Cardinality and Acceptable Abstraction

Roy T. Cook and Øystein Linnebo

Abstract   It is widely thought that the acceptability of an abstraction principle is a feature of the cardinalities at which it is satisfiable. This view is called into question by a recent observation by Richard Heck. We show that a fix proposed by Heck fails but we analyze the interesting idea on which it is based, namely that an acceptable abstraction has to "generate" the objects that it requires. We also correct and complete the classification of proposed criteria for acceptable abstraction.

1   Introduction

An *abstraction principle* (or simply abstraction) is a principle of the form

$$@F = @G \leftrightarrow F \sim G, \tag{Σ}$$

where F and G are second-order variables and $\sim$ denotes an equivalence relation. Some such principles appear acceptable—for instance Hume's Principle:

$$\#F = \#G \leftrightarrow F \approx G, \tag{HP}$$

where $F \approx G$ is the second-order formalization of the claim that F and G are equinumerous, whereas others clearly are not—for instance Frege's Basic Law V:

$$\S F = \S G \leftrightarrow F \equiv G, \tag{BLV}$$

where $F \equiv G$ is the second-order formalization of the claim that F and G are coextensive. Providing a philosophically principled and technically adequate account of what, exactly, separates the "good" principles from the "bad" is known as the *bad company problem*. To solve this problem, defenders of the abstractionist program need to provide a well-motivated criterion of acceptable abstraction.

The dominant approach to the bad company problem has treated the acceptability of an abstraction as a feature of the cardinalities at which it is satisfiable. This

© 2018 by University of Notre Dame   10.1215/00294527-2017-0012

approach is called into question by a recent observation due to Richard Heck,[1] who observes that there are abstractions that seem "fishy" (to use Heck's term) but are satisfiable at exactly the same cardinalities as (HP), which is our paradigm case of a good abstraction.

After explaining Heck's observation, we demonstrate that a fix he proposes still allows "fishy" principles to slip through the net. Next, we discuss how Heck's observation affects previous classifications of possible criteria for acceptable abstraction (Section 2). This results in a complete classification of the extant criteria with regard to their logical strength (Section 3). Finally, we examine whether there might be better ways to develop the idea underlying Heck's proposed fix, namely, that an acceptable abstraction has to *generate* the objects that it requires to be satisfiable (Sections 4 and 6).

A central theme throughout the article will be that previous investigations of acceptable abstraction have been excessively focused on the cardinalities of the domains at which an abstraction principle is satisfiable and have paid inadequate attention to the *way* in which the principle requires the domain to be of the relevant cardinality.

2 Slippery Fish and a Failed Attempt to Catch Them

Before we can explain Heck's observation, we need some explanations and definitions.

Definition 1   An abstraction is *purely logical* if and only if the equivalence relation that figures on its right-hand side is expressible in the vocabulary of pure second- (or higher-) order logic. An abstraction is κ-*satisfiable* if and only if it is satisfiable at a domain of cardinality κ. An abstraction is *unbounded* if and only if it is κ-satisfiable for an unbounded sequence of cardinals κ. An abstraction is *stable* if and only if there is a κ such that it is λ-satisfiable for all $\lambda \geq \kappa$; the least such κ is the *stabilization point*. An abstraction is *strongly stable* if and only if: there is a κ such that it is λ-satisfiable if and only if $\lambda \geq \kappa$. Finally, two abstraction principles are *cardinality equivalent* if and only if: for any cardinal κ, one is κ-satisfiable if and only if the other is.

Heck [3] is favorably disposed to stability (or perhaps strong stability—he wavers a bit here; see pp. 232–33) as a criterion of acceptable abstraction. However, he observes that this criterion lets through certain abstractions that are intuitively fishy. For example, let $\Phi_{\geq \kappa}$ be the second-order formalization of the claim that there are at least κ many objects for some κ that is characterizable in pure second-order (or higher-order) logic. Now consider:

$$@F = @G \leftrightarrow (F \equiv G \vee \Phi_{\geq \kappa}). \qquad \qquad (\text{Fishy}_{\geq \kappa})$$

Although this principle is strongly stable with stabilization point κ, it seems unacceptable because it only provides a single abstract object (the one that is the abstract of any concept). In short, (Fishy$_{\geq \kappa}$) only requires the universe to be sufficiently populous so as to avoid being impaled on Russell's paradox, but does not actually *provide* any of the objects required to achieve this.

Compare the following two ways of arguing that there are infinitely many objects. One option is to invoke (HP), which allows us to prove the existence of zero and of a successor of any given number, thus ensuring that there are infinitely many numbers

(a striking mathematical fact known as Frege's theorem). Another option is to invoke (Fishy$_{\geq \aleph_0}$), which, loosely put, says that either there are at least $\aleph_0$-many objects or else Basic Law V holds. (More carefully, if the domain is finite, then (Fishy$_{\geq \aleph_0}$) is equivalent to (BLV).) Since an inconsistent alternative is no alternative at all, this is simply a postulation that there are infinitely many objects, dressed up in the garb of an abstraction principle. Whereas the first option appears to generate infinitely many objects by honest toil, the second seems to rely on petty theft.

