seen that the following condition defines a doubly invariant, at-most-two-class equivalence $\mathbf{R}_\mathbf{X}$:

$$\mathbf{R}_\mathbf{X}(X, Y) \leftrightarrow ((\mathbf{X}(X) \,\&\, \mathbf{X}(Y)) \vee ({\sim}\mathbf{X}(X) \,\&\, {\sim}\mathbf{X}(Y))) \tag{21}$$

Obviously the complement of $\mathbf{X}$ would give rise to the same equivalence $\mathbf{R}_\mathbf{X}$. Inversely, given any equivalence $\mathbf{R}$ that is doubly invariant and has at most two equivalence classes, it is equally easily seen that the following conditions, wherein U is the universal concept, define complementary invariant superconcepts:

$$\mathbf{X}^+_\mathbf{R}(X) \leftrightarrow \mathbf{R}(U, X) \tag{22a}$$

$$\mathbf{X}^-_\mathbf{R}(X) \leftrightarrow {\sim}\mathbf{R}(U, X) \tag{22b}$$

Assuming the existence of abstracts for doubly invariant, at-most-two-class equivalences is in effect equivalent to assuming the existence of extensions for invariant superconcepts.

Now invariant superconcepts amount to what in abstract logic and model theory are called *generalized quantifiers*. These include, for instance, *none, some, all, at most one, exactly one, at least one, at most two, exactly two, at least two, finitely many, infinitely many, most, evenly many, oddly many,* and so on. The assumption of the existence of logical *objects* corresponding to these quantifiers at once gives us surrogates for natural numbers, in the form of the objects corresponding to *exactly zero, exactly one, exactly two*, and so on. Recognizing these objects corresponds to the historical step, first taken according to some accounts by the Pythagoreans, of switching from the adjectival use of numerals as in "Six disciples conversed with the master", to the nominal use of numerals as in "Six is a perfect number".

Having surrogates for numbers among our objects, concepts applying to them give us surrogates for sets of natural numbers, and we can therefore interpret $\mathbf{P^2}$ in our theory. But actually we can do better, since the quantifier-objects themselves already include surrogates for sets of natural numbers. For instance, the objects corresponding to *evenly many* and *oddly many* can serve as surrogates for the sets of even and of odd numbers. Then using the concepts as surrogates for sets of sets of numbers, we can interpret $\mathbf{P^3}$. □

We may now consider how far the invariance and at-most-two-class conditions in (20) can be relaxed. The first notions we need are the following. Some objects are *few* if they are fewer than (not equinumerous with) all objects, and are *very* few if they are fewer than some objects that are themselves few. A concept Y or a set y is (*very*) *small* if the objects that fall under Y or are elements of y, as the case may be, are (very) few. An equivalence $\mathbf{R}$ is (*very*) *nearly* invariant if there are a (very) few objects such that $\mathbf{R}(\pi X, \pi Y)$ holds whenever $\mathbf{R}(X, Y)$ holds, for all concepts X and Y, and for all permutations π *that leave the given objects fixed*. We can then strengthen (20) to the following:

$$\begin{aligned}&\mathbf{R} \text{ is an equivalence } \ \&\\&\mathbf{R} \text{ is very nearly invariant } \ \&\\&\mathbf{R} \text{ has no more equivalence classes than there are objects} \rightarrow\\&\forall X \exists u @(\mathbf{R}, X, u)\end{aligned} \tag{23}$$

Even with this strengthening, *the general theory of abstraction is consistent*.

Proof In order to prove the consistency of the general theory of abstraction in the version with (23), one constructs a set-theoretic model. Objects will be represented by the elements of some set M. Then concepts will be represented by some subsets of M, two-place relations by some sets of ordered pairs of elements of M, and so on—and change “some” here to “all”, and you get the definition of a *standard* model. To begin with, I assume GCH and show how, for any successor cardinal $\aleph_{\alpha+1}$, to get a standard model with cardinality$(M) = \aleph_{\alpha+1}$, where “(very) small” means $(<)\aleph_\alpha$.

