Which Abstraction Principles are Acceptable? A Limitative Result

Øystein Linnebo and Gabriel Uzquiano

March 19, 2008

1 Introduction

Neo-Fregean logicism attempts to base mathematics on *abstraction principles*, that is, on principles of the form:

$$(*) \qquad \S\alpha = \S\beta \leftrightarrow \alpha \sim \beta,$$

where the variables 'α' and 'β' range over items of a certain sort, '§' is operator taking items of this sort to objects, and '$\sim$' expresses an equivalence relation on this sort of item. At the heart of the neo-Fregean programme lies, for example, the claim that Hume's principle is acceptable as an implicit definition of number:

$$\text{(HP)} \qquad \#F = \#G \leftrightarrow F \sim G,$$

where $F \sim G$ states that the concepts F and G are equinumerous. The neo-Fregean project receives encouragement and support from Frege's Theorem, which states that the axioms of second-order Peano Arithmetic are derivable from Hume's principle and some natural definitions in the framework of second-order logic.

However, for the neo-Fregean programme to be successful, this theorem must be accompanied by an account of implicit definition which explains why Hume's principle is an acceptable implicit definition and why other abstraction principles, such as Frege's inconsistent Basic Law V, are not. In this connection the so-called "bad company problem" arises as the challenge of specifying necessary and sufficient conditions for an abstraction principle to be acceptable as an implicit definition. That is, the neo-Fregeans need to provide the means to discriminate acceptable abstraction principles—from which to derive arithmetic, real analysis and set theory—from unacceptable abstraction principles. However, the quest for necessary and sufficient conditions for acceptability has been tortuous in view of the successive failure of various initially attractive candidates.

In this paper we will be concerned with one of the most promising responses to this problem, which has been quite influential in the recent literature.[1] Say that an abstraction principle Σ is *stable* if and only if there is some cardinal number κ such that Σ is satisfiable in every domain of cardinality greater than or equal to κ. The response in question is that an abstraction principle is acceptable just in case it is stable. This response relies on what Stewart Shapiro in (Shapiro, 2000) calls *the external perspective*, namely the perspective of a bystander who is interested in the prospects of the neo-Fregean project against the background of standard mathematics. We will make free use of Zermelo-Fraenkel set theory with the axiom of choice (ZFC) in the study of various model-theoretic criteria for acceptability. Unless otherwise stated, ZFC will be the metatheory assumed in this paper.

To the best of our knowledge, stability first emerged as a candidate solution to the bad company problem in (Heck, 1992) (494, fn. 5.) But the proposal has only recently gained more currency in the literature. For instance, Bob Hale and Crispin Wright cite a closely related condition as a promising necessary and sufficient condition for acceptability in (Hale and Wright, 2001a). Whatever virtues are identified as the mark of acceptability, they suggest that an acceptable abstraction should be *unmatched* in the sense that it is compatible with all other abstractions with exactly the same virtues. And they carry on to say:

> We conjecture that the unmatched (logical) abstractions will coincide with those

[1]See, for example, (Hale and Wright, 2001a) and (Weir, 2003).

which Kit Fine (1998) called *stable*: abstractions which are satisfiable and such that, if satisfiable at a domain of cardinality κ, are so at every cardinal greater than κ. ((Hale and Wright, 2001a), 427, fn. 14.)[2]

Stability is also interesting in connection with another proposed condition for an abstraction principle to be acceptable. Say that an abstraction principle is *irenic* if and only if it is conservative and jointly satisfiable with any other conservative abstraction principle.[3] Alan Weir has recently shown in (Weir, 2003) that an abstraction principle is irenic if and only if it is stable. So any counterexample to stability as a criterion for acceptability will also serve as a counterexample to irenicity as such a criterion.