Heck proposes that the fishy principles be caught and eliminated by adopting the following modified criterion of acceptability.[2]

Definition 2 An abstraction Σ is *conservatively stable* if and only if (i) Σ is strongly stable with stabilization point κ, and (ii) in any model of Σ, there are at least κ abstracts of the sort characterized by Σ.

However, the following theorem spells trouble.

Theorem 1 *Let Σ be any abstraction principle:*

$$@F = @G \leftrightarrow F \sim G. \tag{Σ}$$

Then there is a purely logical abstraction principle Σ^+ which (i) is cardinality equivalent to Σ and (ii) generates κ abstracts in any model of cardinality κ.

Proof Let $\mathcal{R}(\Sigma)$ be the *ramsification* of Σ.[3] Then let

$$@^+ F = @^+ G \leftrightarrow \big((\mathcal{R}(\Sigma) \wedge F \cong G) \vee (\neg \mathcal{R}(\Sigma) \wedge F \equiv G)\big), \tag{Σ^+}$$

where $F \cong G$ is the second-order formalization of the claim that either F and G are coextensive singletons or neither F nor G is a singleton concept.[4]

Let D be a domain of cardinality κ. Assume that Σ is not satisfiable at D. Then $\neg \mathcal{R}(\Sigma)$ is true at D, whence it follows by cardinality considerations that Σ^+ is also not satisfiable at D. Assume instead that Σ is satisfiable at D. Then $\mathcal{R}(\Sigma)$ is true at D. So Σ^+ will provide κ many abstracts at D, since it will provide κ many distinct abstracts for each of the κ many singleton concepts on D. $\square$

A corollary is now immediate.

Corollary 1 *Assume that Σ is strongly stable. Then Σ^+ is purely logical and conservatively stable.*

A problem with Heck's proposed fix now becomes apparent. Assume that Σ is strongly stable but fishy in the way discussed above. Heck's revised condition of acceptability rejects Σ as unacceptable because it fails to satisfy condition (ii) of the definition of conservative stability. However, the corollary shows that Σ can be used to define another abstraction Σ^+ which satisfies the revised criterion of acceptability. But Σ^+ seems no less fishy than Σ itself. To see this more clearly, let us return to the example of (Fishy$_{\geq \aleph_0}$), which seems unacceptable because it simply demands that the universe be infinite. Consider now the associated abstraction (Fishy$_{\geq \aleph_0}$)$^+$. This in effect says: either the universe is finite and Basic Law V holds or else the universe is infinite and we obtain distinct abstracts for each singleton concept (and a single "dummy" abstract for all other concepts). Since Russell's paradox shows the former disjunct to be unacceptable, (Fishy$_{\geq \aleph_0}$)$^+$ in effect just demands that the universe be infinite—with singleton abstraction just coming along for the ride. But if it was unacceptable merely to lay down a demand of infinity in the form of (Fishy$_{\geq \aleph_0}$),

then it is no less unacceptable to lay down this demand coupled with a benign form of singleton abstraction.

Since the acceptability of Σ stands or falls with that of Σ^+, this leaves us with just two choices. One option is to conclude that the fishy principles which Heck has called to our attention are more slippery than he realizes and that many of them still slip through his revised net. What went wrong? Heck's proposed revision is based on the intuitive idea that the objects required by an abstraction principle have to be properly "generated." This idea is then explicated as the requirement that the objects required by an abstraction principle have to be in the range of its abstraction operator, at least at suitable cardinalities. The moral of our discovery is, on this view, that this explication fails to ensure permissible abstraction. We are therefore left, at least for the time being, without an acceptable solution to this version of the bad company problem, and with a feeling that perhaps we should look for alternative means of formulating an abstractionist account of mathematics that will avoid this particulary vexing issue altogether.[5]

Another option is to conclude that our initial rejection of Heck's fishy principles was too rash. One way to develop such a view would be to emphasize that abstraction principles should not be adjudged acceptable or unacceptable piecemeal, but that we should instead consider collections of abstraction principles (abstraction *theories*) as a whole (see Cook [1] and Fine [2] for expressions of such views). On this sort of approach, the result just proved would not be taken to show that an abstraction principle Σ^+ is fishy because Σ is, but rather to show that Σ is acceptable since Σ^+ is (i.e., since Σ^+ is conservatively stable). Thus, viewed on its own, (Fishy$_{\geq \aleph_0}$) encodes an "ontological obligation"—namely, to provide the infinitely many objects whose existence it requires—that it does not meet. (Fishy$_{\geq \aleph_0}$)$^+$, however, fulfills this obligation, by providing enough objects to "fill" any model of the sort of size required by (Fishy$_{\geq \aleph_0}$). In short, on this sort of view we need not require that abstraction principles be conservatively stable, since we know that for any strongly stable principle, there will be a corresponding conservatively stable principle that will do the "filling."[6]

We make no attempt at a final resolution of these difficult philosophical issues here, but content ourselves with having clarified what is at stake.

3 A Classification Corrected and Completed

Heck's observation shows that a fresh look is needed at some of the previously proposed criteria of acceptable abstraction.