We begin with a series of reductions of the problem. First, to obtain a standard model it will be enough to show that the number of very nearly invariant, not-too-many-class equivalences, and the number of equivalence classes for such equivalences, are $\leq \aleph_{\alpha+1}$. For in that case, the number of pairs consisting of such an equivalence and one of its classes will be $\aleph_{\alpha+1} \cdot \aleph_{\alpha+1} = \aleph_{\alpha+1}$. Then fixing a one-to-one function taking the set of such pairs (or alternatively, simply the equivalence classes) into the set M, we may let the abstract of X with respect to an equivalence be the object of M corresponding under this function to the pair consisting of the equivalence in question and the equivalence class of X with respect to it (respectively, simply to that equivalence class). This will make not only (19) but (19a) (respectively, (19b)) true.

Second, it is actually enough to show that the number of very nearly invariant equivalences $\leq \aleph_{\alpha+1}$. For then, for any such equivalence that has no more equivalence classes than there are objects we will have $\leq \aleph_{\alpha+1}$ equivalence classes, and the *total* number of equivalence classes will be $\leq \aleph_{\alpha+1} \cdot \aleph_{\alpha+1} = \aleph_{\alpha+1}$. Third, it is actually enough to show for that every very small set s the number of s-invariant equivalences is $\leq \aleph_{\alpha+1}$, where s-invariance means invariance with respect to all permutations that fix all elements of s. For every very nearly invariant equivalence is s-invariant for some such s, and GCH implies the total number of such s is $\leq \aleph_{\alpha+1}$. Fourth, it is enough to associate with each s-invariant equivalence a subset, to be called its *signature*, of some set P of cardinality $\leq \aleph_\alpha$, in such a way that distinct equivalences have distinct signatures. For then the number of s-invariant equivalences will be $\leq \exp(\aleph_\alpha)$, which by GCH means $\leq \aleph_{\alpha+1}$.

In pursuit of a suitable definition of signature, let us define for any subsets C and D of M their s-*profile* to be the sextuple consisting of $C \cap s$ and $D \cap s$ and a sequence of four cardinals $\leq \aleph_{\alpha+1}$. Writing M_s as short for the difference $M - s$, the four cardinals in question are to be the cardinalities of the following four disjoint sets, whose union is M_s:

$$C \cap D \cap M_s,\ (C - D) \cap M_s,\ (D - C) \cap M_s,\ M_s - (C \cup D).$$

Let P be the set of all sextuples consisting of two subsets of s and four cardinals $\leq \aleph_\alpha$, so every s-profile is an element of P.

The key observation is that for any two pairs (C, D) and (C', D') that have the same s-profile, there is a permutation π that leaves the elements of s fixed, and has $\pi C = C'$ and $\pi D = D'$. This π is obtained by piecing together the identity function on s, a one-to-one function from $C \cap D \cap M_s$ onto $C' \cap D' \cap M_s$, such as must exist since the two sets have the same cardinality, and similar functions for the obvious further three pairs of sets. The existence of such a π guarantees that, for any given equivalence, if *any* pair C, D with a given s-profile are equivalent, then *every* pair C, D with that s-profile are equivalent. And this means that the equivalence is completely determined by the set of s-profiles for which we *do* have a pair with that profile that are equivalent. This subset of P may be taken as the signature of the equivalence.

To complete the proof it will be enough to show that cardinality$(P) \leq \aleph_\alpha$. To see this, we note on the one hand that since the cardinality of s is $< \aleph_\alpha$, by GCH the number of subsets of s is $\leq \aleph_\alpha$, and on the other hand, that the number of cardinals $\leq \aleph_{\alpha+1}$ is always, even without GCH, $\leq \aleph_\alpha$. The number of sextuples in P is thus $\leq \aleph_\alpha^6 = \aleph_\alpha$.

