Hume's Principle, Bad Company, and the axiom of choice

Sam Roberts and Stewart Shapiro

Abstract

One prominent criticism of the abstractionist program is the so-called *Bad Company* objection. The complaint is that abstraction principles cannot in general be a legitimate way to introduce mathematical theories, since some of them are inconsistent. The most notorious example, of course, is Frege's Basic Law V. A common response to the objection suggests that an abstraction principle can be used to legitimately introduce a mathematical theory precisely when it is *stable*: when it can be made true on all sufficiently large domains. In this paper, we raise a worry for this response to the Bad Company objection. We argue, perhaps surprisingly, that it requires very strong assumptions about the range of the second-order quantifiers; assumptions that the abstractionist should reject.

1 Abstractionist neo-logicism

The abstractionist, neo-logicist program in the philosophy of mathematics began with Crispin Wright's seminal [1983]. Bob Hale [1987] joined the cause, and it continues through many extensions, objections, and replies to objections (see Hale and Wright [2001]).

The program's overall plan is to develop branches of established mathematics using abstraction principles in the form:

$$\forall a \forall b (\Sigma(a) = \Sigma(b) \equiv E(a, b)), \tag{ABS}$$

where a and b are variables of a given type (typically first-order or second-order), Σ is an operator, denoting a function from items of the given type to objects in the range of the first-order variables, and E is an equivalence relation over items of the given type. In what follows, we will sometimes omit the initial quantifiers.

Gottlob Frege ([1884], [1893]) employed at least three equations in the form (ABS). One of them, used for illustration, comes from geometry:

> The direction of l_1 is identical to the direction of l_2 if and only if l_1 is parallel to l_2.

The second was dubbed $N^=$ in Wright [1983] and is now called *Hume's Principle*:

$$\forall F \forall G(\#F = \#G \equiv F \approx G), \tag{HP}$$

where $F \approx G$ is an abbreviation of the second-order statement that there is a one-to-one relation mapping the F's onto the G's. In words, HP states that the number of F is identical to the number of G if and only if F is equinumerous with G. Georg Cantor deployed this principle, albeit not formulated as rigorously, to obtain extensive and profound results, especially concerning the transfinite.

Unlike the direction-principle, the relevant variables, F, G here are second-order. We will postpone the question of what entities these second-order variables range over until section 4. For now, we will follow the literature and refer to them as "concepts", or sometimes "Fregean concepts", but without saying much about what those are.

Frege's [1893] third exemplar of an abstraction principle is the infamous Basic Law V:

$$\forall F \forall G(\epsilon F = \epsilon G \equiv \forall x(Fx \equiv Gx)). \tag{BLV}$$

Like Hume's Principle, Basic Law V is second-order, but unlike Hume's Principle, it is inconsistent.

As is now well-known, Frege's *Grundlagen* [1884] and *Gründgesetze* [1893] contain the essentials of a derivation of the Dedekind-Peano postulates from Hume's Principle.[1] This deduction, now called *Frege's theorem*, reveals that Hume's Principle, together with suitable definitions, entails that there are infinitely many natural numbers. The development of arithmetic from HP is sometimes called *Frege arithmetic*. This theory is taken to be the first success story of abstractionist neo-logicism. The underlying theme is that one can introduce HP as a sort of stipulative, implicit definition and develop arithmetic from that.

There is an ongoing program of attempting to found other, richer mathematical theories on abstraction principles. Here, we will mostly be concerned with (second-order) arithmetic and HP.

[1] Frege [1893] used Basic Law V to derive the two conditionals in HP. The rest of the Dedekind-Peano postulates follow from those.

2 Bad Company

One prominent criticism of the abstractionist program is the so-called *Bad Company* objection. The complaint is that abstraction principles are not in general a legitimate way to introduce mathematical theories, since some of them are inconsistent. The most notorious example, of course, is BLV. An analogous abstraction on well-orderings (or relations generally) is also inconsistent, falling to the Burali-Forti paradox (at least if the logic is classical or intuitionistic).[2]

Neil Tennant [1987, 236-7] raises an early formulation of the objection. He notes that the account of arithmetic developed by Wright [1983] may "with justice be called a naïve theory of number". Tennant adds that this "is not to say that [Wright's theory] is inconsistent; it is only to stress the analogy with the naïve theory of sets", based on the inconsistent BLV.

Most of the contemporary discussion of the Bad Company objection traces to Michael Dummett [1991, 188-189] and George Boolos [1997]. In a retrospective moment, Dummett [1998, 375] observed the following:

> My complaint was the obvious one. In *Gründgesetze* [Frege [1893]], value-ranges are introduced in a manner precisely analogous to that in which Wright argued, in his book, that Frege ought to have introduced the cardinal numbers ... and yet it was so far from being justified as to lead to actual contradiction. It therefore could not be maintained that this procedure is, in and of itself, legitimate.

Dummett concedes that this is not conclusive. Rather it raises a challenge: "Possibly some restriction, distinguishing the case of cardinal numbers from that of value-ranges could be framed." And so the abstractionist is charged with the task of finding appropriate constraints or restrictions to separate the "good" cases—the knowledge-grounding abstraction principles—from the "bad" ones, like BLV and the ordinal abstraction principle.

It will be helpful to distinguish two versions of the objection. First, there is what we might call the *epistemological challenge*: namely, to provide a suitable *justification* for HP that does not also justify abstraction principles incompatible with it, such as BLV (which, of course, is incompatible with

[2] For attempts to use other, weaker logics, see Alan Weir [1998] and the dialetheic treatment in Graham Priest [2006]. Those accounts are not based on abstraction principles, however.

everything). Since HP and BLV are both abstraction principles, it can't be that HP is justified merely in virtue of its being an abstraction principle. So, what is its justification?[3] Dummett [1998, 375] seems to have had this challenge in mind.

In response to this challenge, Wright and Hale suggest that abstraction principles be understood as implicit definitions, which, at least in favorable cases, can be made true by stipulation. Wright notes that, like other implicit definitions, the procedure of introducing concepts and, indeed, abstract objects, via stipulation is defeasible. Things can go wrong. However, it does not have to be explicitly justified on each occasion. It is enough that a given abstraction principle be free from any known defeaters. The proposal is that HP be taken to be an "entitlement" (Wright [2004]). In a slogan, an abstraction principle can be taken to be innocent until proven guilty.

The second version of the Bad Company objection is what we might call the *classificatory challenge*: namely, to provide some informative classification of abstraction principles into the good/acceptable and bad/unacceptable cases. Øystein Linnebo [2009] seems to have this challenge in mind when he says:

> [...] the bad company problem highlights the need for an account of what kinds of abstraction are legitimate. We would like to draw a mathematically informative and philosophically well-motivated line between the acceptable abstraction principles and the unacceptable ones.

It is not clear how the two challenges relate to each other (see Ebert and Shapiro [2009]). A particular justification of HP need not tell us which abstraction principles in general are acceptable. Indeed, Wright's own response to the epistemic challenge does not say anything about the classificatory challenge. And in principle it seems that one could meet the classificatory challenge without providing any justification for HP, or any other abstraction principles, whatsoever.

John Burgess [1984] and George Boolos [1987] observed that HP, and thus Frege arithmetic, is consistent, if second-order Dedekind-Peano arithmetic is. Let D be a property (or set) of natural numbers. If D is finite, of cardinality n, let $\#D$ be $n+1$, and if D is infinite, let $\#D$ be 0. It

[3] This version of the Bad Company objection is best seen as aimed at key epistemological aspects of the abstractionist program. The targets are various theses that some abstraction principles, like HP, are analytic, or all but analytic, or that they are knowable without much epistemological cost, or . . .

is straightforward to verify that, with this definition, HP is provable in a standard deductive system for second-order arithmetic.