Finally, stability is attractive in its own right as a condition for acceptability. A fatal problem with earlier model-theoretic conditions for acceptability has been that there are pairs of abstraction principles which, on the one hand, meet the relevant condition but, on the other hand, fail to be jointly satisfiable in any domain. Since the two abstraction principles are not jointly satisfiable, at least one of them must be unacceptable.[4] But this means that the proposed model-theoretic condition fails to be sufficient for acceptability. Stability represents an apparent advantage in this regard. To see this, say that a family $\mathcal{F}$ of abstraction principles is stable if and only if there is some κ such that every member Σ of $\mathcal{F}$ is satisfiable in every domain of cardinality greater than or equal to κ. Unlike other model-theoretic conditions investigated in the literature, stability ensures that every set of stable principles will itself be stable (and thus *a fortiori* satisfiable), as the following remark makes clear.

Remark 1 Every set $\mathcal{F}$ of stable abstraction principles is itself stable.[5]

[2] In fact, this condition is slightly stronger than what we have called stability because it allows no "gaps" among the cardinalities at which an abstraction is satisfied. For instance, if an abstraction Σ is satisfiable only on $\aleph_0$ and every cardinal greater than $\aleph_4$, then Σ is stable by our definition but fails to meet the requirement in the passage. Thanks to Stewart Shapiro.

[3] A conservative abstraction principle is, very roughly, one which has no new (semantic) consequences for any objects other than the abstracts being introduced. Henceforth, logical notions such as consequence and conservativeness will always be understood in the model-theoretic rather than proof-theoretic sense.

[4] We here rely on an assumption which is standard in neo-Fregean approaches, namely that abstraction principles are supposed to be true in a single, universal domain of objects. Although one may consider lifting this assumption (see (Weir, 2003)), doing so would result in a very different project.

[5] To see this, let $\mathcal{F}$ be a set of stable abstraction principles. Let $\phi(x, \kappa)$ express that κ is the smallest

However, (Uzquiano, 2007) has recently called attention to a serious cost attached to stability as a necessary condition for acceptability. He observes that no stable abstraction principle can hope to recover standard (second-order) set theory when this is taken to include an axiom of infinity and a weak form of replacement. Despite this limitation, there still seems to be good reason for optimism about the prospects for stability as a *sufficient* condition for acceptability. Our present aim is to show that this optimism is misplaced and that stability in fact fails to provide a sufficient condition for acceptability.

We will remain neutral on the difficult question whether the mere acceptability of an abstraction is sufficient for it to be a legitimate implicit definition or whether, alternatively, a certain epistemic burden has to be discharged. One might for instance require that the abstraction be *known* to be acceptable for it to be legitimate to employ it as an implicit definition.[6] This constraint would obviously be far from trivial. But as it happens, our counterexamples to stability as a sufficient condition for acceptability are not epistemic in character and work even when stability is known to obtain.

We will proceed as follows. We will first exhibit a very simple proper class of stable principles which is not satisfiable. Modulo standard set-theoretic assumptions, this example spells trouble for what Kit Fine calls *compromising* neo-Fregeans, that is, for neo-Fregeans who are prepared to take set theory at face value and seek to provide a neo-Fregean foundation for it, but who would also be prepared to accept an alternative foundation.[7] By contrast, our example will not move an *uncompromising* neo-Fregean, who rejects any mathematical theory until it has been recaptured on the basis of abstraction principles. Such a theorist will reject our example as relying on assumptions that remain unsupported, given that no neo-Fregean foundation for standard set theory has yet been developed. In order to deal with her, we will go on to present a somewhat more complicated proper class of stable principles which is inconsistent with assumptions which can be independently

---

cardinal such that the abstraction principle x is always satisfiable in domains of cardinality $\geq \kappa$. Using Replacement, we can find a set C of the cardinal numbers that are thus associated with abstraction principles from $\mathcal{F}$. Let λ be the union of C. Then all abstraction principles from $\mathcal{F}$ are satisfiable in domains of cardinality $\geq \lambda$.

[6] The issue of whether a candidate condition for acceptability must be known to obtain has been raised and discussed by (Ebert and Shapiro, 2007).

[7] See, for example, (Fine, 2002), pp. 10-15.

motivated on purely neo-Fregean grounds.

2 A simple counterexample

Are there stable abstraction principles which are nevertheless unsatisfiable (and thus unacceptable)? What would an example of this phenomenon look like? Given the technical result that any set of stable abstraction principles is itself stable (and thus also satisfiable), we would need a proper class of stable abstraction principles. One may be tempted to dismiss this possibility out of hand on the grounds that the specification of a proper class of abstraction principles would require a proper class of formulas, and no reasonable extension of our language would be able to accommodate *that* many formulas.

