Book Review

Keith Simmons. *Universality and the Liar: An Essay on Truth and the Diagonal
   Argument*. Cambridge University Press, Cambridge, 1993. xii–229 pages.

The literature on truth and the liar paradox has enjoyed a recent resurgence, of which
McGee [6], Koons [4], and Gupta and Belnap [3] are only three, if prominent, examples. Simmons's book is a welcome addition to the field.

   The book's most notable contribution is the author's own "singularity solution"
to the liar paradox, which develops and makes precise Gödel's intuition in [2] that the
paradoxes correspond to "semantic singularities," analogous to division by zero, limit
points where the ordinary semantic rules no longer apply. Along the way, Simmons
manages to develop a general account of diagonal arguments, review and criticize all
the modern approaches to the paradoxes, and present an erudite account of a medieval
singularity solution. As is clear, this is no mean feat.

   The book is organized as follows. Chapter 1 contains a review of some basic
facts about the Liar and related paradoxes, as well as of some possible lines of attack. The chapter also introduces two themes that will be developed at length in the
book, namely the issues of "semantic universality" and diagonalization. The latter
is taken up in Chapter 2, which develops a general theory of diagonal arguments,
explaining what makes some of them *good* (i.e., giving rise to deep mathematical
results such as Cantor's theorem), and some of them *bad* (i.e., giving rise to paradoxes). Although structural similarities between diagonal arguments and paradoxes
have long been noticed, to this reviewer's knowledge this is the first general treatment of the subject. Chapters 3 and 4 review all modern solutions to the Liar paradox, beginning with Kripke's truth-gap approach in [5], and touching upon Gupta's
and Belnap's [3] revision theory of truth, McGee's [6] notion of definite truth, and
Feferman's [1] type-free theory of partial predicates. Chapter 5 presents a medieval
version of the singularity approach, which the author traces back to Ockham, Burley,
and the Pseudo-Sherwood. Chapters 6 and 7 introduce and formally develop the author's own singularity solution. Chapter 8 contains applications of the solution to a
variety of paradoxical phenomena. Chapter 9 closes the book with a long reflection
on semantic universality as a *desideratum* for any model of natural language and the
degree to which such a *desideratum* is met by the singularity approach.

Received October 31, 1995

As mentioned, universality and diagonalization are recurring themes in the book. The notion of universality here involved is the one introduced by Tarski in his *Wahrheitsbegriff* [9], or at least in certain popularizations of it. Roughly, a language is "(semantically) universal" if and only if it can express all (semantic) concepts. Although the notion of universality (not restricted to semantic notions) seems quite unreasonable, the author presents the notion of semantic universality as a desirable trait of any semantic theory, and perhaps a trait that is actually possessed by natural language. Tarski's theorem on the undefinability of truth then implies (among other consequences) the proposition that no formalized language meeting certain minimal criteria can be universal. These criteria are, roughly, that the language be classically two-valued, closed under negation, and capable of drawing fine enough syntactic distinctions. It has been shown that all three of these requirements are necessary for Tarski's theorem.

The theorem itself, as is well known, is an instance of a diagonal argument. Tarski reduces to absurdity the hypothesis of the existence of a predicate $T(x)$ satisfying all the equivalences $T(t) \leftrightarrow \varphi$, where t is a term denoting (the code of) φ. The proof shows how to construct a sentence τ provably equivalent to $\neg T(t)$, where t denotes (the code of) τ. So, τ effectively says of itself (modulo an appropriate coding) that it is not true.

We are skating on thin ice here. The proof of Tarski's theorem is remarkably similar to the reasoning involved in the Liar paradox. This brings to the foreground the question discussed in the second chapter of Simmons's book, viz., the question of the difference between *good* diagonal arguments, giving rise to deep and interesting theorems, and *bad* diagonal arguments, bringing us on the brink of paradox and inconsistency. To this reviewer's knowledge, the treatment given in the book is the most general to date. Simmons identifies all the main constituents of a diagonal argument and shows how paradoxes arise whenever one of the collections involved in the argument is unwarrantedly assumed to be well-defined. The analysis is crisp and to the point; the author shows how a number of theorems involving some form of *reductio* fall within the general framework he devised. As an added benefit, once the stage is set for a formalization of diagonal arguments, it is a simple matter to generalize the approach to higher-order arguments (involving more than two sets).