This completes the proof of the existence of a standard model on a domain of κ objects for any successor cardinal $\kappa = \aleph_{\alpha+1}$, assuming GCH. The proof does not work for limit cardinals, since for such a cardinal κ, either there will be too many very small subsets s, or else there will be too many cardinals $\leq \kappa$. (For the *cognoscenti*, the former problem arises if κ is singular, and the latter if κ is regular.) For $\kappa = \aleph_1$, where "very small" means *finite*, and "small" means *countable*, the foregoing proof gives a standard model assuming only CH.

It is, however, a well-known consequence of Gödel's proof of the interpretabiliy of **ZFC** + CH and **ZFC** + GCH in **ZFC** that any *purely arithmetical* theorem that can be proved to follow from the extra assumption CH or GCH by accepted mathematical means (as codified in **ZFC**), actually can be proved *without* the extra assumption. And the statement that the general theory of abstraction is consistent is reducible, by the kind of coding used in the proof of Gödel's incompleteness theorems, to a purely arithmetical statement. □

A somewhat different-looking though essentially equivalent way of putting the matter would be to say that, even without CH we get a *non*standard model, in which the concept variables, instead of ranging over *arbitrary* subsets of M, will range over *constructible* subsets of M. Looking closely at the proof formulated this last way, it appears that the second-order, schematic version of the general theory of abstraction is interpretable in the theory **ZF***+ V = L, where **ZF*** is **ZF** minus its power set axiom, but with the assumption that at least the power set $\wp(\omega)$ of the set ω of finite ordinals exists, and where V = L is as always the axiom of constructibility. This latter theory itself is interpretable in **ZF***, which in turn is interpretable in $\mathbf{P^3}$. The end result is that the general theory of abstraction appears to have exactly the same interpretability strength as $\mathbf{P^3}$, though the reviewer has not dotted every i nor crossed every t in the proof.

Obviously the assumption of having no more classes than there objects cannot be dropped from the hypothesis of (23). What about the assumption of being very nearly invariant? Can this be weakened to the assumption of being nearly invariant? Unfortunately, if the assumption is thus weakened the proof breaks down, because unless s is *very* small, we cannot be sure the number of subsets $C \cap s$ that appear in profiles will be $\leq \aleph_\alpha$ as needed for the proof.[1] The long and the short of it is, we appear to have reached the "limits of abstraction".

Apart from the material in the two immediately preceding paragraphs, essentially everything in this section has been distilled from the book. The treatment there is considerably lengthier—enough so as to have made the process of distillation seem worthwhile undertaking—for two reasons. First, Fine seeks complete generality in the statement of results. Thus, for instance, where I have considered only the question of the existence of a model, Fine seeks to characterize (without reliance on special assumptions like CH or GCH) *all* cardinals $\aleph_\alpha$ such that there exists a standard model of size $\aleph_\alpha$. Second, he devotes a good deal of attention to two subsidiary topics.

One of these additional topics is the special kind of equivalences he calls *internal*, where the issue whether X and Y are equivalent depends only on what objects fall under X and under Y, and not on what or how many other objects there may be beyond those. For example, being equinumerous is internal, having equinumerous complements is not. Another of the additional topics, not at all unrelated, is how to generate a *minimal* model "from the bottom up", rather than just prove the existence of *some* model "from the top down", as in the foregoing consistency proof. Though this topic has some intrinsic interest, Fine hints that his real motivation for entering into details about the construction is closely bound up with work in progress on what he calls "procedural postulationism", on which he has given public lectures but has not yet at the time of this writing published, and on which it would therefore be premature to comment. Setting details about minimal models and related additional topics aside, therefore, I here conclude my account of the technical results of the book.

What those results suggest is a limit well below higher set theory for what can be obtained by abstraction if one takes the bad company objection fully seriously. An abstractionist who concedes to the author that the objection *ought* to be taken fully seriously, must then either devise some *other* way than through imposing invariance restrictions to deal with the problem of hyperinflation, or else argue that abstractionism offers philosophical benefits worth sacrificing higher set theory to obtain.