It is common ground, in this debate, that consistency is not a sufficient condition for the acceptability of an abstraction principle. There are at least two reasons for this. The most often cited one is that there are consistent abstraction principles that have no infinite models. One such is the *parity principle* in Boolos [1997], and another is the *nuisance principle* presented in Wright [1997]. Both are satisfiable on any finite domain, but not on any infinite domain. If we assume that HP is an acceptable abstraction principle, then our overall theory will commit us to infinitely many objects and, in that case, neither the parity principle nor the nuisance principle are acceptable. And the argument goes in the other direction as well. If either the parity principle or the nuisance principle is acceptable, then HP is not, since the other two abstraction principles each require that the domain be finite.

A second, and perhaps more important reason concerns the goals of logicism, and the various forms of neo-logicism. Frege, in effect, argued that (cardinal) numbers belong to logic since, like logic, arithmetic is universally applicable. If one is dealing with objects at all, then one can ask *how many* of them there are in a given context. The model constructed by Burgess and Boolos consists of only natural numbers, and they show that, on this domain, the theory has the wherewithal to define a cardinality-operator on Fregean concepts of natural numbers.[4] So we only speak of *numbers of numbers*, not numbers of anything else.

If abstractionist neo-logicism is to meet its stated goals, however, it must be consistent to add HP to *any* legitimate theory whatsoever. If HP is to be truly universally applicable, then we need cardinal numbers of people, countries, grains of sand, stars, ... (vagueness aside). And we need cardinal numbers of real numbers, points in various geometric spaces, functions on real numbers, sets ...

In this article, we explore perhaps the most prominent response to the Bad Company objection, namely, that the acceptable abstraction principles are the *stable* ones. In the next section, we outline the stability response and show how it relies crucially on the axiom of choice. In Section 4, we show that this entanglement with the underlying set theory is problematic for certain kinds of abstractionists, but not for others. In Section 5 we show that there

[4] As Burgess and Boolos note, a more natural model would consist of the natural numbers plus one more item, $\aleph_0$. This model contains what are usually taken to be the cardinalities of sets of natural numbers. But, in both cases, it is only sets (or concepts) of cardinal numbers that have cardinalities.

is a deeper and more general entanglement that proves problematic for all abstractionists.

3 Stability and choice

Øystein Linnebo [2009] is an excellent introduction to the Bad Company problem, and an overview of the state of play at that time. He articulates a common proposal:

> A ... suggestion is that an abstraction principle is acceptable if and only if it is *stable*, where an abstraction principle $(*)$ is said to be stable if there is a cardinal number κ such that $(*)$ is satisfiable in models of cardinality λ for any $\lambda > \kappa$.[5]

Some potentially annoying details of the back ground meta-theory are relevant here. Since our interest is in the interaction between stability and various choice principles, we do not want the very definition of stability to presuppose choice. Within the background set theory, the cardinality of a set is usually identified with the least von Neumann ordinal it is equinumerous to, but in the absence of choice it may be that some sets are not equinumerous with any von Neumann ordinal. The most common alternative, due to Dana Scott, is to identify the cardinality of a set with the set of all sets, of minimal rank, that are equinumerous with it.[6] This requires that the sets are organised into a rank structure, and we'd rather not presuppose such a substantial assumption in the mere definition of stability. But luckily, for each of these notions of cardinal number, the above characterization of stability is equivalent to the following, which requires neither a rank structure nor the axiom of choice.

> An abstraction principle $(*)$ is *stable* just in case there is some set x such that $(*)$ is satisfiable over any set into which x is

[5] The notion of stability is mentioned in Heck [1992, 494, note 5]. Alan Weir [2003, 32] shows that an abstraction principle is stable if and only if it is *irenic*, i.e., if it is compatible with every abstraction principle that enjoys a certain conservativeness property (see Wright [1997, 23]). Weir's proof relies on the axiom of choice in the meta-theory.

[6] The Scott "trick" thus presupposes the axiom of foundation. Lévy [1969] shows that in the absence of foundation, there can be infinite but Dedekind finite sets—that is, sets into which each natural number can be one-one mapped but into which there is no one-one mapping from all the natural numbers. It follows that in the absence of foundation, HP can fail to be satisfied on some infinite set, since the satisfiability of HP over a set implies that set is Dedekind infinite.

> injectable.[7]

We will thus take stability to be the thesis that an abstraction principle $(*)$ is acceptable if and only if there is some set x such that $(*)$ is satisfiable over any set into which x is injectable.

It is often noted that HP is stable and, indeed, this is easy to see if we take the background meta-theory to be Zermelo-Fraenkel set theory with choice (ZFC):

> Theorem 1 (ZFC): HP is stable
>
> The proof is a straightforward extension of the above consistency argument. The cardinal in question is $\aleph_0$, the cardinality of the natural numbers. Let A be any set for which $|A| \geq \aleph_0$. By the axiom of choice, we can just let A be a von Neumann cardinal λ, in the $\aleph$-series.
>
> Let $D \subseteq A$. If D is finite, of cardinality n, then let $\#D = n + 1$. If D is infinite, but smaller than λ, then let $\#D$ be the usual cardinality of D, which, of course, is a member of λ. Finally, if D is equinumerous with λ, then let $\#D = 0$. It is easy to verify that, with this definition, HP is true.

The proof makes use of the axiom of choice in the assumption that every set can be well-ordered (and is thus equinumerous with some von Neumann cardinal). Much of the relevant literature follows suit, making free use of the axiom of choice. It is common, for example, to simply assume that all cardinals are comparable: if κ and λ are cardinals, then either $\kappa \leq \lambda$ or $\lambda \leq \kappa$. When cardinals are defined via Scott's trick, however, this claim is equivalent to the axiom of choice.

The use of choice in Theorem 1 is essential:

[7] We should clarify what we mean by "satisfiable over a given set x". There are two salient options. We might mean satisfiable in the full second-order structure over x—one where the monadic, second-order variables range over the full powerset of x, and similarly for dyadic second-order variables. Or else we might mean satisfiable in *some* second-order structure over x, a so-called Henkin model (see Shapiro [1991]). Say that a principle is *standardly-stable* when it is stable in the first sense, and say that it is *Henkin-stable* when it is stable in the second sense. It turns out that there are individually Henkin-stable principles that are jointly inconsistent (see the Appendix for a proof.) For this reason, we will always work with standard stability.

> Theorem 2: There is a model of ZF in which HP is not stable. Specifically, given mild large cardinal assumptions, there is a model of ZF for which: there are arbitrarily large ranks V_α satisfying second-order ZF (i.e. there are arbitrarily large inaccessible ranks) such that HP is not satisfiable over any of them.

See the Appendix for the proof.[8]

What's the significance of this result? It strongly suggests that in order to prove the stability of HP we need to assume the axiom of choice in the form that every set can be well-ordered. Independently of that, it is easy to see that the stability of HP requires the axiom of infinity (since HP is only satisfiable on infinite domains). So the stability of HP seems to require, at a minimum, both the axiom of well-ordering and the axiom of infinity. Moreover, given some mild auxilliary assumptions—like the axiom of separation—the axioms of well-ordering and infinity suffice to prove that HP is stable (along the lines of Theorem 1). So well-ordering and infinity seem to be both necessary and sufficient to establish the stability of HP.

