Notre Dame Journal of Formal Logic
Volume 52, Number 3, 2011

Some Criteria for Acceptable Abstraction

Øystein Linnebo

Abstract Which abstraction principles are acceptable? A variety of criteria have been proposed, in particular irenicity, stability, conservativeness, and unboundedness. This note charts their logical relations. This answers some open questions and corrects some old answers.

1 Introduction

An *abstraction principle* is a principle of the form

$$\S\alpha = \S\beta \leftrightarrow \alpha \sim \beta \qquad (\Sigma)$$

where the variables α and β range over entities of some sort, and where $\sim$ is an equivalence relation on this sort of entity. Frege's inconsistent Basic Law V shows that not every such principle is acceptable. A variety of criteria for acceptable abstraction have been proposed.[1]

This note charts the logical relations between some of the proposed criteria. I answer some technical questions thrown open by the discovery of errors in the proofs and claims of the most systematic study of these issues to date, namely, the deservedly influential [12]. My results are summarized by the following strict implications.

$$\boxed{\begin{array}{ccc} \text{irenic} & \Rightarrow & \text{conservative} \\ \Downarrow & & \Downarrow \\ \text{stable} & \Rightarrow & \text{unbounded} \end{array}}$$

Restricted to the important class of "purely logical" abstraction principles (Definition 2.7), the vertical dimension of the previous diagram is collapsed to yield two equivalences and one strict implication.

$$\boxed{\text{irenic} \Leftrightarrow \text{stable} \Rightarrow \text{conservative} \Leftrightarrow \text{unbounded}}$$

Received June 23, 2010; accepted February 3, 2011; printed July 25, 2011
2010 Mathematics Subject Classification: Primary, 00A30, 03A05, 03C65
Keywords: abstraction, bad company problem, conservativeness, unboundedness, stability, irenicity
 10.1215/00294527-1435492

Throughout the note, I proceed model-theoretically and work against the background of standard ZFC set theory.[2] No attempt is made to assess the philosophical plausibility of the suggested criteria.[3]

2 Conservativeness and Unboundedness

The criterion of *conservativeness* is based on a fairly intuitive idea, namely, that an abstraction principle Σ is acceptable just in case it can be added to any theory without disturbing this theory's claims about the objects with which it is concerned. That is, adding Σ to a theory T may give us additional information about 'new' objects with which Σ is concerned, but not about the 'old' objects with which T is concerned.

The standard way of making this intuitive idea precise is as follows. Let T be a theory in some base language $\mathcal{L}$ that does not contain the abstraction operator §. Let $\mathcal{L}^+$ be the language that results from adding to $\mathcal{L}$ the operator §. Define the predicate 'old(x)' as $\neg\exists\alpha(x = \S\alpha)$. Let φ^{old} be the result of restricting all the quantifiers in φ to 'old' objects, and let T^{old} be the result of replacing every axiom φ of T with φ^{old}.

Definition 2.1 (Conservativeness) An abstraction principle Σ is *conservative over an $\mathcal{L}$-theory T* if and only if for any $\mathcal{L}$-formula φ we have

$$\text{if } T^{\text{old}} \cup \{\Sigma\} \models \varphi^{\text{old}}, \text{ then } T \models \varphi.$$

Σ is *conservative* if and only if it is conservative over any $\mathcal{L}$-theory T provided $\mathcal{L}$ does not contain the operator §.[4]

Note that we are concerned with *model-theoretic* rather than *proof-theoretic* conservativeness. Note also that the only restriction on the base theory T is that its language must not contain the operator §. (Henceforth, this restriction will be left implicit.) It would be inappropriate to impose any other restrictions on the base theory. For the intuitive idea which motivates the criterion of conservativeness is that, whatever theory T scientists might come to formulate, adding an abstraction principle should not disturb the claims about the objects with which T is concerned. To impose any further restrictions would be to go beyond abstractionism's remit and prejudge what theories it might be legitimate for scientists to use.

Another criterion that has been discussed is *unboundedness*, which we define as follows.

Definition 2.2 (Unboundedness) An abstraction principle Σ is *κ-satisfiable* if and only if Σ is satisfiable in a domain of cardinality κ. Σ is *unbounded* if and only if Σ is κ-satisfiable for an unbounded sequence of cardinals κ.