But actually, there is some question whether even Fine has taken the bad company objection as seriously as it should be taken. For it would seem that any philosophical account that would justify admitting abstracts of concepts with respect to equinumerosity $\approx$ would equally justify admitting abstracts of relations with respect to isomorphism $\cong$. And certainly the one equivalence is just as invariant as the other. But inconsistency results from admitting abstracts for isomorphism as A. P. Hazen [17] and Harold Hodes [19] both observed, since such abstracts or isomorphism-types include order-types, which include ordinals. And with the ordinals comes a contradiction, the Burali-Forti paradox.

The restriction of Frege's assumption of the existence of abstracts to the case of logical equivalences has some intuitive appeal, as does the construal of logicality as invariance. By contrast, stopping at equivalences that apply to first-level concepts of *one* place only, seems unprincipled. Boolos has a device, plural quantification, that if concepts of any level were dropped—and with them, I suppose, any aspiration to have the resulting theory called "neo-Fregean"—would provide a substitute for the one-place concepts X, and none for the two-place concepts R, so the issue of why to assume the axioms for the former and not the latter would then not arise. But unfortunately, the two-place concepts are crucially needed to express the notion of invariance.

So far as I can see, in the present context there are just two strategies that might be used to defuse the paradox. One would be to motivate independently somehow the assumption of pairs of objects as objects, and then do away with relations, using concepts (or plural quantification) applying to pair-objects. The other would be to argue somehow that the correct two-place analogue of permutation-invariance is not isomorphism-invariance but some much coarser equivalence with far fewer equivalence classes (perhaps one that gives as the only two-place analogues of quantifiers combinations of one-place quantifiers, such as *for every x there is some y, for most x there are exactly two y, for some x there are infinitely many y*, and so on). But it is not at all clear that either strategy can be made to work.

The Burali-Forti paradox was historically the very first of the set-theoretic paradoxes to become public, years before the Russell paradox. This paradox—or rather, the failure to deal adequately with it—remains the skeleton in the abstractionist closet. Fine mentions it at the very end of his book as another area where more work is needed. It will be very interesting to see what future work of Fine and others will make of it.

4 Purely Philosophical Issues

It remains to consider the alleged philosophical benefits of abstractionism, for which sacrifices may be required. Now the underlying assumption of discussions about the philosophical benefits of introducing natural numbers—I mean introducing them into *discourse*, not necessarily introducing them into *reality*—in one way rather than in another, is that even when two theories are interpretable in each other, and to that extent from a mathematical point of view equivalent, from a philosophical point of view it may be preferable to regard one of them as *primary* and the other as *derivative*, with the former being justified directly, and the latter justified by its interpretability in the former.

In this connection the first thing to note about Fine is that, unlike the Scottish School, he is not especially concerned to *advocate* the abstractionist way of introducing numbers. His article began with the words, "This paper has been written more from a sense of curiosity than commitment," and the same disclaimer, with "present monograph" substituted for "paper", is the opening line of the introduction to the book version. Fine's concern is more to evaluate claims Wright and others have made about the supposed benefits of introducing numbers as abstracts than to make any such claims himself.

There are two problems, especially, that Wright in his book claimed abstractionism could help solve. Each problem was in a sense raised by Frege, and each has also been made the topic of an influential paper by Paul Benacerraf. We begin with Frege's question, "How do we know that the number two is not Julius Caesar?" or "How do we know that Julius Caesar is not a number?" To this Fine adds "And how do we know that the number two is not a Roman?" or "How do we know that no Roman is a number?" Frege held that introducing numbers as equivalence classes, rather than simply as abstract, which is to say, identifying a number with a set

$$\#X = \{Y : X \approx Y\} \tag{24}$$

and then demonstrating (12), as opposed to taking #X as a previously undefined term, and introducing it by an undemonstrated assumption of (12), had the advantage of providing an answer to the Julius Caesar problem. Most of Frege's readers, however, have had difficulty seeing why this step doesn't just push the problem back a stage, leaving us with the question, "How do we know that Julius Caesar is not a set?"