Weir [9, pp. 21–22] considers two notions of conservativeness. To explain them, we first need a definition. Given a formula Φ and a (possibly complex) unary predicate $\Psi(x)$, the *relativization of Φ to $\Psi(x)$*, written $\Phi^{\Psi(x)}$, is defined recursively by

  (i) translating each atomic predication $Ft_1, \ldots, t_n$ as $Ft_1, \ldots, t_n \wedge \Psi(t_1) \wedge \cdots \wedge \Psi(t_n)$;
  (ii) restricting all (first- and second-order) quantifiers to Ψ;
  (iii) letting the translation commute with the truth-functional connectives.[7]

For reasons that will shortly become obvious, we rebaptize Weir's two notions as follows.[8]

Definition 3   Let Σ be an abstraction.

(i) Σ is *strongly conservative* if and only if for any theory T and sentence ϕ we have

$$T^{\neg \exists F(x = @F)}, \Sigma \models \phi^{\neg \exists F(x = @F)} \quad \Rightarrow \quad T \models \phi;$$

(ii) Σ is *weakly conservative* if and only if for any theory T and sentence ϕ we have

$$T^{P(x)}, \Sigma \models \phi^{P(x)} \quad \Rightarrow \quad T \models \phi,$$

where P is a new one-place predicate.

Both notions of conservativeness admit of very illuminating model-theoretic reformulations.

Theorem 2

(a) Σ *is weakly conservative if and only if for every $\mathcal{L}$-model $\mathcal{M}$, there is an $\mathcal{L}_@$-model $\mathcal{N}$ such that $\mathcal{M} \subseteq \mathcal{N}$ and $\mathcal{N} \models \Sigma$.*

(b) Σ *is strongly conservative if and only if for every $\mathcal{L}$-model $\mathcal{M}$, there is an $\mathcal{L}_@$-model $\mathcal{N}$ such that $\mathcal{M} \subseteq \mathcal{N}$, $\mathcal{N} \models \Sigma$, and $\mathcal{N} \setminus \mathcal{M}$ contains all and only the @-abstracts.*

Proof   We begin by observing that it is straightforward to rewrite the conservativeness requirements as follows:

Σ is weakly conservative if and only if for any theory T we have

$$\exists \mathcal{M} \models T \Rightarrow \exists \mathcal{M} \models T^{P(x)}, \Sigma. \tag{1}$$

Σ is strongly conservative if and only if for any theory T we have

$$\exists \mathcal{M} \models T \Rightarrow \exists \mathcal{M} \models T^{\neg \exists F(x = @F)}, \Sigma. \tag{2}$$

(In both implications, the antecedents are understood to quantify over $\mathcal{L}$-models, and the consequents, over $\mathcal{L}_@$-models.)

The observation is proved by considering the contrapositives of the implications found in Definition 3 and the theory $T, \neg\phi$.

To prove (a), assume that Σ is weakly conservative, and consider an $\mathcal{L}$-model $\mathcal{M}$. Let $\mathcal{L}^M$ be the result of adding to $\mathcal{L}$ a constant for each element of the domain M of $\mathcal{M}$. Let T be the elementary diagram of $\mathcal{M}$, that is, the $\mathcal{L}^M$-theory consisting of all the literals true in $\mathcal{M}$. Since T is quantifier-free, we have $T^{P(x)} = T$. Thus, by (1), there is a model $\mathcal{N}$ of T, Σ. Let $\bar{\mathcal{N}}$ be the reduct of $\mathcal{N}$ to the language $\mathcal{L}_@$.[9] Since $\mathcal{N} \models T$, we may assume that $\mathcal{M} \subseteq \bar{\mathcal{N}}$. Since we also have $\bar{\mathcal{N}} \models \Sigma$, this is our desired model. For the other direction, assume the right-hand side of (a). It suffices to prove (1). So assume that there is an $\mathcal{M}$ such that $\mathcal{M} \models T$. Applying the right-hand side of (a) and interpreting P as true of all and only the elements of M yields the desired $\mathcal{N}$ such that $\mathcal{N} \models T^{P(x)}, \Sigma$.

To prove (b), assume first that Σ is strongly conservative, and consider an $\mathcal{L}$-model $\mathcal{M}$. Let $\mathcal{L}^M$ be as before, where $\{a_\gamma\}_{\gamma < \kappa}$ is an enumeration of all the elements of $\mathcal{M}$. Let T be the theory that consists of the elementary diagram of $\mathcal{M}$ and the sentence

$$\forall x \left(\bigvee_{\gamma < \kappa} x = a_\gamma \right). \tag{3}$$

By (2), there is a model $\mathcal{N}$ of $T^{\neg \exists F(x = @F)}, \Sigma$. (Intuitively, T is true in $\mathcal{N}$ when interpreted as being solely about the non-@-abstracts.) Since $\mathcal{N}$ thus models the elementary diagram of $\mathcal{M}$, we may assume $\mathcal{M} \subseteq \mathcal{N}$. Next, by clause (i) of our

definition of relativization, it follows that each a_γ is a non-@-abstract. This means that *all* @-abstracts are contained in $\mathcal{N} \setminus \mathcal{M}$. (Intuitively, there are no "Caesar cases," i.e., cases where an @-abstract is identified with one of the "old" objects already in $\mathcal{M}$.) Finally, since $\mathcal{N} \models (3)^{\neg \exists F(x = @F)}$, every non-@-abstract is in $\mathcal{M}$, which is to say that $\mathcal{N} \setminus \mathcal{M}$ contains *only* @-abstracts. As before, let $\bar{\mathcal{N}}$ be the reduct of $\mathcal{N}$ to the language $\mathscr{L}_@$. Since $\bar{\mathcal{N}} \models \Sigma$, we are done.