An immediate upshot is that if the abstractionist is not entitled to these axioms, the stability response to the Bad Company objection may be undermined. Whether they are entitled to those axioms depends on what perspective they take to abstractionism. We look at this question in the next section.[9]

[8] One upshot of Theorem 2 is that there are models of ZF such that *no* set can be extended with objects so that it both satisfies (i) HP and (ii) second-order ZF. In these models, second-order ZF and HP are kind of like the "distraction principles" in Weir [2003]. Any set that satisfies HP fails to satisfy second-order ZF, and vice-versa. This extends, moreover, to second-order ZF with urelements if we assume that the urelements form a set. We don't know what happens when that assumption is relaxed.

[9] What of the iterative hierarchy itself, which, of course, is not a set, at least not in ZF? The "C" in ZFC is sometimes called an axiom of *local* choice. It states, in effect, that every *set* has a choice function. Zermelo's celebrated well-ordering theorem is that this axiom entails that every *set* has a well-ordering. So the stability of HP, as that is interpreted above, does not entail that it can be satisfied on the iterative hierarchy itself (i.e., on V).

A proof like that of Theorem 1 would go through, for the iterative hierarchy itself, if we had that all so-called proper classes (or all properties whose extension is not a set) are equinumerous with each other. Then we would just need just one "cardinality" for them. In light of the other axioms, the statement that all proper classes are equinumerous with each other is equivalent to the existence of a global well-ordering, since it would entail that the von Neumann ordinals are equinumerous with the universe. But this goes beyond ZFC.

4 Perspectives

There are different perspectives one can take toward the abstractionist project (beyond disinterest). One is that of the abstractionist himself. The focus is on mathematical principles that can be stated and derived in a suitable second-order logical deductive system, augmented with various acceptable abstraction principles. Call this the *internal perspective.*

A paradigm of the internal perspective is Frege's theorem, the derivation of the Dedekind-Peano axioms from HP plus explicit definitions of the primitive terminology of arithmetic. As noted above, the abstractionist takes this result to shed light on the epistemic status of arithmetic, via the epistemic status of the abstraction principle.

A second orientation is that of the established mathematician, observing the program as it unfolds. She is interested in determining which mathematical structures have been *recaptured* by the abstractionist. To explore this terrain, the mathematician inquires into the meta-theoretic properties of the abstractionist systems. She assesses the abstractionist program from the point of view of someone who already has (or assumes they have) a rich, functioning mathematics, including a robust mathematical ontology, a set theory in particular. Let us call this the *external perspective.* The mathematician uses every tool at her disposal, whether the abstractionist is able to reconstruct it or not.

Kit Fine [2002, 10] describes the external orientation toward the abstractionist program:

> We have therefore what might be called an externalist characterization of a given position, one which can be regarded as correct, or even as intelligible, only by someone who does not hold the position ... Given that we adopt the externalist perspective, it is natural to ask: What is the size of our opponent's universe? In response to this question, it seems that the best we can do is to see which cardinals will render true what our opponent takes to be true.

Presumably, the "opponent" of the externalist here is the internalist, although it's not entirely helpful to put things in adversarial terms. Indeed, externalism and internalism aren't mutually exclusive. We could be externalists about some abstraction principles (or some features of a given abstraction principle) and internalists about others.

Following Fine [2002], it is helpful to draw another distinction between two kinds of internal perspective. The *uncompromising* abstractionist only accepts those branches of mathematics that can be recaptured via acceptable abstraction principles. We might call such abstractionists *imperialists.*

In contrast, the *compromising* abstractionist is prepared to accept branches of mathematics, like set theory, that have not been recaptured via acceptable abstraction principles. Of course, if a certain branch cannot be reconstructed on the basis of acceptable abstraction principles, then it won't inherit the epistemic status of those principles. But this does not mean that its basic principles cannot become known at all.[10] Set theory might be justified on the basis of some sort of Kantian or Gödelian intuition, or perhaps on broadly abductive, or pragmatic or holistic grounds, ala Quine. Or set theory may not need any extra-mathematical justification at all (as, for example, is argued in Maddy [2007]). Our compromiser need not take a stand on this issue. She is out to provide an abstractionist foundation for what she can, leaving the rest of mathematics to fend for itself.

In effect, the compromising abstractionist combines the internal and external perspectives. So long as the left hand knows that the right hand is doing, there need be no conflict. When engaged in abstractionist constructions (or reconstructions), the compromiser must be careful not to smuggle in any unwarranted set theory, for this may undermine the epistemic goals of the program. However, they can assess the set-theoretic properties of various abstraction principles, and use the results to guide their abstractionist theorizing.

We now have two distinctions, between the epistemological and classificatory challenges arising from the Bad Company objection, on the one hand, and between imperialist and compromising abstractionists on the other. The significance of our central result initially appears to depend on which of these perspectives one adopts.

We see no easy objection to the claim that stability plays a role in the compromising abstractionist's response to the classificatory challenge. From the point of view of our compromiser, one can use the resources of set theory, or any other mathematical theory, to provide (external) evidence that HP is stable and thus on the good side of the divide.

What of the imperialist? It is conceivable that they might develop an abstractionist theory of sets, perhaps along the lines of Hale [2000] or Shapiro [2003]. One could then formulate the notion of stability internally, in which case it might be made available to them. The idea would be that the imperialist first presents Hume's Principle, as the basis for a theory of cardinality. They then develop a set theory, with its own account of satisfaction, and

[10] Wright has suggested a view like this (pc).

use that to formulate and prove stability.

Even if this plan makes sense philosophically, it is, at best, only a promissory note, turning on developing an internalist friendly set theory. Moreover, it looks like such a set theory would be rather limited. Gabriel Uzquiano [2009] points out that any set theory with countable replacement (stating that every countable collection forms a set) fails to be stable: no such theory is satisfiable on any domain with with countable cofinality. So if stability marks the criterion for an abstraction principle to be acceptable, then even relatively meagre set theories seem to be out of bounds.[11]

To be sure, the set theory required to prove the stability of HP is itself limited—we only need infinity and well-ordering, and some staples like separation. So it may be that a weaker set theory could be developed for this purpose. But let us briefly note one worry with this kind of strategy. Since sets will be abstracted from Fregean concepts on any abstractionist reconstruction of set theory, it is hard to see how such a reconstruction would yield the axiom of well-ordering for sets *unless* concepts were themselves already well-orderable. If they were, it would mean that the universal concept was well-orderable and thus that there was a *global well-ordering*: a well-ordering of absolutely all objects. That is a significant assumption about concepts and one we might reasonably doubt. Indeed, we will argue in the next section that the abstractionist *should* doubt it.

Let us now turn to the epistemological challenge raised by the Bad Company objection. Our central result shows that to prove the stability of HP we already need a set theory that licences the axioms of infinity and well-ordering. If stability is to form part of our *justification* for HP, then its epistemic status will be constrained by that of the underlying set theory. If that set theory doesn't require an abstractionist reconstruction, the epistemic interest of HP is greatly diminished—its epistemic status would in that case be achievable without abstraction. On the other hand, if it does require an abstractionist reconstruction, then we seem to be caught in a vicious circle. The justification of HP would involve our justification of the underlying set theory, and that would in turn require the justification of a set theory in which *that* set theory is provably stable, and so on. Indeed, in each case the relevant set theory would have to be proof-theoretically *stronger* than the previous set theory, since stability implies consistency.