But be that as it may, there is a problem with the identification (24) itself. For many of Frege's readers have had difficulty believing that a number could turn out to be a set. Russell, in the period before he discovered his paradox, independently rediscovered several of Frege's ideas, and among them the idea of identifying a number with a set. But where Frege identified a number with the extension of a *super*concept, making the number two the set of all concepts under which exactly two objects fall, Russell identified a number with the extension of a *concept*, making the number two the set of all sets in which there are exactly two objects as elements. This difference in identifications between two independent investigators counts against both identifications, each of which is in any case implausible, as Russell, at least, if not Frege, acknowledged, lamely excusing himself by saying, "It is . . . more prudent to content ourselves with the class of couples, which we are sure of, than to hunt for a problematical number 2 which must always remain elusive" ([28], p. 172). The artificiality is more conspicuous with later set-theoretic identifications of number, like Zermelo's, which makes the number two to be the singleton of the singleton of the empty set {{{}}}, or von Neumann's, which makes it to be the unordered pair of the empty set and its singleton {{}, {{}}}. Those definitions provoked a classic paper of Paul Benacerraf [2] urging that whatever numbers may be, they are not *sets*. (While Benacerraf starts with von Neumann and Zermelo, his remarks are clearly meant ot apply to Russell and Frege as well.)

Since that paper philosophers of mathematics have pretty generally turned against any such identification, and in that sense have rejected the question, "What are numbers?" There are, however, three forms this rejection may take. *Nominalists* reject the question's existence presupposition, that there *are* such things as numbers. Fine ignores them, and I am prepared, in the present context, to join him in doing so. *Structuralists* reject the question's uniqueness presupposition, that there is such a thing as *the* number zero in *the* system of natural numbers. According to structuralism, an assertion about "the natural number system" and "the number zero" is to be understood as a generalization about *any* system of objects that is ordered in an ω-sequence, and about the initial object in *any* such system. On this view, there is no more such a thing as "the" number zero in "the" natural number system than there is such a thing as "the" identity element in "the" group of order two. Fine dismisses structuralism with a brief critical remark, and I am prepared, in the present context, to concur in its dismissal.

That leaves a third way to reject the question, namely, by answering, "Everything is what it is, and not another thing." In other words, one dismisses the question "What are numbers?" by answering that numbers are objects *sui generis*. Underlying this position is the assumption, diametrically opposed to Frege's, that the category N is divided into any number of sorts N_i, and that an identity $a = b$ involving two terms of category N is automatically accounted false (if not ill-formed and without truth-value), unless the terms are of the same sort. (As Fine points out, some provision will have to be made in a more refined statement of this position for the use of *non*sortal terms like "What Paul is thinking of".) And it would seem that the grammatical point about sortal divisions within the category N is all that is needed to provide an answer to the Julius Caesar problem (and the Roman problem).

This leaves us, however, with the problem, "How do we *know* that 'two' and 'Julius Caesar' are expressions of type N of different sorts?" Wright's suggestion, to put it in my own words, and very roughly, is that different sorts come with different *identity* criteria for the objects denoted by expressions of type N of those sorts, and the fact that the identity criteria provided by (12) for numbers are obviously very different for those for human beings is what indicates that the terms "two" and "Julius Caesar" are of different sorts.

But supposing we accept that objects introduced as abstracts in general, and numbers as introduced by (12) in particular, are not to be identified with any objects *not* introduced as abstracts, there still remains the question how we know that numbers as introduced by (12) are not to be identified with any abstracts introduced by any *other* abstraction principle? Or *do* we in fact know that much? Fine's formal system certainly assumes that abstracts with respect to different equivalences are different if the corresponding equivalence classes are different. But what about cases where the equivalence classes are the same? The formal system is equally compatible with assuming, as per (19a), that in such cases, too, the abstracts are still always different, and with assuming, as per (19b), that in such cases, rather, they are always the same. It is also compatible with assuming that they are sometimes the same, and sometimes different.

