Notre Dame Journal of Formal Logic
Volume 35, Number 4, Fall 1994

Book Review

Gupta, A. and N. Belnap, *The Revision Theory of Truth*, MIT Press, Cambridge, 1993.

1 Introduction Gupta and Belnap's new book on the Liar and related paradoxes, *The Revision Theory of Truth*, is a tour de force of technical expertise and philosophical profundity. It is indispensable for all who are interested in the current state of philosophical work on these ancient puzzles. Combining mathematical elegance with pellucid prose, the authors summarize important progress made in recent years (by Kripke, Woodruff, McGee and the authors, among others) in the use of the mathematical theory of fixed points in studying the paradoxes and contribute to further progress in this field by proposing a novel approach and by powerfully illuminating related problems of philosophical methodology. They offer penetrating criticisms of the major families of existing approaches to the Liar, and they develop a novel approach, claiming that truth is a *circular* concept, by which they mean that the semantic significance of truth can be wholly captured by means of a circular definition, consisting of the class of trivial biconditionals identified originally by Alfred Tarski. In my opinion, there are compelling reasons for rejecting this approach, reasons rooted deeply in the philosophical tradition, but my thinking about these matters has been profoundly affected by reading this superb book.

A considerable part of the book is devoted to describing and criticizing alternative theories of semantic paradox, especially the inconsistency view of Tarski and Chihara [5] and the fixed-point approach (based on the seminal paper by Kripke [11]), but the authors have clearly given higher priority to presenting their own novel approach, which consists of the conjecture that (1) the intension of truth is wholly determined by the Tarski biconditionals, and that (2) these biconditionals constitute a circular definition of the concept of truth. The authors put forward the startling claim that, contrary to the whole philosophical tradition, circularity in definition is not necessarily a defect. They develop several viable semantic and proof theories for circularly defined expressions and demonstrate that the introduction of such circular definitions can (much as can inductive definitions) extend the expressiveness of a language. They apply this theory to a number of philosophical problems unrelated to semantic paradox, including the apparently intertwined identity conditions for place

Received January 17, 1995

and for enduring physical object, and the paradoxes of set-membership. The authors' work on circular definitions is of considerable independent interest, however in the end one may evaluate their theory of truth.

Perhaps the greatest contribution of the book is to the task of defining what a theory of semantic paradox must do. The authors present a number of extremely illuminating distinctions, such as: logical vs. nonlogical concepts of truth, normative vs. descriptive theories of paradox, extensionally equivalent vs. extensionally adequate definitions. They provide a very persuasive explanation of how a mathematically complex model can illuminate an everyday, mathematically unsophisticated notion, like truth. They carefully distinguish the function of philosophic analysis of a concept from that of a psychological account of its cognitive operation.

In the first two sections of this review, I will present, first, a summary of the authors' criticisms of four major alternative accounts, and, second, a brief sketch of their theory, viz., that the Tarski biconditionals constitute a circular definition of truth. In the final section, I will present a number of objections to their theory, a few of which are discussed in the book, but most of which are put forward for the first time in this review. I will argue that the authors have not been evenhanded in their treatment of the problem of generality, faulting hierarchical approaches to the Liar (such as those of Parsons [16], Burge [4], Barwise and Etchemendy [3], and Koons [9]) for the lack of an account of generality in natural language which their own theory does not provide. I will also contend that the primacy of truth, the fact that a grasp of the concept of truth is a prerequisite for linguistic behavior in general and for participating in the practice of stipulative definitions in particular, makes truth an intrinsically indefinable notion. Although a circular definition via the Tarski biconditionals may be even intensionally equivalent to our ordinary concept of truth, it cannot be logically adequate for that concept. As a consequence, there are a number of central features of truth for which Gupta and Belnap can provide no account. In particular, they cannot account for the asymmetry of truth and falsehood. I argue that an explication of truth in terms of proper functions (in the teleological sense) provides a more adequate account, and that this account supports the hierarchical approach to the Liar and related paradoxes.

2 Gupta and Belnap's critique of previous accounts Gupta and Belnap subject four previously existing accounts of semantic paradox to scrutiny. These are: (1) the "simple" solution, which denies the existence of genuine paradox by denying the existence of viciously circular propositions; (2) the inconsistency view of Tarski and Chihara, according to which the conventions governing the use of "true" are inconsistent; (3) the natural hierarchy view, proposed by Parsons, Burge, Barwise and Etchemendy, and myself, according to which the extension of "true" in natural language varies from token to token, the various extensions corresponding roughly to the hierarchy of metalinguistic truth-predicates proposed by Tarski as the appropriate method for avoiding contradiction in the semantics of formal languages; and (4) the simple fixed-point approach, in which the extension of "true" is given by some fixed point of a monotone jump operator, an approach first sketched by Kripke in his [11].

The simple solution is one of the oldest and arguably most popular responses to the challenge posed by the Liar. This solution has been proposed by Bar-Hillel [2], Mackie [12], and recently by Sobel [19], although it has ancient and medieval pre-

cursors. According to these philosophers, it is propositions and not sentence-types or sentence-tokens which are the bearers of truth and falsity. Moreover, these propositions (unlike sentences) are not viciously self-referential (on some accounts, they cannot be self-referential in any way). The attempt to formulate a Liar sentence under these stipulations produces a sentence with a nondenoting singular term:

> (L) The proposition expressed by sentence L is not true.

The sentence L does not express a proposition, and the expression "the proposition expressed by sentence L" fails to denote.

Gupta and Belnap offer five objections to the simple solution. Firstly, they point out that paradoxical sentences can be used truthfully as the complement sentence of a belief attribution (as well as attributions of other so-called "propositional" attitudes). This suggests that such sentences do express something like a proposition. Similarly, paradoxical sentences can be meaningfully modalized, which also supports the hypothesis that such sentences express propositions. Thirdly, paradoxical sentences can occur within truth-functional expressions, and some of the resulting compound sentences are arguably truth (because tautologous): "if all Cretans are liars, then all Cretans are liars." Fourthly, Gupta and Belnap point out that this solution is much more plausible for simple, intrinsic liars, like sentence L above, than for contingent liars, like:

> (C) The proposition expressed by the second indented sentence on the third page of Koons's review of Gupta and Belnap's new book is not true.

In the case of an intrinsically self-referential sentence like L, it is clear that such a sentence could have no coherent use and so it is plausible to suppose that it fails to express a proposition. A sentence like C on the other hand is syntactically and lexically indistinguishable from paradigmatically successful attributions of untruth. Thus, the defender of the simple solution must suppose that an utterance's success in expressing a proposition depends on more than its lexical meaning and immediate context: it may depend on facts arbitrarily remote.