For the other direction, it suffices to prove that the right-hand side of (b) entails (2). The proof of this is straightforward.   □

Corollary 2   *Strong conservativeness implies, but is not implied by, weak conservativeness.*

Proof   The implication is obvious in light of Theorem 2. Heck's observation can be used to show that the converse implication fails. For instance, let T be the first-order sentence:

$$(\exists x)(\exists y)\big(x \neq y \land (\forall z)(x = z \lor y = z)\big).$$

Then $\{T^{P(x)}, (\text{Fishy}_{\geq \aleph_0})\}$ has a model, but $\{T^{\neg \exists F(x = @F)}, (\text{Fishy}_{\geq \aleph_0})\}$ does not.
                                                                                            □

Having cleared up the relation between the two notions of conservativeness, drawing crucially on Heck's observation, some confessions are in order. Linnebo [5] and Cook [1] fail to heed the distinction between strong and weak conservativeness, which results in some statements in both of these works—when taken literally, as behooves serious scientific statements—being false.[10] *Nostra culpa.*

However, the errors are superficial and easily cleared up. Provided that [5, Definition 2.1] is changed so as to attach the label "conservativeness" to what we are now calling weak conservativeness, rather than to strong conservativeness as was in fact done, all the claims and proofs are okay (which some of them would otherwise not be). So although Linnebo [5] confuses definitions, he does so in an entirely systematic way. Likewise, Cook [1] mistakenly claims that weak and strong conservativeness are equivalent and provides the definition of strong conservativeness when (in retrospect) he clearly meant weak conservativeness, but the arguments are correct if a uniform substitution of the same sort is performed—hence little of philosophical substance is affected.

With the bifurcation of conservativeness into a strong and a weak version comes an analogous bifurcation of a condition known as *irenicity*.

Definition 4   An abstraction is *weakly irenic* if and only if it is weakly conservative and cosatisfiable with any other weakly conservative abstraction. An abstraction is *strongly irenic* if and only if it is strongly conservative and cosatisfiable with any other strongly conservative abstraction.

Previous definitions of irenic in the literature (e.g., [1], [5], [9]) are also systematically ambiguous. Disambiguating them along the lines sketched above makes them equivalent to weak irenicity. The main result of [5]—which like all the other results still stands once the above clarifications are enforced—can now be depicted by the

following diagram, where all single arrows mark strict implication.
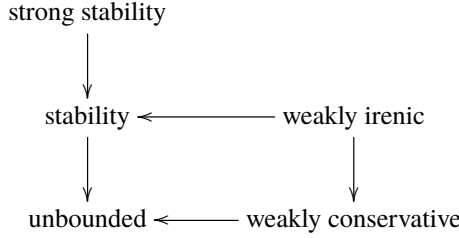
Moreover, under the simplifying assumption of pure logicality—which we will henceforth adopt without explicit mention—the horizontal dimension of the previous diagram collapses, as we become able to prove the converse to each of the above horizontal arrows.[11]

However, the distinctions prompted by Heck's discovery shows that the above classifications are incomplete: Where do the *strong* notions of conservativeness and irenicity fit in? To keep things manageable, we work under the simplifying assumption of pure logicality, which most philosophers sympathetic to this approach to the bad company problem are anyway prepared to grant (see, e.g., [1], [3]). Our result (which we prove in Section 5) is depicted by the following diagram.
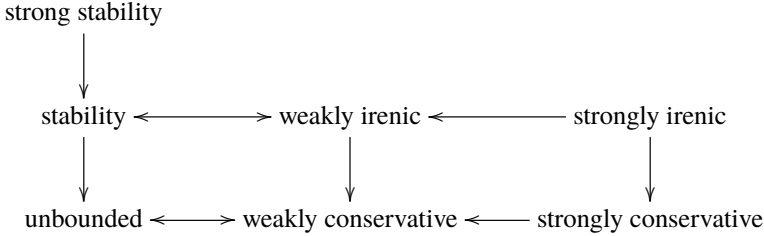
4 Critical Fullness

Recall Heck's proposed notion of conservative stability, which adds to strong stability the requirement that any model of the abstraction should contain κ many abstracts, where κ is the stabilization point. Let us try to analyze the idea(s) behind Heck's proposal to see if we might extract something that works better than the original proposal (which was proved unsuccessful in Section 2). One such idea will be identified and examined in this section, and another, in the final section.

According to the former idea, when an acceptable abstraction requires the domain to be larger than before the abstraction, this should be a result of new abstracts that are actually generated by the abstraction. To make this idea—which we call *full-ness*—precise, we need to attend not only to the stabilization point of a strongly stable abstraction but to a larger class of similar points.