An abstraction principle of the form:

$$(*) \qquad \S\alpha = \S\beta \leftrightarrow \alpha \sim \beta,$$

is, for the neo-Fregean, a stipulation that a certain equivalence relation $\sim$ on items of the sort of α and β provides a criterion of identity of §-abstracts. On this conception of abstraction principles, it is completely immaterial whether the equivalence relation is expressible by a closed formula of the language in question. What matters is the equivalence relation itself, not how it is expressed. So there is nothing that prevents us from using free variables—or parameters as we like to call them—to specify a proper class of equivalence relations. This will allow one to lay down a proper class of stipulations corresponding to a parametrized family of abstractions.

We will now implement this thought to describe a proper class of abstraction principles which individually are stable but which are not jointly satisfiable in any set-sized model. We proceed in two steps. In the first step we identify a proper class of equivalence relations. The second step generates a parametrized family of abstraction principles corresponding to them, which fit the bill by being individually stable but not jointly satisfiable in any set-sized model.

For the first step, consider for each cardinal $\kappa > 0$ the equivalence relation $R_\kappa(F, G)$

which a concept F bears to another concept G just in case there are at least κ objects in the domain or F and G are coextensive. In a domain of cardinality $< \kappa$, no two concepts stand in the relation R_κ unless they are coextensive. But in a domain of size $\geq \kappa$, any two concepts are related by R_κ, which generates a trivial equivalence class of all concepts.

If we can specify a proper class of abstraction principles $\mathcal{A}_\kappa$ associated with each R_κ, then we will have a proper class of stable abstractions principles which are individually but not jointly satisfiable in a set-sized domain. For if $\mathcal{A}_\kappa$ is designed to generate abstracts on the basis of R_κ, then $\mathcal{A}_\kappa$ will collapse into Basic Law V when interpreted in a domain of cardinality $< \kappa$. However, when interpreted in a domain of cardinality $\geq \kappa$, $\mathcal{A}_\kappa$ will generate a single abstract and thus become trivially satisfiable. So $\mathcal{A}_\kappa$ is only satisfiable at a model of cardinality $\geq \kappa$. It follows that proper-class-many abstractions in the family fail to be satisfiable at any set-sized domain D. For if the cardinality of this domain is κ, then any $\mathcal{A}_\lambda$ where $\lambda > \kappa$ will fail to be satisfiable at D.

The second step consists in making use of parameters in order to specify the family of abstractions $\mathcal{A}_\kappa$ associated to each R_κ. To that purpose, we use a free third-order variable $\mathbb{X}$, which acts as a parameter allowed to range over second-level concepts.[8] In effect, the parameter acts as a predicate of concepts. The family of abstraction principles we are after can now be specified relative to assignments of appropriate second-level concepts to the free variable $\mathbb{X}$ below:

$$\S_{\mathbb{X}}F = \S_{\mathbb{X}}G \leftrightarrow [\exists Y\ \mathbb{X}(Y) \vee \forall x(Fx \leftrightarrow Gx)].$$

In particular, consider, for each κ, the assignment to '$\mathbb{X}$' of the second-level concept $\mathbb{C}$ under which a concept F falls if and only if F has cardinality greater than or equal to κ.

A word of clarification is in order. We first interpret each abstraction with respect to the entire universe. That is, the free variable '$\mathbb{X}$' is assigned, prior to considering any set-sized models, a second-level concept $\mathbb{C}$ under which all and only first-level concepts greater than or equal to κ in size fall. When interpreted with respect to a set-sized domain, '$\mathbb{X}X$'

[8] A first-level concept is one under which objects fall; a second-level concept is a concept under which first-level concepts fall; and an $n+1$th-level concept is one under which nth-level concepts fall.

is true relative to an assignment of a concept, i.e., a subset S of the domain, to the variable 'X' if and only if S has cardinality greater or equal to κ. So the parameter functions in effect as a cardinality predicate which applies to a concept, i.e., a subset of the domain, just in case this subset has a certain cardinality absolutely.