Finally, Gupta and Belnap argue that the simple solution falls prey to a strengthened version of the Liar. If we consider a sentence *S* that states: *S* does not express a true proposition, we are forced to apply the simple solution and conclude that *S* does not express any proposition. But, of course, if *S* does not express any proposition, it does not express a true proposition. Thus, *S* itself seems to be a logical consequence of the simple solution. The defender of the simple solution would seem to be simultaneously committed to asserting *S* and to denying that *S* expresses a proposition. This quasi-contradiction could be avoided only by hypothesizing that it is tokens rather than types of sentences that express propositions, and that one of two tokens of the same, nonindexical sentence-type can express a proposition even though the other does not. Such a variation has been proposed recently by Gaifman [6]. However, since this variant of the simple solution is no longer so simple, and since it moves a considerable distance in the direction of the theory of the natural hierarchy, I will take it up further when considering that alternative.

A second approach considered by the authors is the Inconsistency View. As Gupta and Belnap point out, this view has the virtue of simplicity, while still taking the paradoxicality of the Liar seriously. The authors' principal objection to this view

is that it cannot explain ordinary, nonparadoxical uses of "true." They make reference to Curry's paradox to illustrate that one can derive any conclusion whatsoever from instances of the Tarski biconditionals, and that one can derive this conclusion in extremely weak logics, even those, like relevance logic, lacking the principle of *ex falso quodlibet*. For example, there is a direct argument for the existence of God from the Tarski biconditional involving the sentence "If this sentence is true, then God exists." Chihara, a defender of the Inconsistency View, responds in his [5] that although such conclusions do follow from the Tarski biconditionals, which in turn are conventions governing the use of "true," it is not in fact reasonable to do so. However, neither Chihara nor anyone else has proposed a theory of reasonable inference that blocks Curry's paradox while validating ordinary, nonparadoxical uses of the biconditionals and deductive logic.

The third approach criticized by Gupta and Belnap is what I will call the natural hierarchy theory. On this approach, the claim is made that a hierarchy of extensions of "true" is implicit in natural language, a hierarchy analogous to the extensions of "true" in the hierarchy of metalanguages proposed by Tarski as the correct approach to the semantics of formal languages. This approach has been advocated by Parsons [16], Burge [4], and myself [9]. Early suggestions along these lines were made by Ushenko and Donnellan. Recent work by Gaifman [6] and by Barwise and Etchemendy [3] is also closely related to the natural hierarchy approach. Neither Gaifman nor Barwise and Etchemendy make explicit reference to a hierarchy of extensions of "true," but in both cases the most natural interpretation of their work leads in this direction. In both cases, the semantic values of isomorphic tokens can differ. Gaifman's theory essentially distinguishes two levels of "true," one occurring in tokens caught in vicious cycles, and the other occurring in all other tokens, including tokens commenting on pathological tokens from the outside. If we add to Barwise and Etchemendy's theory the assumption that semantic facts supervene on nonsemantic, then (as I have shown in [9]) all occurrences of "true" in Barwise and Etchemendy's Austinian propositions can be assigned an ordinal level in a Tarski-like hierarchy.

Given this approach isomorphic sentence-tokens may differ in semantic value, since the occurrences of the truth predicates in the tokens may be assigned to different levels in the hierarchy. In the case of Burge's work [4], this approach is combined with the use of fixed points pioneered by Kripke. The extension of true_0 corresponds to the extension of "true" in the minimal fixed point of the strong Kleene evaluation scheme. The anti-extension of true_0 consists of all other sentences, including paradoxical ones. The process is then repeated, keeping the interpretation of true_0 fixed, but allowing re-evaluation of "true" at all other levels. The extension of true_1 consists of the extension of "true" in the minimal fixed point of this scheme, and the anti-extension of true_1 again consists of the complement of its extension, relative to the set of sentence-tokens. This process can be extended through the ordinals.

On this view, we have a family of Liar paradoxes, one for each level of the hierarchy. A level-0 Liar is a token that denies its own truth_0. Such a token is not true_0, since it will not belong to the minimal fixed point of the strong Kleene evaluation scheme. However, the token will be true_1, since it correctly asserts that it does not belong to the extension of true_0. Thus, the naive analysis does not, when properly interpreted, lead to a contradiction. Instead, we conclude that the 0-Liar is both untrue_0

and true_1. The appearance of contradiction is created by the fact that the assignment of occurrences of "true" to levels is tacit in natural language. As Gupta and Belnap point out, this account is not complete until a fairly detailed theory of how occurrences of "true" in natural language acquire an appropriate level in the Tarskian hierarchy is added. Burge suggests some informal principles for such an assignment, and in my book I have provided a formal model of one interpretation of these principles, relying primarily on principles of symmetry and charity. I demonstrated that for ordinary, nontheoretical uses of "true" in natural language, only two levels are needed. The authors' principal complaint against the natural hierarchy approach is that it renders certain propositions unsayable, since, as in Russell's ramified type theory, any assertion involving "true" will be limited on this account to a single level in the hierarchy and will thereby lack the sort of generality of which assertions in natural languages like English seem capable. This complaint is a fairly standard one against hierarchical solutions to the Liar. For instance, McGee in his recent book [14] relies very heavily on the intuition that we can make assertions that are not limited to any level or set of levels of a metalinguistic hierarchy. The standard response to this complaint, sketched by Burge and developed further by myself, is to postulate the existence of two types of generality: quantificational and schematic (following Russell's 1908 distinction [17] between "all" and "any"). One can interpret an assertion such as "God knows all truths" as schematic, that is, as simultaneously asserting "God knows all truths_α," for every ordinal α.

Gupta and Belnap object that such an account cannot give the correct reading to an essentially existential assertion, like "God does not know all truths." This utterance could still be interpreted schematically, but then it would be interpreted as saying, for each ordinal α, that there is some truth_α which God does not know, and this is not the reading that Gupta and Belnap intend. However, I agree with Barwise and Etchemendy in doubting that the kind of generality Gupta and Belnap claim for natural language assertions is really possible. Every assertion involves bringing some sort of classificatory scheme to the world. It involves locating the actual world in some kind of logical space (as in Wittgenstein's *Tractatus*). As Simmons has argued [18] the lesson of the diagonal arguments (including the Liar paradox) is that there is no final, all-inclusive scheme of classification, no scheme which cannot be diagonalized out of. Every assertion is about the world-as-surveyable-by some particular scheme, and there is no way for an assertion to generalize over all possible schemes. The sort of absolute existential generality postulated by Gupta and Belnap is not needed in formulating a hierarchical theory of truth, so that approach cannot be charged with defeating itself.

I will argue in Section 4 that Gupta and Belnap's own account limits the generality of natural language in a very similar fashion, so this objection against the natural hierarchy approach does not advantage their approach.

Two chapters of *The Revision Theory of Truth* are devoted to the discussion of the fixed point approach. Gupta and Belnap consider both three- and four-valued logics, four truth-valued schemes altogether: the classical two-valued scheme, the weak Kleene and strong Kleene three-valued schemes, and a four-valued scheme developed by Smiley and Dunn, which is based on a generalization of strong Kleene. They produce a very elegant proof of Tarski's indefinability-of-truth theorem, by considering

classical (two-valued) languages with constants for every open formula in the language. They produce a second version of the theorem by dropping the assumption about names for open formulas and adding the assumptions that the language contains standard names for its formulas and a primitive substitution operator. Finally, they produce the original Tarski theorem, with its use of Gödel numbering, as a generalization of these earlier versions. These theorems establish, not only that truth is not definable in the language of arithmetic, but that no classical language meeting the stipulated conditions contains a materially adequate truth-predicate for itself (as gauged by Tarski's biconditionals).