Definition 5   A cardinal κ is a *critical point* of an abstraction Σ if and only if Σ is κ-satisfiable, and there is a $\gamma < \kappa$ such that, for all λ such that $\gamma \leq \lambda < \kappa$, Σ is not λ-satisfiable.[12] An abstraction Σ is *critically full* if and only if, for each critical point κ of Σ, any model of Σ of size κ contains κ abstracts of the sort characterized by Σ. An abstraction Σ is *Heck conservative* if and only if it is strongly conservative and critically full. An abstraction is *Heck irenic* if and only if it is strongly irenic and

critically full.[13] An abstraction is *Heck stable* if and only if it is strongly stable and critically full.

(Thus, the result of appending "Heck" to an adjective X means: strongly X and critically full.)

Our main result is depicted by the following diagram, which we prove in the next section.
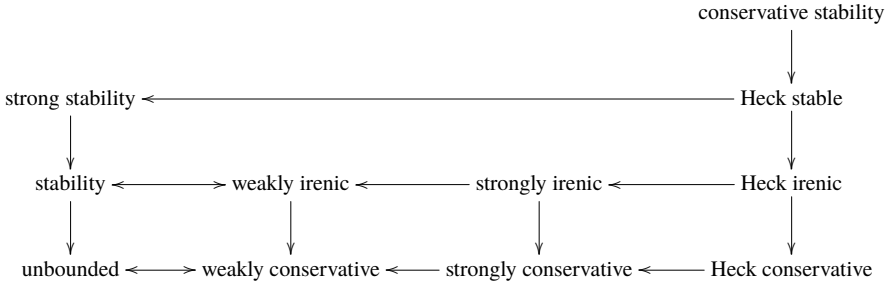
Of course, our result in Section 2 shows strong conservativeness to be insufficient to ensure that an abstraction principle is acceptable, and consequently, so are the weaker requirements in the rightmost column. Might an even stronger fullness requirement help? The most natural option would be to require that *any* model of Σ contain κ many abstracts of the sort characterized by Σ, where κ is the cardinality of the model. However, this fullness requirement would be implausible; in particular, it would be violated by (HP), which generates "only" $\aleph_0$ abstracts on a domain of cardinality $\aleph_1$.

5 Proofs

The leftmost half of the diagram is the main result of [5] (again, with the above corrections in place). First, all of the implications except one are straightforward consequences of the definitions and are left to the reader. The one exception is handled by the following lemma.

Lemma 1   *Strong irenicity implies weak irenicity.*

Proof   This requires us to show that an abstraction Σ compatible with each strong conservative is also compatible with each member of the (provably) larger family of weak conservatives. We show this by appealing to Theorem 1, which ensures that for each weak conservative Γ, there corresponds a cardinality-equivalent strong conservative Γ^+. Since we are assuming that Σ is purely logical, compatibility with Σ is preserved under cardinality equivalence, which completes our proof.   □

Returning to our main proof, we next take care of the horizontal nonimplications.

Lemma 2   *No entry from the first or second column implies any entry from the third column.*

Proof   It suffices to observe that strong stability does not imply strong conservativeness. This is immediate from the Heck's observation: (Fishy$_{\geq\aleph_0}$) is strongly stable but not strongly conservative.   □

Lemma 3 *No entry from the third column implies any entry from the fourth column.*

Proof It suffices to observe that strong irenicity does not imply Heck conservativeness. A slight complication of Heck's idea will work here as well. Consider

$$@F = @G \leftrightarrow \left((\Phi_{<\aleph_0} \wedge X \equiv Y) \vee (\Phi_{\aleph_0}) \vee (\Phi_{\geq \aleph_1} \wedge F \cong G) \right), \qquad (\Sigma_1)$$

where $\Phi_{<\aleph_0}$ is the second-order formalization of the claim that the universe is finite, $\Phi_{\aleph_0}$ is the second-order formalization of the claim that the universe is countably infinite, and $F \cong G$ is as before. Σ_1 is equivalent to (BLV) on finite domains, generates a single object on countably infinite domains, and generates a distinct object for each singleton concept (plus a "dummy" object) on uncountably infinite domains. Hence Σ_1 is strongly irenic but not Heck conservative. □

We now handle the vertical (and, in so doing, the remaining diagonal) nonimplications.

Lemma 4 *No entry on the bottom row implies any entry on the next row up.*

Proof It suffices to show that Heck conservativeness does not imply weak irenicity. Consider

$$@F = @G \leftrightarrow \left((\Phi_{\text{Succ}} \wedge X \equiv Y) \vee (\Phi_{\text{Lim}} \wedge F \cong G) \right), \qquad (\Sigma_2)$$

where Φ_{Succ} is the second-order formalization of the claim that the universe is the size of a successor cardinal, Φ_{Lim} is the second-order formalization of the claim that the universe is the size of a limit cardinal, and $F \cong G$ is as before.[14] Σ_2 is equivalent to (BLV) on successor cardinals and provides singleton abstraction (plus a single dummy object) on limit cardinals. Hence Σ_2 is Heck conservative but not weakly irenic (since not weakly stable). □