Theorem 4.2 of [12] asserts that any abstraction principle that is "unbounded" is conservative. However, the following example shows Weir's assertion to be incorrect, given his (and our) official definitions.[5]

Example 2.3 Let Σ be the abstraction principle,

$$\S F = \S G \leftrightarrow \neg\exists x\, Px \vee \forall u(Fu \leftrightarrow Gu),$$

where F and G are monadic second-order variables and P is an atomic predicate of the base language $\mathcal{L}$. Σ is unbounded because it is satisfiable in any domain by interpreting P as not applying to any object in the domain. However, Σ is nonconservative over the theory T whose sole axiom is $\exists x\, Px$. For on the one hand we have $T^{\text{old}} \cup \{\Sigma\} \models \bot$ (because T^{old} ensures that the domain contains a P, which turns Σ into Frege's inconsistent Basic Law V). But on the other hand we have $T \not\models \bot$.

What went wrong? The problem is that the notion of conservativeness requires that the vocabulary of the base language $\mathcal{L}$ retain its meaning on the 'old' domain, whereas the notion of unboundedness is defined in terms of satisfiability and thus allows this vocabulary to be reinterpreted. Any notion of unboundedness capable of implying conservativeness must ensure that the vocabulary of the base language retains its meaning on the 'old' domain. This suggests to the following definition.

Definition 2.4 (Uniform unboundedness) An abstraction principle Σ is *uniformly unbounded* if and only if for any model M of any base language $\mathcal{L}$ there is a model N of the extended language $\mathcal{L}^+$ such that

(i) N is an extension of M whose 'old' objects are precisely the objects of M,
(ii) N satisfies Σ.

Lemma 2.5 *Any uniformly unbounded abstraction principle is unbounded. But the converse does not hold.*

Proof The first claim is straightforward. The second claim follows from Example 2.3. □

Theorem 2.6 *An abstraction principle is conservative if and only if it is uniformly unbounded.*

Proof Assume Σ is uniformly unbounded. Assume that $T \not\models \varphi$ for some $\mathcal{L}$-formula φ. Then there is a model M such that $M \models T \cup \{\neg\varphi\}$. By the first assumption, M can be extended to a model N of Σ whose 'old' objects are precisely the objects of M. It follows that $N \models T^{\text{old}} \cup \{\Sigma, \neg\varphi^{\text{old}}\}$. This shows $T^{\text{old}} \cup \{\Sigma\} \not\models \varphi^{\text{old}}$. Since $\mathcal{L}$, T, and φ were arbitrary, it follows that Σ is conservative.

Assume next that Σ is conservative. Let M be a model of some base language $\mathcal{L}$. Let $\mathcal{L}_M$ be the enriched base language that adds to $\mathcal{L}$ a distinct constant for every element of M. ($\mathcal{L}_M$ may thus be an uncountable language, as is commonplace in model theory.) Let the $\mathcal{L}_M$-theory T consist of the diagram of M (that is, the set of all atomic sentences and negated atomic sentences in $\mathcal{L}_M$ which are true in M). Since T contains no quantifiers, we have $T^{\text{old}} = T$. Assume $T \cup \{\Sigma\}$ has no model. Then $T^{\text{old}} \cup \{\Sigma\} \models \bot$, whence by Σ's conservativeness, $T \models \bot$, which contradicts $M \models T$. So let N be a model of $T \cup \{\Sigma\}$. Then N can be assumed to be an extension of M whose 'old' objects are precisely the objects of M. Viewed as an $\mathcal{L}$- (rather than $\mathcal{L}_M$-) model, N then shows Σ to be uniformly unbounded.[6] □

Although generalized languages such as $\mathcal{L}_M$ are mathematically acceptable, it may be objected that the present use of such languages is philosophically problematic because the criterion of conservativeness was formulated with ordinary languages in mind, not generalized languages. But the objection is unconvincing. The conservativeness of an abstraction principle Σ is supposed to ensure that Σ can be added to any base theory T whatsoever without disturbing this theory's claims about the objects with which it is concerned. So it would be inappropriate to impose any restrictions on the base theory T (other than that its language not contain the operator §). Indeed, the question of whether there could be beings capable of grasping an infinitary base theory of the sort invoked above is not for pure mathematics to answer. So a philosophical account of pure mathematics must not be made to depend on a negative answer to this question. Moreover, when we chose to study model-theoretic rather than proof-theoretic conservativeness of higher-order theories, we already gave up on any requirement of a close link with actual (or even mildly idealized) human capacities. We are in general no more able to assess questions of higher-order semantic consequence than we are to master a generalized language.[7] However, it is an interesting technical question—to which I don't know the answer—whether Theorem 2.6 can be proved without going beyond an ordinary second-order language.

I next define another assumption and show that its addition enables us to prove a partial converse of the result of Lemma 2.5.

Definition 2.7 (Purely logical) An abstraction principle Σ is *purely logical* if and only if it contains no nonlogical vocabulary except the operator §.