The authors then turn their attention to the use of fixed points by Kripke [11] and Martin and Woodruff [13] to construct nonclassical languages that do contain their own truth-predicates. There are two elements essential to these constructions: the notion of a ground model, which assigns classical interpretations to all the nonlogical constants except the putative truth-predicate G; and a jump operator, based on the Tarski biconditionals and one of the three- or four-valued truth-valuation schemes. Suppose ρ is one of the valuation schemes. Then corresponding to ρ is a jump operator ρ_M, an operation on the possible interpretations of the predicate G, given a fixed ground model M. Given an interpretation g of G, the interpretation $\rho_M(g)$ assigns each sentence A to the extension, anti-extension, neither, both, of G, depending on whether the sentence A is true, false, neither or both in the model $M + g$. In other words, the interpretation $\rho_M(g)$ assigns to the predication of G (the would-be truth-predicate) to a sentence A the very same value that A takes in the model $M + g$. A fixed point of the operator ρ_M is an interpretation $g*$ of such a kind that $\rho_M(g*) = g*$. The interpreted language $\mathcal{L}_g = \langle L, M + g*, \rho \rangle$ is a language containing its own truth-predicate, G.

A scheme ρ has the fixed-point property iff for all languages L with one-place predicates G and all ground models M of L the jump ρ_M has a fixed point. The classical scheme τ does not have the fixed-point property, however the schemes κ (strong Kleene), μ (weak Kleene), and ν (four-valued) have this property. Thus, nonclassical languages based on these schemes can always be constructed so as to contain their own truth-predicate. Classical languages (even languages containing arithmetic) can contain their own truth-predicates, so long as their resources for self-reference are limited. For example, they can contain standard, quotation-like names for their own sentences, but they cannot define truth on arithmetical codes of these sentences. At the same time, nonclassical languages can be constructed that do not have the fixed-point property. Gupta and Belnap prove, for example, that any three-valued scheme with the Łukasiewicz biconditional $\equiv$ fails to have the fixed-point property. The Łukasiewicz biconditional has the following truth table:

$\equiv$	t	f	n
t	t	f	n
f	f	t	n
n	n	n	t

Likewise, any three-valued scheme with exclusion negation fails to have the fixed-point property. Parallel claims hold for four-valued schemes. Thus, only logically incomplete languages have this property. Moreover, even for logically incomplete lan-

guages, there are related semantic notions which such languages cannot contain. The Kripke and Martin-Woodruff constructions are concerned with "weak" truth predicates: predicates whose semantic value mirrors that of the object sentence. Thus, predicating weak truth of a sentence that is neither true nor false should result in a sentence which is neither true nor false. Strong truth, in contrast, is falsely ascribed to sentences that are neither true nor false, as is strong falsity. As Gupta and Belnap point out, no syntactically rich three-valued language can contain its own strong-falsity predicate, since a sentence saying "I am F" can have none of the values t, f or n.

These variants on the Liar establish the incompatibility of logical richness (containing such connectives as the Łukasiewicz $\equiv$ or exclusion negation), semantic richness (containing such notions as strong truth or strong falsity), and syntactic richness (unlimited resources for forming self-referential sentences). Gupta and Belnap point out, however, that this incompatibility depends on giving traditional semantic interpretations (in terms of fixed extensions and anti-extensions) to semantic predicates. By abandoning this sort of semantics for semantic predicates, Gupta and Belnap's approach escapes (in a certain sense) these limitations. The fixed-point approach, in contrast, postulates limitations of both logical and semantic expressiveness in natural language.

Gupta and Belnap argue that there is a gap of logical expressiveness (involving for instance, the Łukasiewicz biconditional and exclusion negation) between natural languages and those languages that are amenable to the fixed-point approach. Natural languages contain these richer logical notions, or, at the very least, it seems coherent to add them. In order to accommodate these expressive resources, the fixed-point theorist is forced to postulate some sort of hierarchy within natural language, resulting in the natural hierarchy approach discussed above. Gupta and Belnap discuss briefly a novel and interesting version of such a hierarchy: a hierarchy of logical notions, like Łukasiewicz biconditional and strong negation. The Łukasiewicz biconditional of level n could be allowed to interact only with truth-predicates of level less than n. As I discussed above, the authors' principal objections to the natural hierarchy approach turn on the need for a theory of levels and on the problem of unqualified existential generality.

Gupta and Belnap press two additional objections against the fixed-point approach. First of all, they point out that it has undesirable consequences in respect of the properties of logical implication. Classical tautologies, such as "if the Liar is true, then the Liar is true," do not come out as true on the fixed-point approach. Costly revisions to logical practice result. At the same time, fixed-point approaches validate counter-intuitive implications. For example, if the fixed-point approach interprets truth via the minimal (or largest intrinsic) fixed-point of the strong-Kleene scheme, a Truth-teller logically implies a Liar.

Finally, the most important family of arguments against the fixed-point approach developed by the authors is based on certain metaphysical assumptions they share on the relationship between semantic and nonsemantic facts. They propose three theses concerning this relationship: the Supervenience Thesis, the Signification Thesis, and the Local Determination Thesis. The Supervenience Thesis consists of the claim that the semantic facts are wholly determined by the nonsemantic facts: when the

latter are fixed, so are the former. The Signification Thesis adds to supervenience a claim about how the semantic facts are fixed; namely, that the signification of truth in a world is wholly fixed by holding the Tarski biconditionals (properly understood) true in that world. Signification is a technical term introduced by the authors and intended to generalize the familiar notion of extension. Thus, the signification of a predicate in a three- or four-valued model could be given by listing both its extension and anti-extension. On the fixed-point approach (and on the simple and natural hierarchy approaches), the truth-predicate has a definite signification of some kind in every world (including the actual one). The Signification Thesis implies that this signification must be the only one compatible (in the given world) with making all the Tarski biconditionals true. Finally, the Local Determination Thesis incorporates the claim that any given semantic fact is determined by a limited range of nonsemantic facts; namely, those in its dependency range (which can be defined recursively). In other words, changing facts involving properties and objects that are unrelated to a given semantic fact should not alter that fact.

In order for the fixed-point approach to constitute a theory of truth, its proponents must do more than demonstrate that there are languages that contain materially adequate truth-predicates (in the sense of being fixed-points of the appropriate jump operator). Instead, the fixed-point theorist must claim that the signification of "true" (the actual predicate of English) is determined by the corresponding fixed points of this operator. A claim of intensional, and not just extensional, equivalence must be made. Gupta and Belnap use the term "T-predicate" to refer to a predicate that meets the condition of material equivalence to truth, and the term "truth-predicate" to refer to a predicate that meets the more stringent condition of intensional equivalence. Gupta and Belnap argue that the Supervenience Thesis can be used to establish that certain languages that contain materially adequate T-predicates cannot contain their own truth-predicates.