One way to break out of this circle is to claim that stability plays an

[11] The imperialist might retreat to a fall back position: stability is only a sufficient condition on abstraction principles. A richer, non-stable but nonetheless acceptable, set theory might then still be developed.

external role in our justification for HP, roughly along the lines of externalist epistemologies. The thought would be that we are justified in believing HP in part because it *is* as a matter of fact stable, even though we may not be justified in believing that it is stable.

As Ebert and Shapiro [2009, §6.2] have argued, however, this "externalist" epistemology is problematic for the abstractionist. To see this, let S be any true statement in the language of arithmetic, say the statement of Fermat's Last Theorem. Developing a technique due to Richard Kimberly Heck [1992],[12] Ebert and Shapiro [2009, §6.2] show how to formulate an abstraction principle HP[S] that is equivalent to the conjunction of HP and the result of restricting the quantifiers of S to the finite cardinal numbers. Moreover, HP[S] is stable if HP is. If the mere stability of an abstraction principle were justification for it, then we could stipulate HP[S], and thus become entitled to S. So we did not need the heroic work of Andrew Wiles to settle Fermat's last theorem!

A weaker externalist position would be the claim that an abstractionist can adopt an abstraction principle if she has *some* justification to believe that it is stable, even if that justification falls short of abstractionist standards. However, this proposal fails for similar reasons. Let S be as above, but now assume that we know S somehow or other. Suppose, for example, that S is Con(ZFC). Assuming we know that HP is stable, it follows from the above argument that we also know that HP[S] is stable. So, by the current weak externalist proposal, we are justified in believing HP[S] and thus *S by abstractionist standards.* In general, this gives us a recipe for bootstrapping pedestrian justifications of arithmetic statements into abstractionist ones. Again, an embarrassment of riches.[13,14]

The issues here are subtle, and we do not claim to have given the last word. We turn now to some considerations that affect all versions of abstractionism.

5 Stability and reflection

The axiom of choice—and in particular, the principle of well-ordering—appears to be necessary to prove that HP is stable (Theorem 2). In the last section, we argued that this was problematic for many perspectives on

[12] See also §5 below.

[13] Of course, we could rely Wright's [2004] notion of entitlement to respond to the epistemological challenge. But that would leave no epistemological role for stability.

[14] Thanks to two anonymous referees for pressing us on these options.

abstractionism. Some, however, seemed unaffected. In particular, there appears to be no good reason why the compromising abstractionist addressing the classificatory challenge shouldn't appeal to stability. In this section, we develop a much more general worry about the connection between stability and choice that we argue affects all perspectives.

Stability is a claim about *sets*. It says that an abstraction principle is acceptable if and only if it can be made true when its second-order quantifiers are *re-interpreted* as ranging over subsets of all sufficiently large sets. At first sight, this should strike you as strange. Why should there be such a tight connection between two prima facie distinct domains: between the acceptability of claims about Fregean concepts, on the one hand, and the behaviour of the subsets of sufficiently large sets, on the other? Why should the latter tell us *anything* informative about the former?

We can draw this point out by invoking a technique due to Heck [1992]. Let φ be any sentence in the language of second-order logic and assume that it is stable. Then the following abstraction principle will also be stable:

$$\#F = \#G \leftrightarrow (\forall x(Fx \equiv Gx) \vee \varphi) \qquad (\mathsf{ABS}_\varphi)$$

since φ is satisfied on a domain precisely when ABS_φ is satisfiable over the domain. Effectively, ABS_φ is Basic Law V conditional on $\neg\varphi$. Since Basic Law V is false, ABS_φ implies φ.

So when φ is stable, ABS_φ is stable and implies φ. By the stability response to the Bad Company objection, that means that when φ is stable, ABS_φ, and thus φ, are acceptable. In general, the stability response implies that any stable claim, in the language of second-order logic, is acceptable.[15]

The stability response thus commits us to a kind of *reflection principle* which says that whatever is true of the subsets of sufficiently large domains is also acceptable for concepts. Conversely, it says that if something is not acceptable for concepts, then it is false for subsets of arbitrarily large domains. Stability thus implicitly assimilates concepts to sets. This has significant consequences.[16]

Consider, first, comprehension principles. Since sets obey the axiom of separation, every instance of the second-order comprehension schema:

$$\forall\vec{y}\forall\vec{G}\exists F\forall\vec{x}(F\vec{x} \leftrightarrow \varphi) \qquad (\mathsf{comp})$$

[15] Note that this argument only relies on the claim that stability is sufficient for acceptability, not that it is necessary.

[16] See Roberts [ms] for a discussion of these consequences in the context of set-theoretic reflection principles.

is satisfied over every set (where φ's free variables are among $\vec{y}, \vec{G}$). So each instance of comp is stable and therefore acceptable according to the stability response.

That is already a substantial commitment. It is well known that the abstractionist needs *some* impredicative comprehension in order to establish Frege's Theorem (that the Dedekind-Peano postulates follow from HP plus the usual definitions). To be precise, the proof of Frege's theorem uses comprehension on Π^1_1-formulas, and one cannot improve (much) on that.[17] But here we see that stability requires the acceptability of full impredicative comprehension, given the axiom of separation for sets.

Next, consider well-ordering principles. As we've argued, the stability of HP requires that every set admit a well-ordering. So the second-order claim that there is a well-ordering of all objects—a global well-ordering—will be stable. It follows from the above observation that it too must be acceptable according to the stability response.

But why think that concepts *are* like sets in these respects? There are two broad kinds of conception we might have about concepts: *combinatorial* conceptions and *definability* conceptions.[18] According to combinatorial conceptions, concepts are given by *arbitrary* "choices" over the first-order domain. Metaphorically, we "run through" the first-order domain choosing for each object whether to keep it or throw it away. Any such choice results in a concept that applies to the chosen things and nothing else. Similarly, we can choose for each pair of objects whether they relate or not. Any such choice results in a relation that relates the first to the second elements of those pairs and nothing else. On this kind of conception, we take it as plausible that both impredicative comprehension and global well-ordering hold.

Sets are typically taken to be a paradigm case of entities governed by a combinatorial conception. But given the ambition of HP to generality, they provide an unsuitable way of understanding concepts. There should be a number of (absolutely) all objects, for example, but there is no set of all objects, by Russell's paradox. Concepts are not sets.

Pluralities are perhaps *the* other paradigm case of entities governed by

[17] Sean Walsh [2012], Corollary 92, shows that full first-order Dedekind-Peano arithmetic cannot be interpreted in a system with HP and only predicative instances of comprehension. It follows from this and Walsh's Proposition 6, that the interpretability strength of the system of second-order arithmetic known as ACA_0 is strictly above that of HP with Δ^1_1-comprehension. Walsh points out that the result can be strengthened so as to replace ACA_0 with (first-order) PA (pc).

[18] See, for example, Maddy [1983].

a combinatorial conception.[19] Unfortunately, we think they also provide an unsuitable way for the abstractionist to make sense of concepts. First, there is considerable controversy over the nature and extent of plural quantification. For example, some have thought that plural quantification is really set quantification in disguise (e.g., Resnik [1988]). A more subtle version of this view takes it that even if some plural locutions can be understood in terms of primitive plural resources, more complicated plural locutions can only be understood by taking them to implicitly quantify over sets.[20] For them, plural locutions of arbitrary complexity cannot be understood independently from set quantification. Moreover, it is a standard view in linguistic semantics that plural quantification is best understood only via set quantification (e.g. Landman [1989]).