We see only one way in which a compromising neo-Fregean may attempt to object. For we have assumed that for a family of abstraction principles to be acceptable, they must be satisfiable—in the usual model-theoretic sense of being satisfied in a *set sized* domain. But techniques are available for talking about satisfaction in proper class sized domains.[9]

By contrast, our example will not necessarily move an uncompromising neo-Fregean. For we have made essential use of external resources provided by our background theory of ZFC. In particular, we have assumed the existence of a concept of size κ for each cardinal κ. But this may be called into question by an uncompromising neo-Fregean who is prepared to suspend judgment about the axioms of ZFC until a neo-Fregean foundation for ZFC is in place. We won't here take a stand on whether either of these possible objections can be spelled out and defended but will leave this as a challenge to neo-Fregeans.[10]

3 A Fregean counterexample

We will now provide a different counterexample to stability as a sufficient condition for acceptability which is immune to the two worries just mentioned and which relies only on assumptions that are acceptable to the neo-Fregeans themselves. This counterexample is based on a family of abstraction principles first studied by (Fine, 2002), who uses it as a counterexample to the idea that an abstraction principle is acceptable provided it

[9] See for example (Rayo and Uzquiano, 1999).

[10] An interesting alternative to the simple counterexample developed in this section was suggested to us by Stewart Shapiro. Consider the generalized cut abstraction (GCA) studied by (Cook, 2002) (see p. 58). Remove the initial universal quantifiers. Replace the pair $\{H, <\}$ with a single parameter R ranging over set sized linear orders. The ensuing parametrized family of abstraction principles can be shown to provide another counterexample with the following two advantages over the one described in this section. It avoids the need for a third-order parameter, making do with a dyadic second-order one. And it is based on abstractions that are closely related to ones in which the neo-Fregeans have themselves been interested. However, this alternative counterexample has the disadvantage of requiring significantly more mathematics to be developed properly. And it remains just as vulnerable to the two objections just outlined.

is *non-inflationary*, in the sense that it is satisfiable at the actual universe. We claim that this family also provides a counterexample to stability as a sufficient condition for acceptability. Since our target differs from Fine's, our argument will differ accordingly. In particular, unlike Fine, we will need to argue that the example provides us with a family of individually stable abstractions which are nevertheless not jointly satisfiable. Moreover, once we do this, we will see how similar counterexamples to the sufficiency of stability for acceptability are ubiquitous once we allow for parametrized families of abstractions.

Consider the equivalence relation $R_X(F, G)$ which holds between two concepts F and G just in case both coincide with X on the domain in question or neither does. The equivalence relation $R_X(F, G)$ is defined by the following formula:

(Def-R_X) $$\forall x[Xx \leftrightarrow Fx \wedge Xx \leftrightarrow Gx] \vee [\neg\forall x(Xx \leftrightarrow Fx) \wedge \neg\forall x(Xx \leftrightarrow Gx)].$$

Whenever a concept C is assigned to the free variable 'X', the formula 'Xx' will be defined for any assignment to the variable 'x' of some object o simply by letting the formula be true just in case o falls under C. (Note that it does not matter what domain D the object o is taken from.) The free variable 'X' thus functions just like a meaningful predicate constant, except that it is allowed to vary. Free variables that function in this way are what we in Section 2 called *parameters*.

Consider now the family of abstraction principle of the form:

($*_X$) $$\S_X F = \S_X G \leftrightarrow R_X(F, G).$$

We claim that this family provides a counterexample to the claim that stability is sufficient for acceptability. Our argument can be broken into the following distinct steps, each of which will be defended in the next section.

Step 1. For any concept X, the abstraction principle ($*_X$) is stable. For each such principle is satisfiable in any model that contains two or more objects.

Step 2. The following generalized abstraction is true:

$$\text{(GA)} \qquad \forall X(\S_X F = \S_X G \leftrightarrow R_X(F, G))$$

We rely here on Step 1 and our assumption that stability is a sufficient condition for acceptability.

Step 3. Consider two equivalence relations **R** and **S** on which abstraction is permissible. Then a necessary condition for the **R**-abstract of F to be identical to the **S**-abstract of G is that the equivalence class of F under **R** is identical to the equivalence class of G under **S**.