Lemma 5 *No entry on the second lowest row implies any entry on the next row up.*

Proof It suffices to show that Heck irenicity does not imply strong stability. Consider

$$@F = @G \leftrightarrow ((\Phi_{<\aleph_0} \wedge F \cong G) \vee (\Phi_{\aleph_0} \wedge X \equiv Y) \vee (\Phi_{\geq \aleph_1} \wedge F \cong G)), \quad (\Sigma_3)$$

with all abbreviations as before. Σ_3 is equivalent to (BLV) on countably infinite domains and provides singleton abstraction (plus a single dummy object) on finite and uncountably infinite domains. Hence Σ_3 is Heck irenic but not strongly stable. □

Lemma 6 *Heck stability does not imply conservative stability.*

Proof Consider

$$@F = @G \leftrightarrow \left((\Phi_{<\aleph_0} \wedge F \equiv G) \vee (\Phi_{\aleph_0} \wedge X \cong Y) \vee \Phi_{\geq \aleph_1} \right), \qquad (\Sigma_4)$$

where $F \cong G$ is, as before, the second-order formalization of the claim that either both F and G fail to be singleton concepts, or F and G are coextensive singleton concepts. Σ_4 is equivalent to (BLV) on finite domains, provides singleton abstraction (plus a single dummy object) on countably infinite domains, and provides a single object on uncountably infinite domains. Hence Σ_4 is Heck stable but not conservatively stable. □

This completes our proof of the correctness of the diagram from the end of the previous Section 4.

Finally, for the sake of completeness, we sketch a proof of the claim from note 13 that Heck conservativeness (irenicity) can be given an alternative but equivalent definition as *weak* conservativeness (irenicity) and critical fullness. The proof of the parenthetical claim is straightforward given the main claim and Lemma 1. For the main claim, it suffices to prove that, in the presence of critical fullness, weak conservativeness implies strong. So assume that Σ is weakly conservative and critically full. Consider an $\mathcal{L}$-model $\mathcal{M}$, and let κ be its cardinality. Since Σ is weakly conservative if and only if it is unbounded, Σ has a model of cardinality $\geq \kappa$. Let $\mathcal{N}$ be a model of this sort which is of minimal cardinality $\lambda \geq \kappa$. Since λ is thus a critical point, $\mathcal{N}$ has λ abstracts. We now claim that if Σ has a model of cardinality λ with λ many abstracts, then Σ also has a model with precisely κ many nonabstracts for any $\kappa \leq \lambda$. Given this claim, we are done, as the κ nonabstracts can be chosen to be a copy of $\mathcal{M}$. To prove the claim, we observe that the equivalence relation associated with our purely logical abstraction Σ does not "know" anything about whether or not an object is an abstract. Hence, the result of composing the abstraction mapping of any model of Σ with a one-to-one function yields another model of Σ. The claim now follows by observing that the λ abstracts of $\mathcal{N}$ can be correlated one-to-one with some chosen all-but-κ many members of $\mathcal{N}$.

6 Monotonicity

In addition to fullness, we claim there is a second idea underlying Heck's proposal of conservative stability as a criterion of acceptable abstraction. To bring this second idea into focus, let us consider why conservative stability is so much stronger than the other criteria we have considered, such as weak, strong, and Heck conservativeness (and weak, strong, and Heck irenicity). The reasons are relatively simple to locate.

Weak conservativeness/unboundedness, weak irenicity/stability, and strong stability are all conditions that are formulated in terms of (or equivalent to conditions formulated in terms of) the nature of the class of cardinals at which an abstraction is satisfiable. An abstraction principle is κ-satisfiable if and only if the equivalence relation does not "carve" up the domain into "too many" equivalence classes—that is, into more than κ-many equivalence classes. Thus, requiring that an abstraction be weakly conservative, or weakly irenic, or strongly stable, amounts to requiring that the principle in question not generate "too many" abstracts. As a result, each of these classes is closed under cardinality equivalence (cf. Definition 1).

Strong conservativeness, strong irenicity, Heck conservativeness, Heck irenicity, and strong Heck irenicity all require more than this. These conditions are all motivated by the thought that it is not enough that an abstraction principle not provide too many objects (i.e., is satisfiable on "enough" cardinals). The thought underlying these categories is that, in addition, acceptable principles ought to provide enough objects to "fill" the domain. In short, strong conservativeness requires that we can always extend a model of a theory T to a "full" model of Σ, while Heck conservativeness requires that we can always do so without moving past the "next" cardinal at which Σ is satisfiable.

Conservative stability, however, adds another consideration into the mix. In addition to requiring that an abstraction Σ provide κ-many abstracts any time the addition

of Σ to a theory will require that we move from a smaller domain to a domain of size κ; conservative stability also requires that Σ will continue to provide that many abstracts as we move to even larger domains. We can arrive at something like Heck's notion of conservative stability by combining strong Heck irenicity with something like the following monotonicity requirement.[15]

Definition 6   An abstraction Σ is *cardinality monotonic* if and only if, for any κ and γ such that $\kappa \leq \gamma$ and Σ is both κ- and γ-satisfiable, the cardinality of the class of abstracts of the sort characterized by Σ on a domain of size κ is no greater than the cardinality of the class of abstracts of the sort characterized by Σ on a domain of size γ.