Consider a classical language **L** containing a one-place predicate G. (In fact Gupta and Belnap distinguish between three kinds of "language": L stands for a language syntactically characterized, $\mathcal{L}$ stands for an "interpreted language" (L plus some extensional model assigning denotations to terms and predicates), and **L** stands for a language intentionally characterized (the primitive expressions of L assigned meanings or intensions). The T-predicate is defined for interpreted languages, such as $\mathcal{L}$, while the truth-predicate only makes sense when applied to an intentionally characterized language, like **L**.) Suppose that the constant a designates (in a ground model M that leaves G uninterpreted) the L-sentence "Ga" (a Truth-teller), and suppose that the preferred jump operator ρ_M has two fixed points: g_0 and g_1, the first making "Ga" true and the other making it false. Suppose that this is the only anomaly in $\mathcal{L}$ (the interpreted language based on L and M). G is therefore a T-predicate for both $\mathcal{L}_{g_1}$ and $\mathcal{L}_{g_2}$ (the interpreted languages based on g_1 and g_2, respectively). Both fixed-points verify all of the Tarski biconditionals. Therefore, the Signification Thesis entails that if G means "true in **L**," it must have the same extension in both interpreted languages, which it does not. Consequently, although G can be a T-predicate for L in M (being interpreted by some fixed point of the classical jump operator), G cannot be a truth-predicate for **L**.

The Signification Thesis entails that, for any language **L** and any one-place pred-

icate G, G can mean "true in **L**" only if the appropriate jump operator for L has a unique fixed point for every ground model of L. This poses a dilemma for the fixed-point operator. If the language L is sufficiently rich in syntactic and logical resources, there will be many ground models for which no fixed point for G exists (for example, ground models with Liars, combined with a classical language). Alternatively, if the language L is sufficiently impoverished logically (for example, containing only strong Kleene connectives), many ground models will have multiple fixed points (caused, for example, by the presence of Truth-tellers).

An even stronger result is forthcoming from the Local Determination Thesis. There are languages and models for which the jump operator has a unique fixed point, and yet the resulting extension cannot be that of truth. For example, consider a classical language L and two ground models M_1 and M_2. In M_1, the constant a denotes the sentences "Ga," and the constant b denotes the sentence "$Gb \rightarrow Ga$." In M_2, the constant a denotes "Ga," and the constant b denotes "$Gb \rightarrow \neg Ga$," and these are the only anomalies in either model. The classical jump operator has one fixed point based on M_1, in which "Ga" is true, and it has one fixed point on M_2, in which "Ga" is false. The constant b is not in the semantic dependency range of the sentence "Ga" in either model. Consequently, the Local Determination Thesis entails that if G is a truth-predicate, "Ga" must have the same truth-value in any acceptable model based on either M_1 or M_2. Consequently, G cannot be a truth-predicate for **L**, even though the appropriate jump operator has a unique fixed point everywhere.

The Signification and Local Determination Theses provide grounds for proving intensional versions of Tarski's indefinability theorem. Tarski proved that certain interpreted languages cannot contain their own T-predicates. The authors prove that certain languages (intentionally characterized) cannot contain their own truth-predicates. The above example also supports a surprising result: there are cases in which the Tarski biconditionals, interpreted as material equivalence, determine a unique signification for truth, and yet this signification can be shown, on metaphysical grounds, to be incorrect (note that in the case of three- or four-valued languages, this means interpreted by means of the Łukasiewicz biconditional). The authors take this as demonstrating, not that the Signification Thesis is false (because it is in conflict with the Local Determination Thesis), but that the Tarski biconditionals must be given a new interpretation.

3 Gupta and Belnap's novel proposal Gupta and Belnap claim that the Tarski biconditionals constitute a circular definition of truth, and that truth is therefore a circular concept. They develop, in Chapter 5 of their book, a general theory of circular definitions, which they apply to truth and other concepts in Chapters 6 and 7. A circular definition is to be understood, first and foremost, as a rule of revision, a rule for generating a new and improved extension for the definiendum, given some provisional hypothesis about that extension. The definiendum, as it occurs in the definition, is interpreted according to the provisional hypothesis, and the definition itself picks out a new extension. The most interesting and difficult question for a theory of circular definitions to answer is how to make the transition from hypothetical to categorical judgments. In some cases, although not in all, it is possible to make the categorical judgment that some object does or does not belong to the extension of a given circular

concept. This transition to categorical judgments must meet two desiderata: the definition should yield a conservative extension of the original language, and the defined concept should be ascribed a sufficiently rich content. (We say that a semantic system S is (strongly) conservative iff for all definitions D, all models M, and all languages L, and all sentences A of L (so A contains no occurrences of the definiendum), if A is valid in M according to system S, then A is true in M.)

The theory of circular definitions will result in a threefold distinction among objects: those that are definitely in the extension of the definiendum, those that are definitely not in the extension, and those whose status is problematic. In this sense, their account bears some similarity to the three-valued approach. However, Gupta and Belnap claim that their account is superior to previous three-valued accounts, in that they are able to explain why the problematic cases are neither definitely fish nor definitely fowl.

The authors develop a variety of semantic theories for circular concepts: the family S_0, S_1, S_2, ..., and the systems $S^{\#}$ and S^{*}. Corresponding to the family S_0, S_1, S_2, ..., is a family of logical calculi: C_0, C_1, C_2, ... In these logical calculi, every line of any Fitch-style natural deduction derivation is labeled by some numerical index. Classical rules can be applied to two lines only if they share the same index. To prove "if A then B" by conditional proof, it is necessary to prove B^i from A^i. Suppose that the predicate "x is G" is defined circularly by "$A(x, G)$." The calculus C_0 consists of classical logic plus three additional rules: DfIr, DfEr, and Index Shift.

$$\text{Df Ir} \quad \frac{A(t, G)^i}{[t \text{ is } G]^{(i+1)}} \qquad \text{Df Er} \quad \frac{[t \text{ is } G]^i}{A(t, G)^{(i-1)}}.$$

Applying the circular definition to introduce G involves shifting the index up by one, and using the definition to (partially) eliminate G involves shifting the index down by one. The rule Index Shift allows one to change the index of any line not containing G to any value whatsoever. Each calculus C_n is $C_0 + IS_n$, where IS_n is a rule that permits inferring $A^{(i+n)}$ from A^i, and conversely.

Given a ground model and a definition D for G, we can define a revision operator $\delta_{D,M}$, that takes a hypothetical extension X for G as input and yields a new extension $\delta_{D,M}(X)$. Applying this operator n times in succession results in $\delta^n_{D,M}(X)$.

> A sentence A is valid in M in S_0 iff there is a natural number p such that, for all $q \geq p$ and all subsets X of the universe, A is true in $M + \delta^q_{D,M}(X)$.
>
> A subset X of the domain is n-reflexive in M iff $\delta^n_{D,M}(X) = X$.
>
> A sentence A is valid in M in S_n iff there is a natural number p such that, for all n-reflexive sets X, A is true in $M + \delta^p_{D,M}(X)$.