Others hold that although plural quantification does not presuppose set quantification, there is nonetheless an intimate connection between pluralities and sets. Sets, on this view, are *given* by pluralities and every plurality specifies a corresponding set.[21] So although plural quantification can be understood independently of set quantification, every plurality happens to be co-extensive with some set. And just as there is no set of all objects, there is no plurality of all objects.[22]

Another reason to be unsatisfied with pluralities concerns the universal applicability of Hume's principle and the relationship between plural and modal logic. The abstractionist wants Hume's principle to deliver numbers for as many collections of objects as possible. Now consider the following example from Williamson [2003]. We have a single knife handle, h, and two knife blades b_0 and b_1. Two possible knifes can be made from these components: one, k_0, comprising h and b_0 and one, k_1, comprising h and b_1. HP should help us to capture this fact. But to do so, there needs to be some concept that can apply to both k_0 and k_1. It is natural to think, however, that they are *incompossible* because they share a handle: they metaphysically cannot co-exist.[23] But pluralities are nothing over and

[19] See Roberts [forthcoming] for discussion.

[20] One of the present authors is sympathetic to this line of thought, whereas the other is not. See Roberts [forthcoming] and Shapiro [1991, §9.1.1].

[21] See, for example, Linnebo [2010].

[22] Indeed, one might think that the best shot at an abstractionist set theory is by accepting something like Basic Law V for pluralities. In that case, plural comprehension would have to be jettisoned and an account given of which instances are true. See Florio and Linnebo [2020] for the beginnings of such an account.

[23] Williamson [2003] ultimately rejects this natural thought. For him, existence is a noncontingent matter. No possible objects could have failed to exist and so any two possible objects will necessarily co-exist.

above the things they comprise (Roberts [forthcoming]) and so any plurality comprising at least k_0 and k_1 cannot exist without each of them. Since they cannot co-exist, it follows that there couldn't have been such a plurality. In other words, if we want to apply HP in these kinds of cases, concepts need to be *intensional* in a way that pluralities are not.[24]

Let us mention one final potential worry for understanding concepts in terms of pluralities. It has often been noted that although plural quantification can be used to make sense of *monadic* second-order quantification, it cannot without supplementation be used to make sense of *relational* second-order quantification (Boolos [1984],[1985]). But the latter is needed to state HP. So, if the abstractionist goes this route, she will need to supplement plural quantification with an ordered pairing operator (so that binary relations can be construed as pluralities of pairs).

To be sure, ordered pairs can be introduced via a (stable) abstraction principle, if we can allow abstraction over four-place relations. In the following, π stands for the ordered-pair operation:[25]

$$\pi(a, b) = \pi(c, d) \equiv (a = c \land b = d)$$

However, like HP, this pair-abstraction delivers an infinity of abstracts: if there are at least two objects, then there are no finite models of this principle.

So if the abstractionist goes this route, they can use the resulting infinity of objects together with the global well-ordering established above to explicitly define an operator witnessing HP, along familiar lines. In so far as we want to ground arithmetic in abstraction, then, this strategy makes HP redundant.[26]

So perhaps the abstractionist should simply abandon a combinatorial conception of Fregean concepts. In recent work, Hale [2010], [2013] does just that (see also Hale [2013a]):

> In the so-called "standard" semantics [for second-order languages], the second-order property variables are interpreted as ranging over the full classical power set of the individual domain. Classically, a subset of a given set is thought of as the result of a sequence of choices—one for each element of the set—whether

[24] See Fritz and Goodman [2017] for further discussion.

[25] See Shapiro [2000].

[26] Of course, grounding arithmetic is not the only reason we might be interested in HP. Arithmetic does not by itself give us a universally applicable notion of number like HP does. Thanks to two anonymous referees for pushing us on this point.

> it is to be or not to be an element of the subset. The choices may be guided by a rule, or determined by some stateable condition for membership, but they may equally well be arbitrary ... In the case that the set is infinite, ... we must suppose an uncountably infinite sequence of such infinite sequences of arbitrary choices. At this point, we engage in theology, and suppose that while no finite being can perform even one such infinite sequence of arbitrary choices, this limitation can be set aside—we can assume that God has done the work for us, or that there is no need for it, because all the sets which would be determined if this hypertask could be performed already exist anyway, just in virtue of the existence of their members. (Hale [2013, 145])

Hale then explicitly articulates and defends what we are here calling a "definability conception" of Fregean concepts:

> ... *properties* and *relations* are what (one- or more-place) predicates stand for. More precisely *first-level* properties, or properties of objects, are what first-level predicates stand for–a first-level predicate being any expression which, applied to a suitable number of singular terms, yields a sentence. (Hale [2010, 405])

Attention is not restricted to the predicates—open formulas—of any actual formal or natural language. We talk about what predicates there *could be*, and thus of what languages there could be:

> Roughly, *properties* and *relations* are those things for which predicates can stand, and a sufficient (and ... in my view, necessary) condition for their existence is that there could be predicates with appropriately determinate satisfaction conditions. (Hale [2013, 134])

There is, of course, a longstanding tradition that adopts this conception of properties, including Whitehead and Russell [1910], Poincaré [1906] and Weyl [1918]. It is alive today in the work of S. Feferman (e.g., [2006]) and others. But these thinkers are clear that we have to do without impredicative definitions for that reason. It seems to be common ground, in that debate, that definability conceptions do not sanction impredicative conception. At the least, the burden of proof is on someone who advocates the acceptability of at least some impredicative definitions on a definability conception of Fregean concepts.

Similarly, there appears to be no reason to expect that a global well-ordering will be definable in some language. Much of the debate over the axiom of choice during the early decades of the twentieth century turned on the matter of definability. Advocates of choice principles explicitly rejected definability while the arguments brought by opponents turned on definability (see Moore [1982]). Again, we take these discussions to show that the burden of proof is on someone who advocates a principle of global well-ordering on a definability conception of Fregean concepts.[27] As we have seen, stability commits the abstractionist both to impredicative definitions and the axiom of global well-ordering. To say the least, sanctioning those principles on a definability conception is problematic.

We do not take our discussion here to be conclusive. In principle, there may be other, as yet undiscovered, conceptions of Fregean concepts that licence impredicative comprehension and global well-ordering, perhaps even conceptions that do not fit neatly into the combinatorial/definability categories. As it stands, however, we know of no such conceptions.[28]

In sum: stability engenders a kind of reflection principle. It says that what is true of the subsets of arbitrarily large domains must also be acceptable for Fregean concepts. Theorem 2 suggests that we need something like an axiom of well-ordering to obtain the stability of HP whatever perspective we take. So it follows from the stability response to the Bad Company objection that the global well-ordering principle must be acceptable. This, we argued, is problematic for the abstractionist.

[27] See Linnebo [2004] and Shapiro [2018] for more on these issues in the context of abstractionism.

[28] It is natural at this point to think of the recent trend in metaphysics that takes concepts at face value, understanding second-order quantification as a new primitive resource not to be explained in other terms. Williamson [2003], for example, has argued that we can understand second-order quantification primitively in this way. (Rayo and Yablo [2001] have argued that some such notion is already present in English.) He has, moreover, endorsed impredicative comprehension on such an understanding, and although he does not directly discuss the issue of a global well-ordering, it seems clear that he would endorse that too. The problem, as we see it, is that a way of understanding concepts—whether primitive or otherwise—does not by itself constitute a conception of them; it does not by itself provide an explanation of which instances of comprehension are true nor whether there is a global well-ordering. As yet, no such conception is forthcoming. In so far as Williamson provides justification for these principles it is of a broadly abductive kind that fits poorly with the abstractionist's epistemic goals. See Roberts [ms] for further discussion. Thanks to a referee for pushing us on this point.