Theorem 2.8 *Any unbounded and purely logical abstraction principle is uniformly unbounded.*

Proof Let Σ be an unbounded and purely logical abstraction principle, and let M be a model of some base language $\mathcal{L}$. Since Σ is unbounded, it is satisfiable in a model N of cardinality larger than that of M. Since Σ is purely logical, it has no nonlogical vocabulary in common with $\mathcal{L}$. (Recall that $\mathcal{L}$ has been assumed not to contain the operator §.) This ensures that N can be taken to be an extension of M whose 'old' objects are precisely those of M. □

Corollary 2.9 *Let Σ be a purely logical abstraction principle. Then Σ is conservative if and only if it is unbounded if and only if it is uniformly unbounded.*

Proof Immediate from Lemma 2.5 and Theorems 2.6 and 2.8. □

It should be noted that the restriction to purely logical abstraction principles is far from trivial. Frege's famous direction abstraction principle—which says that the directions of two lines l_1 and l_2 are identical if and only if l_1 and l_2 are parallel—is nonlogical, as are some of the abstraction principles relied on in the standard abstractionist approaches to the real numbers [4] and [10].[8]

3 Irenicity and Stability

Unfortunately, a conservative abstraction principle need not be acceptable, as a theorem of Weir's shows ([12], Theorem 4.3).

Theorem 3.1 (Weir) *There are pairs of purely logical abstraction principles each of which is conservative but which are not jointly satisfiable.*

A key ingredient of the proof of Theorem 3.1 is the following, well-known theorem.

Theorem 3.2 (Folklore) *In the language of pure second-order logic we can characterize various cardinality properties of a concept X, such as being of size $\aleph_n$ for some natural number n, being of continuum size, being of limit-cardinal size, being of successor-cardinal size, and being of inaccessible size.*

Proof See, for instance, [9], pp. 104–5. □

Proof of Theorem 3.1 Consider the following restricted version of Frege's Basic Law V:

$$\epsilon F = \epsilon G \leftrightarrow (\textsc{Bad}(F) \wedge \textsc{Bad}(G)) \vee \forall x(Fx \leftrightarrow Gx) \tag{RV}$$

where $\textsc{Bad}(F)$ is some $\mathcal{L}$-formula. By Theorem 3.2 we can let $\textsc{Bad}_1(F)$ and $\textsc{Bad}_2(F)$ express that the universal concept—that is, the concept U such that

$\forall x\, Ux$—is, respectively, of successor-cardinal size and limit-cardinal size. The resulting versions of (RV) are easily seen to be satisfiable in all and only domains of, respectively, successor-cardinal size and limit-cardinal size. So the two principles are not jointly satisfiable. But each is unbounded and thus conservative by Corollary 2.9. □

The next definition is a natural response to the problem posed by Theorem 3.1.

Definition 3.3 (Stability) An abstraction Σ is *stable* if and only if there is a cardinal κ such that Σ is λ-satisfiable for all cardinals $\lambda \geq \kappa$. Σ is *strongly stable* if and only if there is a cardinal κ such that Σ is λ-satisfiable just in case $\lambda \geq \kappa$.

Lemma 3.4 *Strong stability implies stability, which implies unboundedness. But neither converse holds.*

Proof The implications are trivial. To establish that stability doesn't imply strong stability, consider the version of (RV) where the condition $\text{B}\textsc{ad}(F)$ is defined so as to be true in domains of cardinality $\aleph_0$ and $\geq \aleph_\omega$ but not in a domain of any other cardinality. To establish that unboundedness doesn't imply stability, we define the condition $\text{B}\textsc{ad}(F)$ so as to be true just in domains whose size is a limit-cardinal. Both conditions can be expressed by Theorem 3.2. □

Theorem 6.1 of [12] asserts that stability is equivalent to another criterion that has been proposed, namely, irenicity.

Definition 3.5 (Irenicity) An abstraction Σ is *irenic* if and only if it is conservative and jointly satisfiable with any other conservative abstraction principle.

However, Weir's proof is flawed for two independent reasons. A minor flaw is brought out by Example 2.3, which shows that an abstraction principle can be strongly stable without being conservative and thus a fortiori without being irenic. Corollary 2.9 suggests that this problem can be avoided by adding the assumption that Σ is purely logical, as will be confirmed below.[9] But even with this added assumption, an observation due to Shapiro shows the proof to contain a second and more important flaw, namely, a failure to distinguish properly between stability and strong stability [8]. All Weir's proof establishes is that (assuming pure logicality) strong stability entails irenicity, which in turn entails stability. The proof thus leaves open the question whether any of the converses hold. This will now be investigated.