Gupta and Belnap demonstrate that the calculus C_0 is sound and complete for the semantic system S_0, and in general, C_i is sound and complete for S_i. The system S_0 is conservative, but for $n > 1$, the system S_n is not conservative. Indeed, the family of systems S_n is not conservative in the limit. At the same time, the system S_0, although conservative, ascribes far too weak a content to the definiendum. These deficiencies in S_0 and in the S_n's motivate the construction of two additional semantic systems, $S^{\#}$ and S^{*}.

Before defining these systems, Gupta and Belnap survey the existing state of knowledge about revision sequences (possibly transfinitely repeated applications of the revision rule). Let Σ be a sequence of hypotheses about the extension of the definiendum G, of length $lh(\Sigma)$. An object d is *stably* G in Σ iff there is an ordinal $\beta < lh(\Sigma)$ such that for all γ, if $\beta \leq \gamma < lh(\Sigma)$, then $d \in \Sigma_\gamma$. The definition for stably non-G is symmetrical. If d is stably G, its status as a G is eventually fixed in the sequence of hypotheses. A hypothesis h *coheres* with Σ iff, for all d, if d is stably G in Σ, then $d \in h$. Such a sequence Σ is a *revision sequence* for revision rule ρ iff for all $\alpha < lh(\Sigma)$, if α is the successor of β, then $\Sigma_\alpha = \rho(\Sigma_\beta)$, and, if α is a limit ordinal, then Σ_α coheres with Σ restricted to α, i.e., for all d , if d is stably G in Σ restricted to α, then $d \in \Sigma_\alpha(d)$. The authors prove that if d is stably G in Σ, then d belongs to all hypotheses cofinal in Σ, and conversely, if $lh(\Sigma) = On$ (i.e., the class of ordinals). (A hypothesis h is cofinal in Σ iff for all ordinals $\alpha < lh(\Sigma)$, there is a β such that $\alpha \leq \beta < lh(\Sigma)$ and $\Sigma_\beta = h$.)

A hypothesis h is *recurring* for revision rule ρ iff h is cofinal in some revision sequence Σ of length On for ρ. Intuitively, recurring hypotheses are hypotheses that can survive some arbitrarily long sequence of revisions. A hypothesis h is α-reflexive for ρ iff there is a revision sequence Σ for ρ such that $\alpha < lh(\Sigma)$ and $\Sigma_0 = \Sigma_\alpha = h$; h is *reflexive* for ρ iff h is α-reflexive for some $\alpha > 0$. Gupta and Belnap prove that all and only recurring hypotheses are reflexive. This shows that the notion of recurring hypothesis can be defined in ZFC, without quantification over proper classes. For any revision rule ρ, recurring hypotheses exist. As the authors note, a theorem by McGee relies on the downward Lowenheim-Skolem-Tarski theorem to establish an upper bound on the cardinality of revision sequences needed to define reflexive hypotheses, given the cardinality of the language involved.

The system $S^\#$ is defined by means of the notion of recurring hypotheses. Sentence A is valid in M in $S^\#$ iff for all hypotheses h that are recurring for $\delta_{D,M}$, there is a number n such that, for all $p \geq n$, A is true in $M + \delta^p_{D,M}(h)$. In other words, A is valid in M iff A eventually becomes stably true, whenever the revision process starts with a recurring hypothesis. $S^\#$ is stronger than S_0, and weaker than S_n, for $n > 0$. In finite situations, S_0 and $S^\#$ are equivalent. The unrestricted rules for definitions, DfI and DfE (unrestricted by reference to indices, unlike DfIr and DfEr), hold in $S^\#$ in categorical contexts, but not generally in hypothetical contexts. In hypothetical contexts, the restrictions on indices of DfIr and DfEr must be respected (as in C_0). $S^\#$ is strongly conservative. Kremer recently demonstrated that $S^\#$ is not axiomatizable and that it is at least Π^1_2 (Antonelli proved that it is exactly Π^1_2, see [10], and [1].)

The system S^* is also constructed by means of recurring hypotheses. A is valid in M in S^* iff A is true in all models $M + h$, where h is a recurring hypothesis of $\delta_{D,M}$. The fundamental idea of S^* is that of stability, while that of $S^\#$ is near stability. This can be brought out by introducing the notion of evaluation sequences. Each hypothesis h induces an evaluation $E(h)$ of pairs consisting of an open formula A with n free variables and an n-tuple of objects: $E(h)(\langle A, d\rangle) = t$ iff A is true of d in $M + h$. Consequently, a revision sequence, which is a sequence of hypotheses, induces a sequence of evaluations. A is valid in M in S^* iff $\langle A, \varnothing\rangle$ is stably t in all On-long evaluation sequences of M, and A is valid in M in $S^\#$ iff $\langle A, \varnothing\rangle$ is nearly stably t in all such sequences. A sentence is nearly stably t in some evaluation se-

quence **E** iff there is a $\beta < lh(\mathbf{E})$ such that for all γ, if $\beta \leq \gamma < lh(\mathbf{E})$, then there is a natural number m such that for all $n \geq m$, A is true in $\mathbf{E}_{\gamma+n}$. Stability simpliciter requires that a sentence become permanently true after some point in the sequence; near stability requires only that, after some point, the sentence becomes permanently true except perhaps for finite periods of fluctuation after each limit ordinal.

The system S^* is weaker than the system $S^\#$, and S^* and S_0 are incomparable. S^*, like $S^\#$, validates classical reasoning. In both systems, the simple, unrestricted rules for definitions are available in categorical contexts. Both are strongly conservative, and both are nonaxiomatizable (Kremer [10] and Antonelli [1] also proved that S^* is Π^1_2). There are two closely related differences between the two systems: (i) $S^\#$ does, and S^* does not, allow use of restricted rules of definition in hypothetical contexts, and (ii) $S^\#$, like S_0 but unlike S^*, is ω-inconsistent. Although Gupta and Belnap admit that ω-consistency is a desirable feature of a semantics for definitions, they argue that it is not absolutely compulsory. From the fact that a circular definition yields ω-inconsistency, it does not follow that accepting this definition means rejecting the standard model of arithmetic. This implication would follow if one assumed that circularly defined concepts must be assigned some definite, coherent extension, but this is just what the authors reject. (I will say more on this issue in the next section.)

Systems $S^\#$ and S^* generate a very illuminating set of distinctions among sentences, distinctions analogous to those that might be made in studying the Liar and related semantic paradoxes. A sentence is *categorical* iff it is either stably true in all evaluation sequences or stably false in all such sequences (in the case of $S^\#$, we would define these concepts in terms of near stability, rather than stability simpliciter). A sentence is *paradoxical* iff it is not stable in any evaluation sequence. A sentence is *Truth-teller-like* iff it is stable in all sequences and stably true in some but not all. Indeed, as the authors point out, there is a striking similarity between the semantic phenomena generated by the theory of circular definitions and that generated by theories of truth. All of the strategies used to cope with the semantic paradoxes could be applied to the theory of circular definitions. The authors take this similarity as a strong reason for supposing that truth itself is a circular concept, and that the semantic paradoxes are simply instances of the more general phenomenon of pathological cases generated by circular definitions.