Step 4. When this necessary condition is applied to the family of abstraction principles $(*_X)$, we get the following principle:

$$\text{(1)} \qquad \S_X F = \S_Y G \rightarrow \forall H(R_X(F, H) \leftrightarrow R_Y(G, H))$$

But this principle is inconsistent, as will be proved in Section 4.4. The thought on which the proof is based is easily described. It follows from (1) that whenever two concepts F and G are distinct, then so are the abstracts $\S_F F$ and $\S_G G$. Such abstracts thus behave much like extensions of concepts. This enables us to exploit the reasoning of Russell's paradox to derive an explicit contradiction.

4 The argument

The above argument will now be spelled out in greater detail.

4.1 Defending Step 1

The only real worry concerning this step is that the abstraction principles $(*_X)$ contain a parameter X. Perhaps the neo-Fregans can object that they are only interested in abstraction principles that are specifiable by closed formulas, not ones whose specifications make use of free variables (or "parameters" as we are also calling them).

But this objection would be desperate and unmotivated. The entire literature on the neo-Fregean approach relies on the definition of an abstraction principle with which we began. And according to this definition, an abstraction principle is any principle that assigns objects to items of a certain sort in such as way as to respect some equivalence relation $\sim$ on these items. To add now that $\sim$ must somehow be specifiable by a closed formula of a higher-order language would be ad hoc and, it seems to us, rather alien to the spirit of the neo-Fregean project. As already observed, what matters for the success of the stipulation is the equivalence relation itself, not how (if at all) it can be linguistically expressed.

Moreover, this concern with the equivalence relations themselves rather than their linguistic expressions is crucial to many parts of the neo-Fregean project. For instance, a satisfactory answer to the bad company problem cannot be restricted to the equivalence relations that are expressible in some given language but must be robust enough to work for any equivalence relation, no matter how it is expressed. And the same goes for the version of the Julius Caesar problem which asks when an abstract associated with one abstraction principle is identical with one associated with another.

4.2 Defending Step 2

The key to defending Step 2 is the claim that acceptability ensures truth. For by Step 1 we know that each abstraction principle $(*_X)$ is stable. And we have assumed for *reductio* that stability is sufficient for acceptability. So if acceptability ensures truth, then each $(*_X)$ will be true. And since a universally quantified formula is true whenever each of its instances is true, (GA) will then be true as well.

Let's therefore scrutinize the claim that acceptability ensures truth. Applied to an individual abstraction principle, this claim is pretty much analytic. What it *is* for an abstraction principle to be acceptable is just for it to be such that it can legitimately be laid down as a stipulation which ensures that the principle is true. What could possibly be objected to?

The only possible objection, it seems, concerns the move from an individual abstraction principle to a whole family of abstraction principles. The acceptability of a family of abstraction principles ensures that *each* principle can be laid down as a stipulation. But does this guarantee that *all* the abstraction principles can be *simultaneously* laid down as stipulations? Given existing neo-Fregeanism, it is hard to see why not. For one of the neo-Fregeans' main desiderata for the notion of acceptability is precisely that all acceptable abstractions should be compatible. The alternative would be to make the acceptability of an abstraction principle depend on what other such principles have already been laid down. But this would lead to a relativized notion of acceptability which is completely foreign to existing neo-Fregeanism.

4.3 Defending Step 3

Step 3 consists in the claim that two abstracts can be identical only if they are associated with identical equivalence classes of concepts. Let's make this claim more explicit. Let $\S_{\mathbf{R}}(F)$ be the abstract of F with respect to the equivalence relation $\mathbf{R}$. Consider two equivalence relations $\mathbf{R}$ and $\mathbf{S}$ on which abstraction is permissible. Then the claim amounts to the following necessary condition for identity:

(NCI) $$\S_{\mathbf{R}}(F) = \S_{\mathbf{S}}(G) \rightarrow \forall H(\mathbf{R}(F, H) \leftrightarrow \mathbf{S}(G, H))$$

This necessary condition is very plausible. In fact, (NCI) figures as an uncontroversial minimal assumption in the only technically precise discussions of this question that we are aware of.[11]