In short, an abstraction is cardinality monotonic if and only if moving to a larger domain never decreases the number of abstracts provided by that principle. Heck suggests that the requirement that acceptable abstractions be cardinality monotonic might be well motivated:

> How could the fact that there are more objects of *other* sorts affect how many abstracts there are? [3, p. 233, footnote 10]

And he may well be right—there may indeed be good reasons for requiring that this additional condition hold of abstraction principles. As we saw in Section 2, however, this is still not enough to rule out principles which are no less (but also no more[16]) fishy than the ones Heck called to our attention.

A stronger notion of monotonicity naturally suggests itself, however. Let $\Sigma[\mathcal{M}]$ be the set of abstracts generated when the abstraction Σ is applied to a model $\mathcal{M}$. Thus, cardinality monotonicity requires that for all models $\mathcal{M} \subseteq \mathcal{N}$, the cardinality of $\Sigma[\mathcal{M}]$ not exceed that of $\Sigma[\mathcal{N}]$. But we can also impose the stronger requirement that $\Sigma[\mathcal{M}] \subseteq \Sigma[\mathcal{N}]$. Here is a way to implement this idea.

Definition 7   An equivalence relation $\Phi(X, Y)$ is *intrinsic* if and only if:

$$\Phi(X, Y) \quad \text{iff} \quad \Phi^{X \cup Y}(X, Y),$$

where $\Phi^{X \cup Y}$ is the result of restricting all the quantifiers of Φ to objects that fall under either X or Y. An abstraction is *monotonic* if and only if the associated equivalence relation is intrinsic.

It is easily seen that this monotonicity requirement excludes all the fishy ones that we have considered throughout this article. The requirement also excludes the abstraction known as New V, which is like Basic Law V except that all concepts equinumerous with the universe are mapped to a single dummy abstract. Some may find this exclusion undesirable. A great variety of mathematically and philosophically interesting abstractions are not excluded, however, for instance: (HP), (BLV), and all of the abstractions used in the "construction" of ordered pairs, integers, rational numbers, as well as real numbers via either Dedekind cuts or Cauchy sequences.

It is straightforward to verify that monotonicity entails cardinality monotonicity. But the requirement of monotonicity does little or nothing to restrict the cardinalities at which an abstraction can be satisfiable.[17] We do not know whether the requirement has any significance for the question of fullness.

We leave for another occasion the question of whether the bad company problem can be solved by means of the requirement of monotonicity combined with some stability requirement and/or some requirement of fullness.[18]

Notes

1. See the postscript to "On the Consistency of Second-Order Abstraction Principles," reprinted in [3, pp. 230–36].

2. This is one of a number of proposals tentatively suggested by Heck on p. 233. We shall return to other variants below.

3. The *ramsification* of an abstraction principle Σ is the result of replacing the abstraction operator with a free variable of appropriate type and then binding that variable with an initial existential quantifier. Notice that the ramsification of a second-order abstraction principle uses third-order logic. However, this ascent to third order can be avoided, if desired, by tweaking the ramsification method as in the proof of Lemma 3.6 of [5].

4. It is easy to tweak the definition of $\cong$ to address some possible concerns. If singleton abstraction is felt to be problematic, F and G could be permitted to be coextensive finite concepts. And if desired, one could stay closer to Σ by conjoining to the disjunct $\mathcal{R}(\Sigma) \wedge F \cong G$ the further requirement $F \sim G$. The resulting equivalence would relate each singleton concept to itself and no other concept, and behave exactly like $\sim$ otherwise.

5. This is the view of ØL, who favors a very different approach to the bad company problem, based on iterated application of individually benign steps of predicative abstraction. See [6] for an application of this approach to (a plural version of) Frege's Basic Law V; see also [4] and [8]. This approach retains the intuitive idea that abstracts have to be "generated" from below but develops it in a different way, akin to the way in which sets are "generated" from their members on the influential iterative conception of set.

6. This is the view of RC. If defensible, this view will enable him to defend his proposed solution to the bad company problem in [1] against the challenge posed by Heck's discovery.

7. This definition differs slightly from Weir's, which instead of clause (i) translates atomic predications homophonically. Our slight modification is defensible, for at least two reasons. First, it is consonant with the idea we are trying to capture, namely that $\Phi^{\Psi(x)}$ should say that Φ holds when we restrict our attention to all and only objects satisfying $\Psi(x)$. For instance, for "t_1 is heavier than t_2" to hold when we restrict our attention to physical objects, t_1 and t_2 have got to *be* physical objects. Second, Weir's definition and ours are equivalent if we assume that the formula (or theory) being relativized contains a conjunct (or axiom) of the form $\exists x (x = t)$ for each singular constant t that occurs in this formula (or theory). Since the logic we are considering is not free, this assumption is implicit in the formula (or theory) and thus unproblematic.

8. Weir's name for strong conservativeness is "Field-conservativeness" because of its similarity to a notion of conservativeness advocated by Hartry Field. His name for weak conservativeness is "Caesar-neutral conservativeness" because it is compatible with any answer to Frege's infamous Julius Caesar-problem.