Gupta and Belnap contend that the Tarski biconditionals constitute an infinitistic, circular definition of truth. The definition would take the following form:

$$x \text{ is true } =_{Df} [(x = \ulcorner A_1 \urcorner \,\&\, A_1) \vee (x = \ulcorner A_2 \urcorner \,\&\, A_2) \vee \dots].$$

Clearly, they are not claiming that we can stipulate that, for certain special purposes, we shall mean by "true" what is so defined. Instead, they offer the Tarski biconditionals as a descriptive definition of our ordinary concept. According to the authors, a descriptive definition can be evaluated on three different levels: (1) for extensional equivalence, (ii) for intensional equivalence, or (iii) for cognitive synonymy. Extensional equivalence is an extremely weak condition, and the claim that the Tarski biconditionals determine an extension equivalent to that of truth would meet with widespread acceptance. At the other extreme, few if any would claim that the infinitistic Tarski definition is cognitively indistinguishable from truth. Gupta and Belnap argue that a philosophically viable and interesting analysis of truth should meet the

second condition: intensional equivalence. However, it is clear that, in order to claim that truth is a circular concept, one must do more than demonstrate that it is intensionally equivalent to a circularly defined concept. Gupta and Belnap's Signification Thesis alone is sufficient to establish intensional equivalence, and this thesis could easily be accepted by those (like myself) who reject the circularity of truth itself. We must identify a fourth level of definitional adequacy, stronger than the condition of intensional equivalence but weaker than the impossibly high standard of cognitive indistinguishability. In the next section, I will propose such a standard (a condition of "logical adequacy") and argue that the circular definition cannot meet this standard.

Gupta and Belnap develop two theories of truth, $T^{\#}$ and T^{*}, based respectively on the semantic systems $S^{\#}$ and S^{*}. The stably true and nearly stably true sentences of a model M (based on the classical revision rule τ) are represented by V^{*}_{M} and $V^{\#}_{M}$. They also present a third system, inspired by McGee's work on maximally consistent sets of sentences. A τ-sequence Σ of M is *maximally consistent* iff for all α, $\{A : \Sigma_{\alpha}(A) = t\}$ is maximally consistent. T^{C} is a theory of truth based on S^{*}, limited to maximally consistent sequences.

> $V^{C}_{M} = \{A : A$ is true in all models $M + h$, where h is cofinal in a maximally consistent τ-sequence of $M\}$.

Each of these sets (V^{*}_{M}, $V^{\#}_{M}$, V^{C}_{M}) have the following properties: each are consistent and closed under logical consequence. Each contains every instance of the following law:

> (TL) $\forall xy(In(x, \ulcorner A(z) \urcorner, y)\ \&\ T(x) \rightarrow A(y))$.

In all three systems, the formulas $T^{n}(d)$ and $T^{m}(d)$ (i.e., n and m iterations of the truth-predicate), where $n \neq m$, are equivalent only in categorical contexts. In each system, there will be instances of $\neg(T^{n}d \leftrightarrow T^{m}d)$.

The three systems differ in their treatment of the following three laws:

> $(T\neg)$ $\quad \mathrm{Neg}(x, y) \rightarrow [T(x) \leftrightarrow \neg T(y)]$
> $(T\&)$ $\quad \mathrm{Conj}(x, y, z) \rightarrow [T(x) \leftrightarrow (T(y)\ \&\ T(z))]$
> $(T\forall)$ $\quad \mathrm{UQ}(x, y) \rightarrow [T(x) \leftrightarrow \forall uv(In(u, y, v) \rightarrow T(u))]$.

(In fact, rather than $(T\neg)$ McGee's theorem requires merely the following weaker principle, $(T\neg W)$: $\mathrm{Neg}(x, y) \rightarrow [T(x) \rightarrow \neg T(y)]$.)

These laws are the familiar rules postulated by Tarski for the definition of the truth of logically complex sentences. The law $(T\neg)$ implies that the negation of a sentence is true iff the sentence is not true. $(T\forall)$ implies that if a universal generalization is true, so must be every one of its instances. The universal closures of all three laws belong to $V^{\#}$, $(T\neg)$ and $(T\&)$, but not $(T\forall)$, belong to V^{C}, and none of these belong to V^{*}.

The price which $V^{\#}$ pays for including all three laws is that of ω-inconsistency. This is implied by an important theorem of McGee [14]:

> Any set Γ of sentences of $\mathcal{L}^{+}$ (including "T") that
> (i) contains the truths of $\mathcal{L}$,
> (ii) is closed under first-order consequence,
> (iii) contains $T^{\ulcorner}A^{\urcorner}$ if it contains A, and
> (iv) contains (TL), $(T\neg W)$, $(T\&)$, and $(T\forall)$ is ω-inconsistent.

V^C, in contrast, preserves ω-consistency, but sacrifices ($T\forall$). The proof of McGee's theorem depends on the construction (by Gödelian methods) of a sentence B stating, in effect, "There is some finite number n such that the result of embedding this sentence (B) in n T-predications is not true." Any set of sentences satisfying conditions (i), (ii) and (iv) must contain B itself, and therefore, if it satisfies (iii) also, it must contain $T^n B$, for every B.

The central feature of the revision theory of truth is the acceptance of all Tarski biconditionals as true, so long as the main connective is understood to be one of definitional, and not material, equivalence. Gupta and Belnap acknowledge that accepting the material Tarski biconditional in some cases (such as that of the Liar) leads to contradiction. Accepting the definitional biconditional in such cases does not lead to contradiction, since the use of such biconditionals in hypothetical contexts (which is crucial to the derivation of the Liar paradox) is either forbidden (as in V^* or V^C) or severely restricted (as in $V^\#$).

Gupta and Belnap point out a number of distinctive advantages of their approach. Their theory preserves the underlying logic of the nonsemantic language without any distortion. For instance, one does not have to change a two-valued language into a three-valued one in order to include a self-referential semantics. Secondly, they do not have to place any restrictions on the logical or syntactic resources of the language. There is no bar to including strong negation or the Łukasiewicz biconditional. Their approach enables the construction of a very fine-grained taxonomy of semantic pathology. Finally, they provide a principled explanation of the existence of semantic pathology in the cases of the Liar and the Truth-teller, by unifying these phenomena with the wider class of pathologies arising from circular definitions.