Appendix

> **Theorem 2**: If ZFC + there is a proper class of inaccessible cardinals is consistent, then so is ZF + there is a proper class of inaccessible cardinals, HP is not satisfiable over any inaccessible rank, and each inaccessible rank satisfies second-order ZF.

Proof. Let M be a ctm containing what it thinks is a proper class of inaccessibles.[29] For simplicity, we assume that $M \vDash V = L$. Working in M, for each regular cardinal κ we let $\mathbb{P}_\kappa$ be the partial order consisting of characteristic functions from κ with domains of size less than κ: $\mathbb{P}_\kappa = \{f : \kappa \to \{0,1\} : |\mathsf{dom}(f)| < \kappa\}$. This is just the partial order that adds a Cohen "real" over κ. Let $\mathbb{P}$ be the Easton product of these partial orders.[30]

It is standard that when G is a generic for $\mathbb{P}$, $M[G]$ is a model of ZFC. Similarly, it is standard that $M[G]$ preserves regularity and continua (i.e. for all $\alpha \in M$, $(2^\alpha)^M = (2^\alpha)^{M[G]}$). It follows that it also preserves inaccessibility.

Now, $\bigcup G$ associates each regular cardinal with one of its subsets. Let A be the set of all these subsets. That is, $A = \{x : \exists\alpha(\mathsf{reg}(\alpha) \wedge \forall y(y \in x \leftrightarrow \bigcup G(\alpha)(y) = 1))\}$. It is easy to see that the elements of A are precisely the sets denoted by names that are given by functions f of the form: for some regular cardinal α, (i) $\mathsf{dom}(f) = \alpha$ and (ii) $\forall\beta < \alpha(f(\beta) = \{\langle\alpha, \{\langle\beta, 1\rangle\}\rangle\})$ (in other words, f maps each $\beta < \alpha$ to the function from α to the function from β to 1). Now consider the names that are just like these except that they map finitely many $\beta < \alpha$ to $\{\langle\alpha, \{\langle\beta, 0\rangle\}\rangle\}$. Let A' be the sets denoted by these names. A simple density argument shows that the α-sized sets in A' are precisely those denoted by such names where $\mathsf{dom}(f) = \alpha$. Let $A'(\alpha) = \{x \in A' : |x| = \alpha\}$. So the elements of $A'(\alpha)$ are precisely the sets denoted by those names.

It is straightforward to show that $\langle M[G], A'\rangle$ satisfies ZFC with its schemas of separation and replacement extended to formulas involving a predicate for A'. In particular, it is straightforward to show that the truth and definability lemmas extend to such a language. For simplicity, we will identify $\langle M[G], A'\rangle$ with $M[G]$.

Working in $M[G]$, we define $HOD(A')$ and observe that it models ZF.

[29] We work with the following choice free notion of inaccessiblity: κ is inaccessible just in case for all $\alpha < \kappa$ and $f : \mathcal{P}(\alpha) \to \kappa$ the least upper bound of $\mathsf{rng}(f)$ is less than κ.

[30] That is, $\mathbb{P} = \{f : \mathsf{dom}(f) \subseteq \mathit{Reg} \wedge \forall\alpha \in \mathsf{dom}(f)(f(\alpha) \in \mathbb{P}_\alpha) \wedge \forall\alpha(\mathsf{reg}(\alpha) \to |\mathsf{dom}(f) \cap \alpha| < \alpha\}$.

Moreover, we note that since $M \vDash V = L$, $M \subseteq HOD(A')^{M[G]} \subseteq M[G]$ and thus each agrees on regularity, continua, and inaccessibility. Since $V_\alpha^{HOD(A')^{M[G]}} \subseteq V_\alpha^{M[G]}$, each inaccessible rank in $HOD(A')^{M[G]}$ will satisfy second-order ZF. We will show that $HOD(A')^{M[G]}$ is the required model.

Now, working in $HOD(A')^{M[G]}$, let κ be inaccessible. There will be set, B, of κ^+ subsets of $\kappa \cap Reg$ that disagree with each other arbitrarily high in κ. For each element $X \in B$, let $X' = \bigcup_{\alpha \in X} A'(\alpha) \subseteq V_\kappa$. It is easy to see that when $X \neq Y$ for $X, Y \in B$, $X' \neq Y'$. So $B' = \{X' : X \in B\}$ has size κ^+ too. We will now show that no two elements of B' are equinumerous.

Assuming that is correct, if HP were satisfiable over V_κ, we would get a one-one map from κ^+ into V_κ. Since $HOD(A')^{M[G]}$ and $M[G]$ agree on regular cardinals and $V_\kappa^{HOD(A')^{M[G]}} \subseteq V_\kappa^{M[G]}$ that would mean we'd have in $M[G]$ a one-one map from κ^+ into V_κ. But in $M[G]$, $|V_\kappa| = \kappa$, since we have the axiom of choice and κ is inaccessible. So we would have a one-one function from κ^+ into κ in $M[G]$, which is impossible.

So let X and Y be two elements of B' and assume that g is a one-one function between them in $HOD(A')^{M[G]}$. It follows that g is definable in $M[G]$ by some condition φ using ordinal parameters, parameters from A', and a predicate for A'. Let $\alpha_0, ..., \alpha_n$ be its ordinal parameters and $\alpha_1^*, ..., \alpha_n^*$ their canonical names, and let $x_0, ..., x_m$ be its parameters from A' and $f_0, ..., f_m$ their names, as above. Let β be the least upper bound of $\{\mathsf{dom}(f_i) : \mathsf{dom}(f_i) < \kappa\}$. By the definition of B, there is some regular cardinal α such that $\beta < \alpha < \kappa$ where $A'(\alpha)$ is a subset of X but disjoint from Y or vice versa. Without loss of generality, we can assume it is a subset of X and disjoint from Y. Let $x \in A'(\alpha)$. Then g maps x to some y in $A'(\delta)$ for some regular cardinal δ strictly greater or strictly less than α. That is:

$$M[G] \vDash \varphi(\alpha_0, ..., \alpha_n, x_0, ..., x_m, x, y)$$

$$M[G] \vDash \forall z \in HOD(A')(\varphi(\alpha_0, ..., \alpha_n, x_0, ..., x_m, z, y) \to z = x)$$

Let f_x and f_y be names of x and y respectively, as above. It follows that for some $p \in \mathbb{P}$:

$$M \vDash p \Vdash \varphi(\alpha_0^*, ..., \alpha_n^*, f_0, ..., f_m, f_x, f_y)$$

$$M \vDash p \Vdash \forall z(\varphi(\alpha_0^*, ..., \alpha_n^*, f_0, ..., f_m, z, f_y) \to z = f_x)$$