Fine produces one example where one may be strongly tempted to identify abstracts with respect to distinct equivalences. Do we *really* want to distinguish the number two as introduced by (12), involving $\approx$, from the "number two" as introduced by a variant (12*), involving a variant $\approx^*$, which holds between two concepts when they are both finite and are equinumerous, or are both infinite? That is to say, do we *really* want to distinguish the number two that is part of the system of cardinal numbers, finite and transfinite, from the number two that was part of the pre-Cantorian system, with just one infinity ∞ rather than a series of transfinite $\aleph$s? It would perhaps not be a fatal objection to the *sui generis* theory to have to concede that two grammatical categories N_i and N_j may sometimes overlap, but the first line of response would surely be to question whether we must identify the two twos. After all, we *do* distinguish the natural number 2 from the positive integer $+2$ and the rational number $+2/1$ and the real number $2.000\ldots$ and the complex number $2.000\ldots + 0,000\ldots i$ and so on. Or at least, logicians make these distinctions, as do symbolic computation programs such as Mathematica. So why not make a distinction in the other case as well? But indeed, Fine does not insist that we *must* identify the two twos. Merely the fact that we can *raise the question* whether we should, and that no technical results push us one way or the other suggests that a complete solution to the Julius Caesar problem has not yet been found.

Another and much larger problem is one made prominent by *another* paper of Benacerraf [3]. This is the problem of how we can have *knowledge* of numbers, which was already of concern to Frege in the form of the question, "How can we apprehend logical objects?" One reason Wright's book made as great an impression as it did is that it suggested a solution. Again in my own words, and again very roughly, the proposed answer is that the objecthood of numbers consists in the singular termhood of numerals, or more precisely, of expressions of the form $\#X$, while the apprehension of them consists in the apprehension of the truth of statements involving such terms.

And this last is no more problematic than the apprehension of the truth of certain statements *without* such terms, namely, the statements about concepts and equinumerosity to which the statements *with* such terms are equivalent by (12). Such was, to a first approximation, Wright's interpretation of Frege's "context principle" to the effect that it only makes sense to ask after the meaning of an expression in the context of a sentence.

At one extreme, no nominalist will think Wright's answer sufficient. An account of numbers giving a central role to their introduction through (12) will, for the nominalist, only be an account of the introduction of numerals, and of terms of the form "the number of". The nominalist will never admit that the numbers themselves have been apprehended through the introduction of numerals and other terms unless it can be shown that the numbers themselves *caused* the introduction of those linguistic items. At the other extreme, no pragmatist will think Wright's answer necessary. Wright may or may not be correct about how number terms were or could have been or could be introduced, but for the pragmatist, *however* they were introduced, since positing them has proved useful in the development of commonsense and scientific theories with practical applications, no further justification of their retention is needed. Even if their genealogy proves to be quite sinister, that is irrelevant from a pragmatist point of view. The fact that they *now* are governed by a definite usage and have shown a definite utility is enough to legitimize them, be their origins however base.

Debates over the context principle take place in the space between the two extremes of nominalism and pragmatism. What tends to make such debates inconclusive is that the anti-nominalist and anti-pragmatist presuppositions of the debaters are often not made explicit. In reading Fine's long discussion of the context principle, the reviewer throughout had the feeling that each point made would be telling against one or another class of opponents, with one or another kind of presupposition, but that it would be hard to locate a philosophical position from which *all* the points made would seem pertinent. This is not to say that Fine ever claimed otherwise.

And be all that as it may, a confirmed pragmatist like the reviewer is perhaps not the best person to attempt an evaluation of Fine's contributions to the on-going—I almost wrote "interminable"—debate on this topic. Neither is a journal devoted to formal logic the best place for such an attempted assessment. Moreover, even if I were the person and this were the place, this is not the time. For as Fine himself hints both in his preface and at the end of chapter 2, the real point of Fine's critique of Wright and associates on the context principle cannot be expected to become wholly clear until such time as his own, rival positive views on "procedural postulation" are made available in print. Let us hope that that time will not be long in coming.

Note

1. On this point, as observed by the reviewer and acknowledged by the author in correspondence, there is a misstatement in one lemma in the text (Theorem 5 p. 159 in the book, p. 598 in the article), and this affects a number of propositions that depend on this lemma, which need to be amended. In the terminology used here, the amendment would consist essentially just in inserting "very" appropriately.