Moreover, (NCI) is a consequence of two principles to which the neo-Fregeans are already committed. First, their response to the Julius Caesar problem relies on the following

[11] See (Fine, 2002), p. 47 and (Cook and Ebert, 2005), p. 136. All these authors are concerned with a stronger principle, one labelled ECIA_2 in (Cook and Ebert, 2005):

(ECIA_2) $$\S_{\mathbf{R}}(F) = \S_{\mathbf{S}}(G) \leftrightarrow \forall H(\mathbf{R}(F, H) \leftrightarrow \mathbf{S}(G, H))$$

They show that it is a difficult and controversial question whether *the converse* of (NCI) holds, that is, whether two abstracts are identical whenever they are associated with identical equivalence classes of concepts. But (NCI) itself is tacitly assumed in both works.

"*Grundgedanke*."[12]

Grundgedanke: When two objects a and b are subject to different criteria of identity, then a and b are distinct.

Next, it is central to the neo-Fregean conception of abstraction principles that the criteria of identity for the resulting abstracts be completely determined by the equivalence relations on which they are based. But this leaves open the question of when two equivalence relations $\mathbf{R}$ and $\mathbf{S}$ determine one and the same criterion of identity. Neo-Fregeans have considered a very demanding condition for this by suggesting that it should be *conceptually necessary* that $\mathbf{R}$ and $\mathbf{S}$ be coextensive.[13] But this is clearly an overkill for our purposes. We can make do with the much weaker assumption that $\mathbf{R}$ and $\mathbf{S}$ cannot determine one and the same criterion of identity unless they are coextensive. This weaker condition is captured by the following bridge principle, which connects equivalence relations and criteria of identity.

Bridge principle: Let $\mathbf{R}$ and $\mathbf{S}$ be two equivalence relations on which abstraction is permissible. Then the criterion of identity for $\mathbf{R}$-abstracts cannot be identical to that for $\mathbf{S}$-abstracts unless $\mathbf{R}$ and $\mathbf{S}$ are coextensive.[14]

We claim that these two principles imply (NCI). To see this, assume that the consequent of (NCI) is false. If $\mathbf{R} = \mathbf{S}$, then for some H either $\mathbf{R}(F, H)$ and $\neg\mathbf{R}(G, H)$, or $\neg\mathbf{R}(F, H)$ and $\mathbf{R}(G, H)$. In either case it follows from the abstraction principle based on $\mathbf{R}$ that $\S_{\mathbf{R}}(F) \neq \S_{\mathbf{R}}(G)$ and thus $\S_{\mathbf{R}}(F) \neq \S_{\mathbf{S}}(G)$. If instead $\mathbf{R} \neq \mathbf{S}$, then by extensionality $\mathbf{R}$ and $\mathbf{S}$ are non-coextensive. So by the bridge principle, the associated criteria of identity are distinct, whence by the *Grundgedanke* the two abstracts are distinct as well. This proves the contrapositive of (NCI), which establishes our claim.

This is a simple and logically valid derivation of (NCI) from premises endorsed by the neo-Fregeans. But modulo a plausible assumption, even the above bridge principle is more

[12]See (Hale and Wright, 2001b), where this characterization is used.

[13]See (Hale and Wright, 2001b), p. 391.

[14]In fact, the neo-Fregeans consider the much stronger requirement that it be *conceptually necessary* that the two equivalence relations be coextensive (p. 391).

than we need. This is significant, as it might be objected that the neo-Fregeans should not only jettison their account of what it takes for two equivalence relations to provide one and the same criterion of identity but even demur at the above bridge principle.[15] For they may still wish to identify abstracts that are associated with different equivalence relations. For instance, they may wish to identify the natural numbers obtained by means of Hume's principle with those obtained by means of *Finite Hume*, which is like Hume's principle on finite concepts but which refrains from assigning any abstracts to infinite concepts. We don't think the *Grundgedanke* is negotiable. But the mentioned identification is easily seen to be permissible if the bridge principle is weakened to the claim that the criterion of identity for **R**-abstracts cannot be identical to that for **S**-abstracts unless **R** and **S** are coextensive *wherever both relations are defined*. However, even this weakened bridge principle suffices for a defense of (NCI). For instance, (NCI) can still be derived if the plausible assumption is made that when the criterion of identity for **R**-abstracts is identical to that for **S**-abstracts, then $\S_{\mathbf{R}}(F) = \S_{\mathbf{S}}(F)$ for any concept F on which both relations are defined.[16]