Next, Simmons takes up the modern solutions to the liar. His presentation is informative and thorough. The author starts with what is perhaps the most natural reaction when confronted with the Liar paradox: the denial that any sentence is either true or false. Among the theories that make this intuition precise, known as *truth-gap theories*, perhaps the best known is Kripke's [5] "fixed point" proposal. Kripke shows how, in the framework of Kleene's (weak or strong) 3-valued logic, one can define, relative to a particular "ground model," a monotone "jump" operator taking (anti-)extensions for truth into (anti-)extensions for truth. Any fixed point of the jump (and, in particular, the *least* such) provides an interpretation for the truth predicate. Simmons's main criticism of truth-gap theories is that no such theory can do justice to the universality of natural language: if we assume the language to be universal, then it is possible to "close the gap" and generate a new paradox of the form "this sentence is either not true or gappy." So, although the object language contains its own truth predicate, there is no way to express the "complement of T" within the language.

Therefore, not only is the language not universal, but we are forced up a metalinguistic hierarchy *á la* Tarski.

Along with Kripke's fixed point proposal, one of the most interesting approaches to the theory of truth is the so-called "revision theory" of Gupta and Belnap. In Chapter 4, Simmons generalizes his critique of the truth-gap theory to "any purportedly non-Tarskian approach to the Liar paradox" by asking two subordinate questions (p. 62): (*i*) "does the theory give rise to semantic concepts that can be expressed only in the metalanguage, on pain of paradox?" and if so, (*ii*) "are these semantic concepts available to the ordinary speaker, independently of philosophical reflection on the Liar?"

These questions, according to Simmons, are particularly pressing also in the case of the Herzberger-Gupta-Belnap approach. (For mysterious reasons Belnap's contributions to the Revision Theory of Truth, beginning with the "limit rule," are passed over in silence by Simmons, who refers to the theory as "Herzberger's and Gupta's Stable Truth.") The upshot of Simmons's otherwise crisp presentation of the theory is that the object language cannot express, on pain of paradox, the concept of *stable truth*. It follows that the language is expressively *incomplete*, a feature that is all the more problematic if, as the proponents of the Revision Theory have suggested, the notion of stable truth is one that is available to the philosophically untrained speaker. Similar criticisms apply to Feferman's [1] notion of truth and McGee's [6] notion of *definite truth*.

A different sort of proposal is embodied by the so-called "inconsistency views." These are the views, for instance, put forward by Rescher and Brandom [8] and Priest [7]. According to such views we have to "bite the bullet," as it were, and accept the fact that natural language is inconsistent, and that therefore any formal language capable of capturing the natural concept of truth will in turn be inconsistent or paradoxical. The challenge then is to allow such an inconsistency without trivializing the language (a task, as is clear, that requires giving up classical logic in one way or another). Rescher and Brandom accomplish this by resorting to modal logic, and Priest by resorting to para-consistent logic. In particular, according to Priest, the claim that the Liar is paradoxical should itself be viewed as paradoxical. Simmons points out, however, how the treatment by Rescher and Brandom also fails to be expressively incomplete, whereas Priest's treatment, although escaping a Tarskian hierarchy, pays too high a price in terms of the inconsistencies we are asked to accept.

In Chapter 5 Simmons reviews a class of medieval solutions to the Liar, which he presents as precursors to and inspirers of his own singularity solution. Although the historical and exegetical accuracy of Simmons's treatment lie outside this reviewer's field of expertise, the discussion there seems to be highly erudite and to the point.

The author's own singularity proposal is outlined in Chapter 6 and further developed in the following three chapters. As mentioned, one of Simmons's central claims is that hierarchical theories do not seem to present a natural account of our semantic concepts. On the contrary, Simmons singles out four central claims of the singularity proposal and argues that they provide for a far more natural theory.

The first claim is that semantic pathology may be analyzed in terms of the notion of *groundedness*. In this, the singularity proposal is in agreement with a number of the hierarchical accounts: first and foremost Kripke's fixed point approach (in which the

notion of groundedness plays a central rôle), but also the Revision-theoretic approach, with its notion of stable truth.

The second claim is that solutions to the Liar must be *context sensitive*: the objects of analysis are not sentence types but sentences in context. There is a sense in which the sentence:

$$(1) \qquad (1) \text{ is not true}$$

is paradoxical, whereas

$$(2) \qquad (1) \text{ is not true}$$

is not paradoxical and, in fact, true. Since (1) and (2) are tokens of the same sentence, the difference between them can be brought out only within the framework of a context sensitive approach.