Working in M, we know that $|\mathsf{dom}(p(\alpha))| < \alpha$. So let β be the least ordinal greater than every ordinal in $\mathsf{dom}(p(\alpha)) \cap \alpha$ and let π be the automorphism of $\mathbb{P}$ that swaps any $q \in \mathbb{P}$ with the unique $r \in \mathbb{P}$ which is exactly like q except that $r(\alpha)(\beta) = 0$ just in case $q(\alpha)(\beta) = 1$ and $r(\alpha)(\beta) = 1$ just in case $q(\alpha)(\beta) = 0$. As is standard, π extends to an automorphism of the whole forcing language so that in particular:

$$\pi(p) \Vdash \varphi(\pi(\alpha_0^*), ..., \pi(\alpha_n^*), \pi(f_0), ..., \pi(f_m), \pi(f_x), \pi(f_y))$$

By definition of π, it is the identity function on p, $\alpha_0^*, ..., \alpha_n^*, f_0, ..., f_m$, and f_y. So, we have:

$$p \Vdash \varphi(\alpha_0^*, ..., \alpha_n^*, f_0, ..., f_m, \pi(f_x), f_y)$$

But $\pi(f_x)$ and f_x will disagree on β: it will always be the case that either $\pi(f_x)$ contains β and f_x doesn't or vice versa. In addition, $\pi(f_x)$ will always be in A'. So:

$$p \Vdash \pi(f_x) \neq f_x \wedge \pi(f_x) \in A'$$

This contradicts the fact that p thinks φ only maps one element in $HOD(A')$ to f_y. □

> **Theorem**: There are individually Henkin-stable abstraction principles that are jointly inconsistent.

First we prove a very general independence result.[31]

Lemma 1 (ZFC)**.** *For any finite first-order language $\mathcal{L}$, there is a φ in the language of second-order logic over $\mathcal{L}$ such that for any infinite model of $\mathcal{L}$, there is an expansion to a model of second-order logic that makes φ true and one that makes φ false.*

Proof. For simplicity we work with the language of pure second-order logic. It will be easy to see how the proof generalises to the language of second-order logic over a given first-order language $\mathcal{L}$.

Let $\mathcal{L}^2_{PA}$ be the language of second-order logic together with a predicate N intended to express "is a natural number" and $<$ intended to express the less-than relation on the natural numbers. The usual proof of the diagonal lemma can be reconstructed to show that for any formula $\psi(x) \in \mathcal{L}^2_{PA}$ with

[31] Thanks to Kameryn Williams for suggesting use of the diagonal lemma to prove this kind of result.

only x free there is a sentence $\varphi \in \mathcal{L}^2_{PA}$ such that second-order logic proves that whenever $\langle D, R\rangle$ is a standard model of second-order Peano Arithmetic ($\mathsf{PA2}$):

$$[\varphi \leftrightarrow \psi(``\varphi")]^{\langle D,R\rangle}$$

where $\chi^{\langle D,R\rangle}$ is the result of replacing occurrences of $N(x)$ in χ with $x \in D$ and $x < y$ with $R(x, y)$.[32]

Let $\psi(x)$ be the claim that (1) there is a three-place relation T coding a pairing function on the first-order domain, (2) there is a two-place relation S coding a collection of classes,[33] (3) S contains a class of all and only the Ns and, relative to T, a class coding $<$, (4) the model of $\mathcal{L}^2_{PA}$ whose first-order domain is the class of all objects and whose second-order domain is given by S relative to T—which we can denote $\langle O, S, N, <\rangle$—satisfies all instances of the comprehension principles in $\mathcal{L}^2_{PA}$, the axiom of global choice, etc,[34] (5) $\langle O, S, N, <\rangle$ falsifies the formula x, and finally (6) the notion of being a standard model of $\mathsf{PA2}$ is absolute for $\langle O, S\rangle$.[35]

Let φ be the formula on the left-hand-side of the diagonal lemma for $\psi(x)$ and let $\mathcal{M}$ be an infinite standard model of second-order logic. We will first show that $\exists D, R(\mathsf{PA2} \wedge \varphi)^{\langle D,R\rangle}$ is true in $\mathcal{M}$. Since $\mathcal{M}$ is infinite, let $\langle D, R\rangle$ be a standard model of $\mathsf{PA2}$. Now suppose $\neg\varphi^{\langle D,R\rangle}$. Then by the definition of φ, all T and S satisfying conditions (1), (2), (3), (4), and (6) will fail to satisfy (5). In other words, "φ" will be true in all such models.

But by a simple Löwenheim-Skolem argument we can get an elementary substructure of $\mathcal{M}$ with the same first-order domain whose classes and relations are coded by some S relative to some T satisfying (1), (2), (3), (4), and (6). But in any such model, "φ" will be false because it's false in $\mathcal{M}$. Contradiction. So, $\varphi^{\langle D,R\rangle}$ must hold after all.

Now, we will show that "$\exists D, R(\mathsf{PA2} \wedge \varphi)^{\langle D,R\rangle}$" is false in some submodel of $\mathcal{M}$. Let D and R be such that $(\mathsf{PA2} \wedge \varphi)^{\langle D,R\rangle}$ and by the diagonal lemma, let S and T satisfy conditions (1)-(6). We will show that the submodel $\langle O, S, D, R\rangle$ is precisely the model we need. For suppose it satisfied "$\exists D', R'(\mathsf{PA2} \wedge \varphi)^{\langle D',R'\rangle}$". It would follow that $\langle O, S, D', R'\rangle$ satisfies "φ" and by condition (6), that $\langle D', R'\rangle$ is really a standard model of $\mathsf{PA2}$. Thus, $\langle O, S, D, R\rangle$ would be isomorphic to $\langle O, S, D', R'\rangle$ and it would follow that

[32] The claim that $\langle D, R\rangle$ is a standard model of $\mathsf{PA2}$ is just $\mathsf{PA2}^{\langle D,R\rangle}$.

[33] A class X is said to be a member of S when there is some x for which $X = \{y : S(x, y)\}$.

[34] Because there are only class-many classes in S, we can explicitly define satisfaction for $\mathcal{L}^2_{PA}$ formulas in $\langle O, S, N, <\rangle$.

[35] In other words, for classes X and Y in S, $\langle O, S\rangle \vDash \mathsf{PA2}^{\langle X,Y\rangle}$ just in case $\mathsf{PA2}^{\langle X,Y\rangle}$.

$\langle O, S, D, R\rangle$ satisfies "φ" contradicting our assumption (5) that "φ" is false in $\langle O, S, D, R\rangle$. $\square$

Our theorem is now immediate. Let φ be as in the proof of lemma 1. Given any infinite first-order model $\mathcal{M}$ we know there is a second-order expansion (the standard model over $\mathcal{M}$) that makes $\exists D, R(\mathsf{PA2} \wedge \varphi)^{D,R}$ true and one that makes it false. Now consider the pair of abstraction principles:

$$\#X = \#Y \leftrightarrow (\forall x(Xx \equiv Yx) \vee \exists D, R(\mathsf{PA2} \wedge \varphi)^{D,R}) \qquad (\mathsf{ABS}_0)$$

$$\#X = \#Y \leftrightarrow (\forall x(Xx \equiv Yx) \vee \neg\exists D, R(\mathsf{PA2} \wedge \varphi)^{D,R}) \qquad (\mathsf{ABS}_1)$$

It follows that they are both Henkin-stable, although jointly inconsistent.

ACKNOWLEDGEMENTS: Many thanks to Saharon Shelah, Harvey Friedman, Tim Bays, Kameryn Williams, and two anonymous referees for helpful discussion and suggestions.

References

Boolos, G. [1984]. “To be is to be a value of a variable (or to be some values of some variables)”, *Journal of Philosophy 81*; ; reprinted in Boolos [1998], 54-72.