4.4 Defending Step 4

Step 4 asserts that the following principle is inconsistent:

$$(1) \qquad \S_X F = \S_Y G \rightarrow \forall H (R_X(F, H) \leftrightarrow R_Y(G, H))$$

We now show why. Let ϵF abbreviate $\S_F F$. Next we substitute F and G for respectively X and Y in (1) and employ this abbreviation to get:

$$(2) \qquad \epsilon F = \epsilon G \rightarrow \forall H (R_F(F, H) \leftrightarrow R_G(G, H)).$$

[15] We are indebted to an anonymous referee here.

[16] To see this, assume again that the consequent of (NCI) is false: that there is some H such that either $\mathbf{R}(F, H)$ and $\neg\mathbf{S}(G, H)$, or $\neg\mathbf{R}(F, H)$ and $\mathbf{S}(G, H)$. We wish to show $\S_{\mathbf{R}}(F) \neq \S_{\mathbf{S}}(G)$. Assume that **R** and **S** provide the same criterion of identity; otherwise by the *Grundgedanke* we are done. Then by the weakened bridge principle, **R** and **S** are coextensive wherever both are defined. But since they are both defined on H, either $\mathbf{R}(F, H)$ and $\neg\mathbf{R}(G, H)$, or $\neg\mathbf{R}(F, H)$ and $\neg\mathbf{R}(G, H)$. In either case it follows that $\S_{\mathbf{R}}(F) \neq \S_{\mathbf{R}}(G)$. Since **R** and **S** have been assumed to provide the same criterion of identity, it follows by the plausible assumption from the text to which this note is attached that $\S_{\mathbf{R}}(F) \neq \S_{\mathbf{S}}(G)$.

Then we observe that the consequent of this conditional is provably equivalent to $\forall x(Fx \leftrightarrow Gx)$.[17] When this observation is applied to (2), we get one direction of Frege's Basic Law V, namely:

$$(\mathrm{V}^-) \qquad \epsilon F = \epsilon G \rightarrow \forall x(Fx \leftrightarrow Gx).$$

This is in fact the direction of Basic Law V that is operative in Russell's paradox. To see this, let $x \in y$ abbreviate $\exists G(y = \epsilon G \wedge Gx)$. Consider now the comprehension axiom:

$$(R\text{-Comp}) \qquad \exists R \forall x(Rx \leftrightarrow x \notin x)$$

This axiom says that the formula $x \notin x$ defines a concept R. Let $r = \epsilon R$. Without invoking any further comprehension axioms we can then derive:[18]

$$(3) \qquad Rr \leftrightarrow \neg Rr$$

But in classical logic this leads to a contradiction, as we set out to show.

Is there any element in this derivation that a neo-Fregean may reject? It would be pointless to reject the definitions made in the course of the argument, as these could be eliminated if desired. Other than the definitions we have relied on: classical first-order logic with identity, the standard introduction and elimination rules for the second-order quantifiers, and the single comprehension axiom (R-Comp). Since the neo-Fregeans are friends of classical first-order logic and regard the introduction and elimination rules for the quantifiers as more or less analytic, the only possible option would be for them to reject (R-Comp). But how plausible would this be? One might attempt to reject the comprehension axiom (R-Comp) on the ground that it is impredicative. For when our

[17] To see that the former implies the latter, instantiate the quantifier $\forall H$ with respect to F and apply the definition of $R_G(G, F)$. To see that the converse holds, we need to show that when F and G are coextensive, then $\forall H(R_F(F, H) \leftrightarrow R_G(G, H))$. But this follows immediately from the observation that all occurrences of F and G in this latter formula are extensional.