The third claim is that the predicates 'true' and 'false' are themselves context-sensitive terms, shifting their extensions according to context. Contextual approaches to the Liar have been developed by Parsons, Burge, Barwise and Etchemendy, and Gaifman. Consider the sentence

$$(3) \qquad \text{The sentence written on the blackboard in room 101 is not true.}$$

Suppose that someone inscribed this sentence on the blackboard in what is in fact room 101, although he believed he was in room 102. Then this sentence, as we know, is paradoxical, albeit contextually so. Moreover, we can carry out a piece of strengthened-liar reasoning and conclude that since the sentence is paradoxical, it is neither true nor false, and hence in particular not true, which makes sentence (3) true after all. Simmons distinguishes three stages in this piece of reasoning. At the first stage we reason to the conclusion that (3) is pathological. At the second stage, *given this conclusion*, we infer

$$(4) \qquad (3) \text{ is not true.}$$

At the third stage, *given the conclusions reached at the previous two stages*, we conclude

$$(5) \qquad (3) \text{ is true.}$$

The extension of 'true' in (4) and (5) is different from its extension in (3), a fact that helps explain the pathological nature of (3) and the non-pathological nature of (4) and (5). Sentences (4) and (5) are called by Simmons *explicit reflections*, and they provide the beginning of what he calls the *reflective hierarchy*. (Of course, Simmons will claim that the reflective hierarchy is non-hierarchical, or at least not in the same sense as Tarski's, Kripke's, and Gupta's and Belnap's hierarchies: more on this below.)

The fourth claim is what really sets apart Simmons's proposal from other context-sensitive approaches: it's the claim that the pathological nature of, say, (3), ought to impose *minimal* restrictions of the extension of 'true'. This is the principle of *Minimality*: exceptions to the applicability of the semantical predicates 'true' and 'false' should be kept to a minimum. Essentially, we should allow only those exceptions as are necessary to avoid pathology. Accordingly, sentence (3) falls outside of the range of applicability of the occurrence of 'true' in (3), but not outside the range of applicability of occurrences of 'true' in (4) or (5). In turn, sentence (3) will fall in the

extension of the occurrence of 'true' in (5) and in the anti-extension of the occurrence of 'true' in (4).

The author also develops a formal account of how singularities are to be detected. First, each sentence is analyzed by means of a syntactic tree which makes explicit the connection not only between any truth-functional or quantificational compound and its components, but also between a truth (or falsity) predication and the sentence to which it applies. The tree is then *pruned* by terminating each branch at the first occurrence of a grounded sentence. If the resulting tree still contains infinite branches, then a singularity has been detected. When further reflecting on the results thus achieved, Minimality is invoked to exclude from the extension of an occurrence of 'true' all and *only* its singularities. Not least among the merits of the book is the fact that the formal account is developed to a high degree of rigor.

Let us now step back and consider what has been accomplished. The singularity approach is certainly very appealing, for reasons that were already clear to Gödel, in a passage quoted by Simmons (p. 116):

> It might even turn out that it is possible to assume every concept to be significant everywhere except for certain 'singular point' or 'limiting point', so that the paradoxes would appear as something analogous to dividing by zero. […] Such a system would be most satisfying in the following respect: our logical intuitions would then remain correct up to certain minor corrections, i.e. they could be then considered to give an essentially correct, only somewhat 'blurred', picture of the real state of affairs.

Simmons makes a good case for the merits of his approach, along Gödel's lines. There is, however, also an opposite point of view, which carries some intuitive punch, and which could be articulated as follows. There is an intrinsic danger in the singularity proposal, due to the fact that the precise identification of the singularities might appear to be *ad hoc*, and not motivated in a general way. Ideally, one would want a *general* and *principled* account of why a singularity is a singularity, other than the fact that the ordinary semantic rules fail there. In other words, with the singularity approach one runs the risk of throwing one's explanatory arms up in the air, giving up on a real explanation of the singularities. We know the Liar and other pathological sentences are singularities, in that they do not obey the usual semantic rules: but *why* are they so? Proponents of the singularity approach must be careful to answer this question, and although it is perhaps possible to distill an answer from Simmons's book, one still wishes that Simmons had at least better acknowledged its pressing character.