Boolos, G. [1985]. “Nominalist platonism”, *Philosophical Review 94*, 327-344; reprinted in Boolos [1998], 73-87.

Boolos, G. [1987]. “The consistency of Frege’s Foundations of arithmetic” in *On being and saying: Essays for Richard Cartwright*, edited by Judith Jarvis Thompson, Cambridge, Massachusetts, The MIT Press, 3-20; reprinted in Boolos [1998], 54-72.

Boolos, George [1997]. “Is Hume’s principle analytic?”, in *Language, thought, and logic*, edited by Richard Heck, Jr., Oxford, Oxford University Press, 245-261; reprinted in Boolos [1998], 301-314.

Boolos, George [1998]. *Logic, logic, and logic*, Cambridge, Massachusetts, Harvard University Press.

Burgess, John P. [1984]. Review of Wright [1983], *The Philosophical Review 93*, 638-640.

Dummett, M. [1991]. *Frege: Philosophy of mathematics*, Cambridge, Massachusetts, Harvard University Press.

Dummett, M. [1998]. “Neo-Fregeans: in Bad Company?”, in *The philosophy of mathematics today*, edited by M. Schirn (editor), Oxford, Oxford University Press, 369-387.

Ebert, Philip and Stewart Shapiro [2009]. “The good, the bad, and the ugly” *Synthese 170*, 415-441.

Feferman, Solomon [2006], “Predicativity”, in *The Oxford handbook of philosophy of mathematics and logic*, edited by Stewart Shapiro, Oxford, Oxford University Press, 590-624.

Fine, Kit [2002], *The limits of abstraction*, Oxford, Oxford University Press.

Florio, Salvatore and Øystein Linnebo [2020]. “Critical plural logic” *Philosophia Mathematica 28*, 172-203.

Frege, Gottlob [1884]. *Die Grundlagen der Arithmetik*, Breslau, Koebner; *The foundations of arithmetic*, translated by J. Austin, second edition, New York, Harper, 1960.

Frege, Gottlob [1893]. *Grundgesetze der Arithmetik 1*, Olms, Hildescheim; *Basic laws of arithmetic*, translated by Philip A. Ebert and Marcus Rossberg, Oxford, Oxford University Press, 2013.

Fritz, Peter and Jeremy Goodman [2017]. "Counting incompossibles" *Mind 126*, 10631108.

Hale, Bob [1987]. *Abstract objects*, Oxford, Basil Blackwell.

Hale, Bob [2000]. "Abstraction and set theory", *Notre Dame Journal of Formal Logic 41*, 379-398.

Hale, Bob [2010]. "The bearable lightness of being", *Axiomathes 20*, 399-422.

Hale, Bob [2013]. 11Properties and the interpretation of second-order logic", *Philosophia Mathematica (3) 21*, 133-156.

Hale, Bob [2013a]. *Necessary beings: an essay on ontology, modality, and the relations between them*, Oxford, Oxford University Press.

Hale, Bob, and Crispin Wright [2001]. *The reason's proper study*, Oxford, Oxford University Press.

Heck, R. [1992]. "On the consistency of second-order contextual definitions", *Nous 26*, 491-494.

Landman, Fred [1989], "Groups, I", *Linguistics and philosophy 12*, 559-605; "Groups II", *Linguistics and philosophy 12*, 723-744.

Lévy, A., [1969]. "The definability of cardinal numbers", in *Foundations of mathematics: symposium papers commemorating the sixtieth birthday of Kurt Gödel*, edited by Jack J. Bulloff, Thomas C. Holyoke, and S. W. Hahn, Berlin, Spring-Verlag, 15-38.

Linnebo, Øystein [2004], "Predicative fragments of Frege arithmetic", *Bulletin of Symbolic Logic 10*, 153-174.

Linnebo, Øystein [2009] "Introduction", Special issue devoted to the Bad Company issue, *Synthese 170*, 321-329.

Linnebo, Øystein [2010], "Pluralities and sets", *Journal of Philosophy 107*, 144-164.

Maddy, Penelope [1983], "Proper classes", *Journal of Symbolic Logic 48*, 113-139.

Maddy, Penelope [2007]. *Second philosophy: a naturalistic method*, Oxford, Oxford University Press.

Moore, Gregory H. [1982], *Zermelos axiom of choice: its origins, development, and influence*, New York, Springer-Verlag.

Poincaré, H. [1906], "Les mathmatiques et la logique", *Revue de Mtaphysique et de Morale 14*, 294-317.

Priest, Graham [2006]. *In contradiction: a study of the transconsistent*, second, revised edition, Oxford, Oxford University Press, 2006; first edition, Dordrecht, Martinus Nijhoff Publishers, 1987.

Rayo, A. & Yablo, S. [2001], "Nominalism Through de-Nominalization", *Noûs 35*, 74-92.

Resnik, M. [1988], "Second-order logic still wild", *Journal of Philosophy 85*, 75-87.

Roberts, S. [ms], "Reflection principles: a survey".

Roberts, S. [forthcoming], "Pluralities as nothing over and above", *Journal of Philosophy.*

Shapiro, Stewart [1991]. *Foundations without foundationalism: a case for second-order logic*, Oxford, Oxford University Press.

Shapiro, Stewart [2000], "Frege meets Dedekind: a neo-logicist treatment of real analysis", *Notre Dame Journal of Formal Logic 41*, 335-364.

Shapiro, Stewart [2003]. "Prolegomenon to any future neo-logicist set theory: abstraction and indefinite extensibility", *British Journal for the Philosophy of Science 54*, 59-91.

Shapiro, Stewart [2018]. "Properties and predicates, objects and names: impredicativity and the axiom of choice", in *Being necessary: themes of ontology and modality from the work of Bob Hale*, edited by Ivette Fred-Rivera and Jessical Leech, Oxford, Oxford University Press, 2018, pp, 92-110.

Tennant, Neil [1987]. *Anti-realism and logic*, Oxford, Oxford University Press.

Uzquiano, Gabriel [2009]. “Bad company generalized”, *Synthese 170*, 331-347.

Walsh, Sean [2012]. “Comparing Peano arithmetic, Basic Law V, and Hume’s Principle”, *Annals of Pure and Applied Logic 163* (2012), 1679-1709.

Weir, Alan [1998]. “Nave set theory is innocent”, *Mind 107*, 763-798.

Weir, Alan [2003]. “Neo-Fregeanism: an embarrassment of riches”, *Notre Dame Journal of Formal Logic 44*, 13-48.

Weyl, H. [1918], *Das Kontinuum*, Verlag von Veit & Comp, Leipzig; translated by S. Pollard and T. Bole as as *The Continuum*, Dover, 1994.

Whitehead, A. N., and B. Russell [1910], *Principia Mathematica 1*, Cambridge, Cambridge University Press.

Williamson, Timothy [2013], *Modal Logic as Metaphysics*, Oxford, Oxford University Press.

Wright, Crispin [1983] *Frege’s conception of numbers as objects*, Aberdeen, Aberdeen University Press.

Wright, Crispin [1997]. “On the philosophical significance of Freges theorem”, in *Language, thought, and logic*, edited by Richard Heck, Jr., Oxford, Oxford University Press, 201-244; reprinted in Hale and Wright [2001], 272-306.

Wright, Crispin [2004]. “Warrant for nothing (and foundations for free)”, *Proceedings of the Aristotelian Society, Supplementary Volume 78*, 167-212.