Lemma 3.6 *Let Σ be an abstraction principle that is not stable. Then there is a conservative and purely logical abstraction principle Γ that is not jointly satisfiable with Σ.*

Proof By Corollary 2.9 it suffices to find a purely logical abstraction principle Γ that is satisfiable at precisely those cardinalities where Σ is not satisfiable. I claim that it is possible to formulate a condition $\text{B}\textsc{ad}(F)$ which expresses that Σ isn't satisfiable on the universal concept U. (The variable 'F' thus occurs vacuously in the condition.) Assuming this claim, we can let Γ be the resulting version of (RV). To see this, assume first that Σ isn't κ-satisfiable. Then any concept F on a domain of size κ will be $\text{B}\textsc{ad}$, which makes Γ trivially κ-satisfiable. Next, assume that Σ is κ-satisfiable. Then no concept F on a domain of size κ will be $\text{B}\textsc{ad}$, which means that on such domains, Γ is like Basic Law V and thus not κ-satisfiable.

It remains to prove the claim. A simple way to do so is to let $\mathcal{R}(\Sigma)$ be the Ramseyfication of Σ, which is available in a language of order one higher than that of Σ. Then $\textsc{Bad}(F)$ can be chosen to be $\neg\mathcal{R}(\Sigma)$. However, if desired, the claim can also be proved without having to ascend to a language of order higher than Σ. Consider the case where Σ is second-order; the other cases are analogous. Then a dyadic relation R can be used to code an assignment of objects to selected monadic concepts by letting $\forall u(Fu \leftrightarrow Rux)$ mean that x is assigned to the concept F. Let $\sim$ be the equivalence relation on which Σ abstracts. Say that F *is associated with* x *under* R if and only if F bears $\sim$ to some concept F' that R associates with x. The claim that Σ is satisfiable can then be expressed as the claim that there is a dyadic relation R such that every concept F is associated with an object x under R, and such that two concepts F and G are associated with the same object under R just in case $F \sim G$. □

Theorem 3.7 *Any irenic abstraction principle is stable. But the converse does not hold.*

Proof Assume Σ is not stable. Then by Lemma 3.6 there is a conservative abstraction principle Γ with which Σ is not jointly satisfiable, which shows that Σ is not irenic. Example 2.3 shows that the converse does not hold. □

However, as in the case of Theorem 2.8, a converse holds under the added assumption of pure logicality.

Theorem 3.8 *Let Σ be a purely logical abstraction principle. Then Σ is stable if and only if it is irenic.*

Proof By the previous theorem, it suffices to prove that every stable and purely logical abstraction principle is irenic. So assume Σ is stable and purely logical. Since stability implies unboundedness, Σ is conservative by Corollary 2.9. Let Γ be another conservative abstraction principle. We need to show that Σ and Γ are jointly satisfiable. Let κ be a cardinal such that Σ is satisfiable in any domain of cardinality $\geq \kappa$. But by Theorem 2.6 and Lemma 2.5, Γ is unbounded, which ensures that there is a cardinal $\lambda \geq \kappa$ such that Γ too is satisfiable in domains of cardinality λ. It follows that Σ is irenic. □

Corollary 3.9 *Conservativeness does not imply irenicity or stability.*

Proof By Theorem 3.7, it suffices to show that conservativeness does not imply stability. Assume it did. Then by Theorem 3.8, any unbounded abstraction that is purely logical would be stable. But we know this not to be so from the example in the proof of Lemma 3.4. □

Corollary 3.10 *There are purely logical and irenic abstraction principles that are not strongly stable.*

Proof This is immediate from Theorem 3.8 and the observation that there are purely logical abstraction principles that are stable but not strongly stable. The observation is established in the proof of Lemma 3.4. □

4 Conclusion

An inspection of the results established above shows that the logical relations between the proposed criteria are indeed as depicted in the two boxed diagrams at the beginning of this note.

Notes

1. See, for instance, [1], [13], and [12], and, for an overview, [6].

2. This approach relies on what Stewart Shapiro in [10] calls *the external perspective*, namely, the perspective of a bystander who is interested in the prospects of the abstractionist projected against the background of standard mathematics.

3. See, however, [11] and [7], which challenge the status of stability and strong stability as, respectively, necessary and sufficient conditions for acceptability. See [3] for a defense of (strong) stability.

4. [12] discusses a slightly different definition as well. Say that Σ is *CN-conservative* (for "Caesar-neutral") if and only if the same definition holds except with 'old' replaced by a new predicate O about which nothing is assumed. Unpacking definitions, one easily sees that conservativeness implies CN-conservativeness. Note 6 shows that the two notions are in fact equivalent.