[18] A suitable instantiation of the quantifiers of (R-Comp) yields $Rr \leftrightarrow r \notin r$. Assume Rr. Then $r \notin r$, which by unpacking of definitions yields $\neg Rr$. So assume instead $\neg Rr$. Then $r \in r$, which by (V^-) and unpacking of definitions yields Rr.

definitions are unpacked, the concept R is defined by the formula '$\neg\exists G(x = \epsilon G \wedge Gx)$', which quantifies over a totality to which the defined concept R would itself belong.[19] However, there is nothing in the neo-Fregean programme, as it is currently set out, which would support a ban on impredicative definitions of concepts.[20]

5 Concluding remarks

We have argued that stability is not a sufficient condition for acceptability by the lights of a compromising neo-Fregean. As for uncompromising neo-Fregeans, our conclusion is disjunctive. Either stability is not a sufficient condition for an abstraction principle to be acceptable. Or the uncompromising neo-Fregeans will have to appeal to ideas that are quite foreign to their programme, such as the relativized notion of acceptability mentioned in connection with Step 2, a weakening of the necessary condition for abstracts mentioned in connection with Step 3, or the predicativist ideas mentioned in connection with Step 4.[21]

A more general conclusion follows as well. Any sufficient criterion of acceptability will have to be extremely strict, at least if we resist relativizing the notion of acceptability and continue to work against the background of a traditional view on concepts according to which the concepts on any domain D correspond to (at least) the elements of the full powerset of D. The problem is that on any domain D there are more equivalence classes of entities from D than there are elements of D itself. (The example from the previous section shows that there are at least as many equivalence classes as there are elements in the powerset of D.) This means that the vast majority of equivalence classes must fail to determine abstracts. And this in turn means that the vast majority of equivalence relations aren't acceptable for abstraction.[22] This is surprising and somewhat disturbing.

[19] An alternative but related option would be to reject (R-Comp) on the ground that it violates the requirement that the process of individuation be well-founded. See (Linnebo, 2007).

[20] Indeed, the standard proof of Frege's Theorem makes essential use of such definitions. See (Linnebo, 2004).

[21] Moreover, if the neo-Fregeans were to attempt one of the escape routes associated with the second disjunct, they would still face the task of developing and defending an objection to the attempted counterexample of Section 2.

[22] Note that this holds regardless of whether the domain D is a domain of objects, concepts, or whatever.

In particular, what could be more innocent and acceptable-looking than abstraction on equivalence relations that partition the relevant domain into just two equivalence classes?

References

Cook, R. T. (2002). The state of the economy: Neo-logicism and inflation. *Philosophia Mathematica (3)*, 10:43–66.

Cook, R. T. and Ebert, P. (2005). Abstraction and identity. *Dialectica*, 59(2):121–139.

Ebert, P. and Shapiro, S. (2007). The good, the bad and the ugly. Forthcoming in Synthese.

Fine, K. (2002). *The Limits of Abstraction*. Oxford University press, Oxford.

Hale, B. and Wright, C. (2001a). *The Reason's Proper Study: Essays towards a Neo-Fregean Philosophy of Mathematics*. Clarendon Press.

Hale, B. and Wright, C. (2001b). To bury caesar. . . . In Hale, B. and Wright, C., editors, *The Reason's Proper Study: Essays towards a Neo-Fregean Philosophy of Mathematics*, pages 335–396. Oxford University Press.

Heck, R. (1992). On the consistency of second-order contextual definitions. *Noûs*, 26:491–4.

Linnebo, Ø. (2004). Predicative Fragments of Frege Arithmetic. *Bulletin of Symbolic Logic*, 10(2):153–74.

Linnebo, Ø. (2007). Bad Company Tamed. Forthcoming in Synthese.

Rayo, A. and Uzquiano, G. (1999). Toward a theory of second-order consequence. *The Notre Dame Journal of Formal Logic*, 40:315–325.

Shapiro, S. (2000). Frege meets Dedekind: A neo-logicist treatment of real analysis. *Notre Dame Journal of Formal Logic*, 41(4):335–64.

Uzquiano, G. (2007). Bad Company Generalized. Forthcoming in Synthese.

Weir, A. (2003). Neo-Fregeanism: An embarrassment of riches. *The Notre Dame Journal of Formal Logic*, 44(1):13–48.