In order to make good on their claim that their approach achieves a substantial theoretical unification of diverse problems, Gupta and Belnap must provide a convincing case for the existence of circular concepts other than the concept of truth. In the last chapter of the book, they consider a number of plausible candidates: set membership, necessity, belief, and body/space identity conditions. In the case of set membership and in that of necessity, any circularity is parasitic on the circularity (if any) of truth. Their discussion of the relationship between the semantic and set-theoretic paradoxes, however, is quite illuminating. They point out that devices such as the iterative hierarchy of Zermelo-Frankel set theory, like the Tarskian hierarchy of metalanguages, provide technical solutions to the engineering problem of doing mathematics in the neighborhood of paradox, but it does not resolve the underlying philosophical dilemma. The postulation of the existence of sets defined by abstraction is clearly related, at its roots, to Frege's conception of sets as the extension (or exemplification-range) of a concept or property. The relation of exemplification of a property (an intensional notion) is conceptually prior to the relation of membership in such an extension. Property-exemplification leads to paradoxes exactly analogous to the paradox of the Liar: one need only consider the property of not being self-exemplifying. The analogy is so strong that it is quite plausible to claim that there is only one paradox here. The revision theory of truth is applicable to the property-exemplification Liar: since exemplification is circularly defined, the relation has no definite significance. Instead, which properties are exemplified by which things changes from one stage of revision to the next.

It is the principle of extensionality (the notion that co-extensive sets are identical) that sets Russell's paradox apart from the semantic and the property-exemplification Liars. If set-membership varies from stage to stage (as does property-exemplification), and if sets are assumed to be definite entities with definite identity conditions, then the revision theory of truth leads to conflict with the principle of extensionality. Instead, Gupta and Belnap propose a version of Russell's No-class theory of sets. Sets are taken to be theoretical fictions, consisting of a refusal to distinguish (for certain purposes) between co-extensive properties. Since property-exemplification lacks a definite, fixed extension, the very identities of sets defined by abstraction are potentially indeterminate.

Gupta and Belnap claim that necessity is a circular concept whose circularity is distinct from that of truth, but the basis for the distinctness claim is hard to make out. They claim that existing explanations for the truth of modal propositions points to some sort of circularity. For example, in possible-worlds semantics, the truth of $\Box p$ at world w is explained in terms of the truth of p in all (accessible) worlds, including (in systems as strong as T) world w itself. If the proposition p is a circular proposition, like the modal Liar "this sentence is necessarily not true," then the circularity is made apparent. However, this seems to be merely the application of the theory of the circularity of truth in general to the special case of truth in a world. It is hard to see how a genuinely distinct case of conceptual circularity has been produced.

In the case of belief, Gupta and Belnap point out that belief is a circular concept according to the doctrine of Functionalism. The Functionalist claims that to believe that p is to have a certain characteristic disposition to act, given other beliefs and desires. However, as the authors admit, most defenders of Functionalism would deny that this definition is inherently circular; instead, they would claim that it is merely part of a simultaneous definition of a system of mental concepts in terms of the causal relationships between internal neural states, perceptual inputs and motor outputs.

The existence of Believer and Rational-Believer paradoxes—such as "this sentence is not believed by k" or "this sentence could not be reasonably believed by k"—poses a serious challenge to Gupta and Belnap's theory. Given the failure of their appeal to Functionalism, Gupta and Belnap lack a convincing explanation for any supposed circularity in the concept of belief. Rational or rationalizable belief, in contrast, would appear to be infected by a kind of circularity. One rationally believes p when one does so while giving all due weight to every relevant consideration of which one is aware, which considerations may include questions about the extension of the concept of rationality itself. In order to decide what I must believe in a certain kind of competitive setting, it may be incumbent on me first to decide what you must believe, but this in turn may depend on my deciding what you must believe about what I must believe, which (if your information about my situation is accurate) may simply amount to deciding what I must believe. However, it is not at all clear that this circularity, a circularity of rational grounds or evidence within a particular situation, has anything to do with any circularity in the *concept* of rationality itself.

Finally, Gupta and Belnap find a quite plausible candidate for circularity in the concepts of spatial location and enduring physical object. They contend that it is quite natural to specify the identity-conditions of enduring physical object in terms of spatio-temporal continuity of physical qualities, which presupposes definite iden-

tities for spatial locations. At the same time, it is natural to specify the identity conditions of spatial locations in terms of spatial relations to physical landmarks, which are enduring physical objects. In many cases, these intertwined specifications lead to no indeterminacy. In others, Truth-teller-like pathologies can emerge, for example, in the famous imaginary universe of Black, in which two qualitatively identical iron spheres rotate around their common center of gravity. In such universes, the identities of the spheres, and of the places they occupy, are underdetermined by the usual conventions. This is an intriguing suggestion, and it opens up a very inviting line for future research.

Gupta and Belnap note (in Chapter 6) a corollary of their approach for the treatment of modality. They remind their readers of a well-known argument by Montague against the syntactic treatment of modality. Montague [15] referred to work on the Paradox of the Knower by himself and Kaplan in [8]. He generalized these earlier results by noting that nearly every popular modal logic validated the very principles required for the construction of the Liar-like Paradox of the Knower, except one: standard modal logics treat necessity as a connective or statement operator, not as a predicate of sentence-like structures. This feature of standard modal logics severely limits their resources for self-referential attributions of modality, blocking, for example, the construction of a sentence equivalent to "this sentence is not necessarily the case." Montague took this fact as providing compelling ground for treating modality as an operator, and not as a predicate of syntactically definable structures.

However, as Gupta and Belnap point out, once we have an acceptable way for dealing with the Liar, the need for avoiding analogous paradoxes, such as that of the Liar, becomes much less pressing. Moreover, it is possible to add operator-like constructions (in particular, a substitution operator) to standard modal logics, thereby enriching their capacities for self-reference and opening up again the potential for the Paradox of the Knower. Analogously, it is possible to treat necessity as a predicate, so long as the resources for self-reference are limited in some other way.

4 Objections to Gupta and Belnap's approach In this section, I would like to consider six objections to Gupta and Belnap's claim that truth is a circular concept, three of which they present and discuss in their book, one of which is at least implicitly acknowledged by the authors as a difficulty, and two of which are original, appearing for the first time in this review. The three discussed by Gupta and Belnap are: (1) the argument that their account is too complex to be a plausible representation of our naive concept of truth, (2) the problem of the Strengthened Liar—"this sentence is not categorical," and (3) the impossibility, on their account, of the existence of a semantically universal language.

A frequently heard objection to much recent work on the Liar is that the mathematically sophisticated constructions they involve cannot possibly be needed to model a simple, everyday concept like that of truth. Gupta and Belnap offer decisive responses to this objection. The fundamental idea of their approach—that truth is a circular concept—is a relatively simple one. Moreover, as they point out, any complexity in their constructions is a consequence of their attempt to provide semantic rules that are adequate for any arbitrary situation, including situations involving infinitely elaborated cross-reference. For ordinary, finite situations, much simpler con-

structions would suffice. Finally, the authors point out that there is an important difference between the description of a concept and the description of the cognitive psychology of a concept. The second sort of description may ignore applications of the concept that exceed the practical capacities of human users, but must not ignore systematic patterns of error. The burdens are exactly reversed in the case of the first sort of description, which is what Gupta and Belnap offer for truth.