5. Weir's definitions are the same as mine: see pp. 21 and 24. However, Weir (p.c.) informs me he was implicitly restricting himself to what I call "purely logical" abstraction principles (see my Definition 2.7), for which his assertion is correct (see my Theorem 2.8). Here I examine what happens without this substantial restriction and explicitly state the restriction when it is made.

6. The argument just given also shows that CN-conservativeness (as defined in Note 4) implies uniform unboundedness. Since the latter notion implies conservativeness, which (as observed in Note 4) implies CN-conservativeness, all three notions are in fact equivalent. This justifies my choice henceforth to focus exclusively on ordinary conservativeness.

7. Another option is to prove the left-to-right direction of Theorem 2.6 using the generalized quantifier "there are κ or more x such that ...".

8. However, say that an abstraction principle is *essentially logical* if and only if it is either purely logical or its only nonlogical expressions are abstraction operators governed by other abstraction principles that are essentially logical. (Thanks here to Stewart Shapiro, who attributes the idea to Crispin Wright.) The abstraction principles used in the mentioned constructions of the reals have this property. Unfortunately, the analogue of Theorem 2.8 that results from replacing the assumption of pure logicality with essential logicality is false, as can be seen by replacing '$\neg\exists x\, Px$' in Example 2.3 with the claim that any two #-abstracts are identical, where '#' is governed by Hume's Principle.

9. Recall from Note 5 that Weir claims he was implicitly restricting himself to purely logical abstraction principles.

References

[1] Boolos, G., "The standard of equality of numbers," pp. 234–54 in *Meaning and Method: Essays in Honor of Hilary Putnam*, edited by G. Boolos, Harvard University Press, Cambridge, 1995. Reprinted in [2]. MR 1376398. 337

[2] Boolos, G., *Logic, Logic, and Logic*, edited by R. Jeffrey, Harvard University Press, Cambridge, 1998. Zbl 0955.03008. MR 1675856. 337

[3] Cook, R., "Conservativeness, stability, and abstraction," forthcoming in the *British Journal for Philosophy of Science*. 337

[4] Hale, B., "Reals by abstraction. Abstraction and neo-logicism," *Philosophia Mathematica. Series III*, vol. 8 (2000), pp. 100–123. Zbl 0968.03010. MR 1768940. 334

[5] Hale, B., and C. Wright, *The Reason's Proper Study. Essays towards a Neo-Fregean Philosophy of Mathematics*, The Clarendon Press, New York, 2001. Zbl 1005.03006. MR 2037746. 338

[6] Linnebo, Ø., "Introduction [to a special issue on the bad company problem]," *Synthese*, vol. 170 (2009), pp. 321–29. Zbl 1182.03018. MR 2540116. 337

[7] Linnebo, Ø., and G. Uzquiano, "Which abstraction principles are acceptable? Some limitative results," *The British Journal for the Philosophy of Science*, vol. 60 (2009), pp. 239–52. Zbl 1174.03002. MR 2525783. 337

[8] Shapiro, S., "The company kept by cut abstraction (and its relatives)," *Philosophia Mathematica*, vol. 19 (2011), pp. 107–38. 335

[9] Shapiro, S., *Foundations without Foundationalism. A Case for Second-Order Logic*, vol. 17 of *Oxford Science Publications*, The Clarendon Press, New York, 1991. Zbl 0732.03002. MR 1143781. 334

[10] Shapiro, S., "Frege meets Dedekind: A neologicist treatment of real analysis," *Notre Dame Journal of Formal Logic*, vol. 41 (2000), pp. 335–64. Special Issue: Logicism and the Paradoxes: A Reappraisal, II. Zbl 1014.03013. MR 1963486. 334, 337

[11] Uzquiano, G., "Bad company generalized," *Synthese*, vol. 170 (2009), pp. 331–47. Zbl 1186.03015. MR 2540117. 337

[12] 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. 331, 332, 334, 335, 337

[13] Wright, C., "On the philosophical significance of Frege's theorem," pp. 201–44 in *Language, Thought, and Logic. Essays in Honour of Michael Dummett*, edited by R. G. Heck, Jr., The Clarendon Press, New York, 1997. Reprinted in [5]. Zbl 0938.03508. MR 1724691. 337

Acknowledgments

I am extremely grateful to Roy Cook for extensive comments and discussion, without which this note would not have existed. Thanks also to Stewart Shapiro and Alan Weir for valuable comments. Research for this note was supported by the AHRC-funded project *Foundations of Structuralism* and an ERC Starting Grant (241098). Their assistance is gratefully acknowledged.

Birkbeck
University of London
UNITED KINGDOM
o.linnebo@bbk.ac.uk