9. That is, $\bar{\mathcal{N}}$ is just like $\mathcal{N}$ except that it "forgets" about the interpretation of all the constants in $\mathcal{L}_@^M$ that are not found in $\mathcal{L}_@$.

10. We can be more specific. In the case of [5], the proofs of Theorems 2.6 and 2.8 assume that the label "conservative" is attached to strongly conservative abstractions, although it is in fact attached to weakly conservative ones. These two "theorems" are therefore false when understood in accordance with the official definition, as are several of the later results which depend on these two "theorems"—although, as stated, everything is fine once the definitions are corrected. Similarly, [1] introduces Field-conservativeness (equivalent to strong conservativeness) in Section 2, but the taxonomy of abstraction principles developed immediately afterward mobilizes weak conservativeness (and is in fact incorrect as written in terms of strong/Field-conservativeness, since not all unbounded abstractions are strong/Field conservative—see Corollary 2 above). All results go through if the definition of weak conservativeness is uniformly substituted for the one given.

11. In fact, one additional assumption is needed for these results: that we can extend the language with the addition of κ-many constants, for any arbitrarily large κ. Since these conservativeness criteria are formulated to show that abstraction principles that satisfy them are "acceptable" in so far as they will not interfere with the truths in *any* theory, in *any* language, this assumption seems not only unobjectionable, but in fact philosophically mandated.

12. Notice that any stabilization point is a critical point, but not vice versa.

13. Heck suggests something very close to what we have called Heck irenicity as a weaker alternative to conservative stability in footnote 9 of page 233 of the postscript in question. In fact, as we shall prove in the next section, Heck conservativeness (irenicity) could have been given an alternative but equivalent definition as *weak* conservativeness (irenicity) and critical fullness.

14. See, for instance, [7, pp. 104–5], for the mentioned second-order formalizations.

15. It is worth noting that monotonicity is a stronger condition than conservative stability. Let $\approx_{\text{fin}}$ behave like HP on finite concepts but equate all infinite concepts, and let $\cong$ be as above. Consider:

$$@X = @Y \leftrightarrow (\Phi_{\text{Lim}} \wedge X \approx_{\text{fin}} Y) \vee (\Phi_{\text{Succ}} \wedge X \cong Y).$$

This abstraction principle has models of size κ for any infinite κ, and every model contains at least $\aleph_0$-many abstracts. Thus it is conservatively stable. But it is not monotonic: Any domain of size κ where κ is a limit cardinal contains $\aleph_0$-many abstracts, but every domain of size κ where κ is a successor cardinal contains κ-many abstracts.

16. Recall the two different responses canvassed at the end of Section 2.

17. Consider for instance abstraction on dyadic relations that are well-ordered, although all relations of certain specified "exceptional" cardinalities are mapped to a single dummy object.

18. Recall, though, the two approaches to the problem that were outlined at the end of Section 2.

References

[1] Cook, R. T., "Conservativeness, stability, and abstraction," *British Journal for the Philosophy of Science*, vol. 63 (2012), pp. 673–96. Zbl 1319.03013. MR 2965803. DOI 10.1093/bjps/axr047. 64, 66, 67, 72, 73

[2] Fine, K., *The Limits of Abstraction*, Oxford University Press, Oxford, 2002. Zbl 1043.00005. MR 2054384. 64

[3] Heck, R., *Frege's Theorem*, Oxford University Press, Oxford, 2011. Zbl 1245.00015. MR 3154918. 62, 67, 71, 72

[4] Linnebo, Ø., "Bad company tamed," *Synthese*, vol. 170 (2009), pp. 371–91. Zbl 1180.03007. MR 2540119. DOI 10.1007/s11229-007-9265-7. 72

[5] Linnebo, Ø., "Some criteria for acceptable abstraction," *Notre Dame Journal of Formal Logic*, vol. 52 (2011), pp. 331–38. Zbl 1252.03015. MR 2822493. DOI 10.1215/00294527-1435492. 66, 68, 72, 73

[6] Linnebo, Ø., "The potential hierarchy of sets," *Review of Symbolic Logic*, vol. 6 (2013), pp. 205–28. Zbl 1283.03085. MR 3069837. DOI 10.1017/S1755020313000014. 72

[7] Shapiro, S., *Foundations without Foundationalism: A Case for Second-Order Logic*, Oxford University Press, Oxford, 1991. MR 1143781. 73

[8] Studd, J. P., "Abstraction reconceived," *British Journal for the Philosophy of Science*, vol. 67 (2016), pp. 579–615. MR 3509984. DOI 10.1093/bjps/axu035. 72

[9] Weir, A., "Neo-Fregeanism: An embarrassment of riches," *Notre Dame Journal of Formal Logic*, vol. 44 (2003), pp. 13–48. Zbl 1071.03006. MR 2060054. DOI 10.1305/ndjfl/1082637613. 64, 66

Cook
Department of Philosophy
University of Minnesota
Minneapolis, Minnesota 55455
cookx432@umn.edu

Linnebo
Department of Philosophy
IFIKK
Oslo, Norway
linnebo@gmail.com