References

[1] Benacerraf, P., and H. Putnam, editors, *Philosophy of Mathematics. Selected Readings*, 2d edition, Cambridge University Press, Cambridge, 1983. MR 85h:00015. 249

[2] Benacerraf, P., "What numbers could not be," *Philosophical Review*, vol. 74 (1965), pp. 47–73. Reprinted in [1], pp. 272–94. MR 30:1929. 246

[3] Benacerraf, P., "Mathematical truth," *Journal of Philosophy*, vol. 70 (1973), pp. 661–79. Reprinted in [1], pp. 403–20. MR 58:4979. 247

[4] Boolos, G., "Iteration again," *Philosophical Topics*, vol. 17 (1989), pp. 5–21. Reprinted in [6], pp. 88–104. 231, 236

[5] Boolos, G., "The consistency of Frege's *Foundations of Arithmetic* (in *On Being and Saying: Essays in Honor of Richard Cartwright*, pp. 3–20, The MIT Press, Cambridge, 1987)," pp. 211–33 in *Frege's Philosophy of Mathematics*, Harvard University Press, Cambridge, 1995. Reprinted in [6], pp. 183–201. Zbl 0900.03062. MR 1376397. 237

[6] Boolos, G., *Logic, Logic, and Logic*, edited by R. Jeffrey, Harvard University Press, Cambridge, 1998. Zbl 0955.03008. MR 2000b:03005. 249

[7] Burgess, J. P., "Review of *Frege's Conception of Numbers as Objects*," *Philosophical Review*, vol. 93 (1984), pp. 638–40. 237, 239

[8] Burgess, J. P., and A. P. Hazen, "Predicative logic and formal arithmetic," *Notre Dame Journal of Formal Logic*, vol. 39 (1998), pp. 1–17. Zbl 0967.03048. MR 2000e:03019. 235

[9] Collins, G. E., and J. D. Halpern, "On the interpretability of arithmetic in set theory," *Notre Dame Journal of Formal Logic*, vol. 11 (1970), pp. 477–83. Zbl 0185.01601. MR 45:4970. 235

[10] Demopoulos, W., editor, *Frege's Philosophy of Mathematics*, Harvard University Press, Cambridge, 1995. Zbl 0915.03004. MR 96h:03015. 250

[11] Fernandes, A. M., and F. Ferreira, "Groundwork for weak analysis," *The Journal of Symbolic Logic*, vol. 67 (2002), pp. 557–78. Zbl 1015.03056. MR 2003b:03084. 234

[12] Ferreira, F., and K. F. Wehmeier, "On the consistency of the Δ^1_1-CA fragment of Frege's *Grundgesetze*," *Journal of Philosophical Logic*, vol. 31 (2002), pp. 301–11. Zbl 1012.03059. MR 2003e:03115. 236

[13] Fine, K., "The limits of abstraction," pp. 503–629 in *The Philosophy of Mathematics Today (Munich, 1993)*, edited by M. Schirn, Oxford University Press, New York, 1998. Zbl 0922.03011. MR 2000g:03009. 227

[14] Friedman, H. M., "Higher set theory and mathematical practice," *Annals of Mathematical Logic*, vol. 2 (1970/1971), pp. 325–57. Zbl 0215.32702. MR 44:1556. 234

[15] Hájek, P., and P. Pudlák, *Metamathematics of First-order Arithmetic*, Perspectives in Mathematical Logic. Springer-Verlag, Berlin, 1998. Zbl 0889.03053. MR 2000m:03003. 234

[16] Hale, B., and C. Wright, *The Reason's Proper Study: Essays towards a Neo-Fregean Philosophy of Mathematics*, The Clarendon Press, Oxford, 2001. Zbl 1005.03006. MR 2037746. 237

[17] Hazen, A., "Review of *Frege's Conception of Numbers as Objects*," *Australasian Journal of Philosophy*, vol. 63 (1985), pp. 251–54. 237, 244

[18] Heck, R. G., Jr., "The consistency of predicative fragments of Frege's *Grundgesetze der Arithmetik*," *History and Philosophy of Logic*, vol. 17 (1996), pp. 209–20. Zbl 0876.03032. MR 98i:03079. 235