Let us compare Simmons's approach to a theory that shares some of its features, while at the same time embodying quite different intuitions. I am referring here to Russell's *Ramified Theory of Types*. In such a theory we have a nested hierarchy of propositional functions (not of sets or classes), in which each function having the truth values as converse domain is assigned a *type* (recursively determined by the type of its arguments), as well as an *order* defined in such a way as to exceed the order(s) of the propositional functions over which its variables range. In the Ramified Theory of Types, the truth predicate can be defined using propositional quantification by putting $Tx =_{df} \forall p (p = x \supset p)$. However, the derivation of the Liar paradox is blocked since there are no unrestricted quantifiers: rather, the quantifier $\forall p$ can range only over

propositions of a certain given order. One could indeed think of the order of the variable p as contextually specified, although this was by no means Russell's intent. We would then have a theory that has points of contact with Simmons's proposal, at least in that paradoxes are blocked by excluding pathological sentences from the range of applicability of the corresponding truth predicates. However, such an exclusion is now *general* and *principled* and not due merely to the fact that doing otherwise generates inconsistencies.

As mentioned, to some extent there are conflicting intuitions here. On the one hand the singularity approach has the advantage of keeping "exceptions" to the universal applicability of the truth predicates to a minimum. On the other hand, there is always the risk that such exceptions turn out to be *ad hoc* and not motivated in a general fashion. Obviously, this need not mean that Simmons's approach actually runs into such a risk, but he should be mindful of such a possibility.

The comparison to Russell's Theory of Ramified Types can help clarify matters in another respect. We have seen that Simmons's account gives rise to a *Reflective Hierarchy*: given a pathological sentence such as the Liar, we can produce a first-level reflection by recognizing that, *given its pathological nature*, it cannot be true. Further, given the pathological nature of the Liar and the conclusion expressed in the form of a first-level reflection, we can recognize that the Liar says of itself that it is not true, which is also the conclusion reached in the first level reflection, and hence that it is true after all. This process can be iterated, giving rise to the Reflective Hierarchy. However, it is crucial to Simmons's project that this hierarchy be adequately differentiated from the usual hierarchical approaches.

The point, as Simmons says, is that "the level of a sentence σ in the hierarchy is *not* a measure of an extension of an occurrence of 'true' in σ" (p. 125). Given the occurrences of 'true' in the two reflections on the Liar, the extension of neither one includes the extension of the other. However, both will contain sentences from every level of the reflective hierarchy, since they will contain any true reflection from any level of the hierarchy. These are essentially "anti-Tarskian" features that set the Reflective Hierarchy apart from the hierarchical approaches Simmons has been criticizing.

Then what does the level of σ in the hierarchy measure? Simmons's answer is that "*The level of a sentence merely indicates the level of the highest reflection that the sentence involves*" (p. 125). In other words, the reflective hierarchy is intensional, not unlike Russell's order hierarchy. It is not a hierarchy of extensions (as the type hierarchy can be construed to be), but a hierarchy of expressive resources.

There are other consequences of the singularity proposal that Simmons points out. First of all, the contextualization of the extension of 'true' brings along a restriction on the availability of the Tarski biconditionals. Let c be some context, and let true_c be the extension of 'true' in that context. Then the schema

$$t \text{ is true}_c \text{ if and only if } \varphi,$$

(where t is a term denoting φ) is applicable only to sentences that do not themselves contain occurrences of 'true_c'.

Moreover, Simmons claims that the singularity theory can incorporate the very notions of *groundedness*, *singularity*, and *truth* into the object language for which it

provides an account. This is indeed quite remarkable, since an analogous move in the case of, say, Kripke's truth-gap proposal has been shown to reinstate the paradox. The point is that a sentence such as

(6)   (6) is ungrounded

is certainly ungrounded, and yet has a definite truth value. Indeed, "it is a prominent feature of the singularity theory that ungrounded sentences, sentences without a truth value in certain contexts, do have a truth value in other contexts" (p. 160). How can Simmons get away with this? Part of the reason seems to be that, contrary to truth, ungroundedness is *not* a context-sensitive term. A sentence is grounded or ungrounded *simpliciter*, regardless of context. Similarly, the very notion of singularity can be introduced in the theory, but not as a context-sensitive term.

The situation is even more surprising with the notion of truth, which is context-sensitive. It appears that as soon as such a notion is incorporated in the language one can formulate the *Superliar*:

(7)   (7) is not true in any context.