[19] Hodes, H. T., "Logicism and the ontological commitments of arithmetic," *The Journal of Philosophy*, vol. 81 (1984), pp. 123–49. MR 88d:03005. 237, 244

[20] Martin, D. A., "Review of W. V. Quine, *Set Theory and Its Logic*," *Journal of Philosophy*, vol. 67 (1970), pp. 111–14. 231

[21] Montagna, F., and A. Mancini, "A minimal predicative set theory," *Notre Dame Journal of Formal Logic*, vol. 35 (1994), pp. 186–203. Zbl 0816.03023. MR 95h:03117. 235

[22] Parsons, C., "Frege's theory of number," pp. 180–203 in *Philosophy in America*, edited by M. Black, Cornell University Press, Ithaca, 1965. Reprinted with a postscript in his *Mathematics in Philosophy*, Cornell University Press, Ithaca, 1983 and in [23], pp. 182–210. 230, 250

[23] Parsons, C., "Frege's theory of number," pp. 182–207 in *Frege's Philosophy of Mathematics*, edited by W. Demopoulos, Harvard University Press, Cambridge, 1995. Originally published in [22]. Zbl 0915.03004. MR 1376396. 250

[24] Parsons, T., "On the consistency of the first-order portion of Frege's logical system," *Notre Dame Journal of Formal Logic*, vol. 28 (1987), pp. 161–68. Reprinted in [10], pp. 422–31. Zbl 0637.03005. MR 88h:03002. 236

[25] Quine, W. V., "Reply to Martin," *Journal of Philosophy*, vol. 67 (1970), pp. 247–48. 231

[26] Quine, W. V. O., *Mathematical Logic*, Harvard University Press, Cambridge, 1951. Revised edition. Zbl 0579.03001. MR 13,613f. 236

[27] Russell, B., "My mental development," pp. 37–50 in *The Basic Writings of Bertrand Russell*, edited by R. E. Egner and L. E. Denonn, Simon & Schuster, New York, 1961. 231

[28] Russell, B., "*Introduction to Mathematical Philosophy* (selections)," pp. 160–82 in *Philosophy of Mathematics*, Cambridge University Press, Cambridge, 1983. Zbl 0548.03002. MR 85h:00015. 246

[29] Russell, B., *Collected Papers vol. 4: Foundations of Logic 1903–1905*, edited by A. Urquhart, Routledge, London, 1994. 236

[30] Tarski, A., *Undecidable Theories*, Studies in Logic and the Foundations of Mathematics. North-Holland Publishing Company, 1953. In collaboration with A. Mostowski and R. M. Robinson. Zbl 0053.00401. MR 15,384h. 235

[31] Tennant, N., *Anti-Realism in Logic: Truth as Eternal*, Oxford University Press, Oxford, 1987. 237

[32] Urquhart, A., "Russell's zig-zag path to the ramified theory of types," *Russell*, vol. 8 (1988), pp. 82–91. 236

[33] Wehmeier, K. F., "Consistent fragments of *Grundgesetze* and the existence of non-logical objects," *Synthese*, vol. 121 (1999), pp. 309–28. Zbl 0952.03068. MR 2001f:03114. 236

[34] Wright, C., *Frege's Conception of Numbers as Objects*, vol. 2 of *Scots Philosophical Monograph Series*, Aberdeen University Press, Aberdeen, 1983. Zbl 0524.03005. MR 85g:00035. 237

Acknowledgments

The author is grateful to Kit Fine, Harvey Friedman, Richard Heck, Øystein Linnebo, and Charles Parsons for comments and information, and especially to Neil Tennant (for numerous comments on and corrections to an earlier draft) and A. P. Hazen (for historical information on the Szmielew-Tarski theorem and on Wang's revision of Quine's inconsistent system, as well as other help).

John P. Burgess
Department of Philosophy
Princeton University
1879 Hall
Princeton NJ 08544-1006
jburgess@princeton.edu