The singularity theorist wants to claim that the Superliar is not paradoxical, but has a definite truth value: indeed the Superliar is *true*! This notion of truth however, is *context-independent*. It is somewhat baffling to be told at this point that there is, in English, also a notion of context-independent truth, which had not made its appearance before. This reviewer cannot help being left with the impression that Simmons is pulling a rabbit out of the hat. It is true that on p. 174 Simmons claims that this notion is not an altogether different notion of truth, but that 'true' can sometimes be used context-independently. But this does not quite seem to be the case. After having made such a big deal of the fact that 'true' is context-sensitive, it just does not seem appropriate to add, as an afterthought, that there also are context-independent *uses* of 'true:' these are not just *uses*, but rather an altogether different semantic notion from contextual truth, even if Simmons confines it to the meta-language.

How does the singularity proposal fare with respect to semantic universality? It will be recalled that semantic universality is a desirable trait of any semantic account. Simmons claims that natural language is not semantically universal, at least in the following sense: there is no predicate expressing *universal truth*, a predicate, that is, that applies to *all* the truths. Thanks to the Minimality principle, however, we are quite close: the exceptions to the applicability of the truth predicate in any context are kept to a minimum. In any context, 'true' applies to all truths, except its singularities. But this is a limitation we can live with, according to Simmons: "what cannot be expressed within one stretch of semantic discourse can always be expressed in another. We can say everything there is to say, but not all at once" (p. 182).

It is worth noting, by way of conclusion, that the singularity approach, for all its virtues, still has to show its fecundity by fruitfully interacting with other branches of formal inquiry. All the recent theories of truth have established relations to other bits and pieces of mathematics. For instance, Kripke's truth-gap proposal is inspired by certain constructions in the theory of inductive definitions, and in turn it can be used to shed light on the complexity of the fixed points of such definitions and the length of

their closure ordinals. Similarly, the revision-theoretic approach of Gupta and Belnap can be used to analyze circularly defined notions in other regions of mathematics, such as computability and non-well-founded set theory.

No such interaction appears to be taking place, yet, in the case of the singularity approach. Perhaps then the following analogy might turn out to be useful. The singularity approach is motivated by the the desire to keep the exceptions to the universal applicability of 'true' to a minimum. Now, in computer science and artificial intelligence people have developed a number of formalisms designed to represent non-monotonic reasoning, in which a certain inference rule is applied whenever possible, except where its application would lead to inconsistencies. This is a way of keeping the number of exceptions to the rule to a minimum, and has therefore a somewhat analogous inspiration to the singularity proposal. Similarities between non-monotonic reasoning and theories of truth have been noticed before, but it has turned out to be somewhat difficult to get a good grasp of precisely wherein the analogy lies. Perhaps the singularity proposal and formalisms for non-monotonic reasoning can each help shed some light on the other, highlighting at the same time the kind of processes and phenomena underlying both.

REFERENCES

[1] Feferman, S., "Toward useful type-free theories, I," *The Journal of Symbolic Logic*, vol. 49 (1984), pp. 75–111. Zbl 0574.03043 MR 85i:03068 1, 1

[2] Gödel, K., "Russell's mathematical logic," pp. 125–53 in *The Philosophy of Bertrand Russell*, edited by P. A. Schilpp, Library of Living Philosophers, vol. 5, Northwestern University Press, Evanston, 1944. 1

[3] Gupta, A., and N. D. Belnap, *The Revision Theory of Truth*, MIT Press, Cambridge, 1993. Zbl 0858.03010 MR 95f:03003 1, 1

[4] Koons, R. C., *Paradoxes of Belief and Strategic Rationality*, Cambridge University Press, Cambridge, 1992. MR 93d:03029 1

[5] Kripke, S., "Outline of a theory of truth," *Journal of Philosophy*, vol. 72 (1975), pp. 690–716. Zbl 0952.03513 1, 1

[6] McGee, V., *Truth and Paradox: An Essay on the Logic of Truth*, Hackett, Indianapolis, 1991. Zbl 0734.03001 MR 92k:03004 1, 1, 1

[7] Priest, G., "The logic of paradox," *Journal of Philosophical Logic*, vol. 8 (1979), pp. 219–41. Zbl 0402.03012 MR 80g:03007 1

[8] Rescher, N., and R. Brandom, *The Logic of Inconsistency*, Rowman and Littlefield, Totowa, 1979. Zbl 0598.03001 1

[9] Tarski, A., "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica*, vol. 1 (1936), pp. 261–405. Zbl 0013.28903 1

Gian Aldo Antonelli
Department of Philosophy
Yale University
P.O. Box 208306
New Haven, Connecticut 06520–8306
email: aldo.antonelli@yale.edu