The second objection to Gupta and Belnap's approach that they explicitly address is the reliance on the Strengthened Liar or Liar's Revenge phenomenon. Any proposed solution to the Liar that employs theoretical resources in its metalanguage that it denies to the object language under study will be vulnerable to this objection, since there will be versions of the Liar paradox, expressible by means of the additional resources available in the metalanguage, whose solution is not provided for by the original solution itself. A successful solution to the semantic paradoxes of this type must provide a schematic solution, that can be applied and reapplied to level after level in the hierarchy. This is exactly the defense offered by Gupta and Belnap. In their case, it is the notion of categoricalness that is available in their metalanguage but not in the object languages for which they provide a systematic account. They are concerned with languages for which truth is the only problematic concept, but in analyzing this concept, they introduce new, admittedly circular concepts (such as pathologicalness, paradoxicality, and categoricalness) which are also susceptible to paradox.

For example, using the resources of the authors' metalanguage, we can construct a sentence A that states, "Sentence A is either not true or not categorical." If we say that A is categorical, then it must be either definitely true or definitely false. It cannot be both categorical and definitely true, since then both of its disjuncts would be false. It cannot be both categorical and definitely false, since then its first disjunct would be definitely true. Therefore, A must not be categorical. But, if A is definitely not categorical, then its second disjunct is definitely true, and A must be definitely true (and thus categorical). Gupta and Belnap respond that their revision theory can simply be reapplied to the metalinguistic notion of categoricalness. This notion is also inherently circular, resulting in cases that are not definitely categorical nor definitely not categorical. The above Strengthened Liar argument can then be blocked in a number of places. Most centrally, the argument depends upon using the definition of categoricalness in hypothetical contexts, and on doing so without reference to stages of revision. At most, what the argument succeeds in showing is that if A is judged to be categorical at one stage, it will be judged to be not categorical at the next stage, and vice versa.

Once the revision theory is applied to the basic notion of categoricalness, a new concept of hypercategoricalness is needed in a new metalanguage in order to provide an adequate account of the semantics of first-order categoricalness. This new concept yields yet new paradoxes, and so on, ad infinitum, the Tarskian hierarchy lives!

Gupta and Belnap's response is itself subject to a further objection, the third objection addressed in the book. McGee [14] and Simmons [18] have both argued that an adequate response to the semantic paradoxes must solve the Problem of Semantic Sufficiency, that is, it must show how to give the semantics of some rich language wholly within that language itself. They claim that this must be done, since we know

that there are rich languages that have this capacity, namely, natural human languages. Natural language is "universal," in the sense that anything whatsoever that can be expressed can be expressed in it.

Gupta and Belnap reply that there is no good reason to believe that any natural language is universal in this sense. It is clearly false that any given natural language can, without alteration or addition, express anything whatsoever. Could the theory of quantum mechanics be expressed in Old English, without the importation of any new vocabulary or new meanings for old vocabulary? To say that anything can be expressed in any natural language, once sufficient resources have been added to it, is to trivialize the claim and render it innocuous in relation to Gupta and Belnap's proposal. Gupta and Belnap are quite happy to accept the hypothesis that, for any semantic concept *C*, there is a language that contains its own *C*-concept, but they are unwilling to embrace the thesis that there is a single language that contains, for every semantic concept *C*, its own *C*-concept.

As a response to the supposed Problem of Semantic Sufficiency, Gupta and Belnap's assertions are quite plausible and cogent. However, there is at least an apparent incompatibility between their response to McGee and Simmons's objection and their own principal objection to the natural hierarchy approach. As a reminder for the reader, I will repeat Gupta and Belnap's objection to the theory that natural language contains (implicitly) a hierarchy of extensions for the truth-predicate. The authors argued that certain kinds of semantic generality are in fact possible in natural language but cannot be allowed for in the natural hierarchy account. For example, one can state "There are truths the devil does not know," intending to state, not that at every level α in the hierarchy there are truths$_\alpha$ that the devil does not know, but that there is some level α and some truth$_\alpha$ that the devil does not know. This the natural hierarchy account cannot acknowledge as a genuine possibility, at least not if the quantification over levels is supposed to include the resources of any possible metalinguistic evaluation of the utterance. But what is the authors' basis for claiming that such generality is in fact possible in natural language? If this were a cogent objection to the natural hierarchy approach, an equally cogent objection could be lodged against Gupta and Belnap's solution. I could insist that in uttering sentence *A*, "Sentence *A* is either not true or not categorical," I intend the predicate "categorical" to include, not just the first-order notion of categoricalness used in Gupta and Belnap's book, but every sort of higher-order, metalinguistic categoricalness. In stating that *A* is not categorical, I mean that it is not categorical at any level in the Tarskian hierarchy of metalanguages. To be consistent, Gupta and Belnap must insist that such intentions cannot be fulfilled by any such speech act. An exactly analogous reply to Gupta and Belnap's original objection is available to the defender of the natural hierarchy approach.

In response to this argument, Gupta and Belnap press two points (private correspondence, September 5, 1994). First, they argue that the situation with "categoricalness at any level" is not identical to that with "true." The expression "categorical" is not a part of "*ordinary* English": they propose adding it to English. In contrast, "true" is already incontestably part of English. However, if, once the expression "categorical" has been added to English, we (as speakers of the enhanced version of English) have exactly the same intuition about the generality of certain uses of "categorical" as we do about parallel uses of "true," then these intuitions should be treated on a par:

the fact that one intuition concerns a well-established word and the other concerns a novel one does not seem relevant. Gupta and Belnap must dismiss the intuition about the potential generality of uses of "categorical" as illusory. To be consistent, they should admit that parallel intuitions concerning "true" are at least suspect.

Secondly, Gupta and Belnap argue that:

> The disagreement is over the signification of truth (in cases where truth is the only problematic element *n* the object language). The bearing of the treatment of "categoricalness" on this fundamental issue seems to us to be minimal.

But the bearing of this issue on the fundamental question is quite clear. There are two competing accounts of the signification of truth: the Burgean natural-hierarchy account and Belnap-Gupta revision theory. Gupta and Belnap contend that the Burgean account does violence to certain intuitions about the meaning of sentences like "some truths are not known by the devil." I am arguing that we have identical intuitions about sentences like "some categorical truths are not known by the devil," once Gupta and Belnap introduce the expression "categorical." Gupta and Belnap must override the very intuition in this case that the Burgean theory overrides in the case of simple truth. Hence, this objection to the Burgean account canceled by an equal and opposite objection to their own account.

There is one objection to the revision theory of truth that is at least implicitly acknowledged by Gupta and Belnap to point to a weakness in their theory. Gupta and Belnap admit that their approach runs afoul of the dilemma posed by McGee's theorem concerning ω-inconsistency. The four laws of truth, TL, $T\neg W$, $T\&$ and $T\forall$, are principles that any adequate theory of truth must validate, yet McGee's theorem demonstrates that any theory doing so that does not allow for contextual shifts in the extension of truth and that affirms $T\varphi$ whenever it affirms φ suffers from ω-inconsistency. McGee's theorem is a powerful reason for embracing Burge's version of the natural hierarchy view. On Burge's approach, the ω-inconsistency demonstrated by McGee's proof is, like the simple inconsistency demonstrated by the standard Liar argument, only apparent. On my version of Burge's theory, we would interpret the paradoxical sentence B as claiming that, for some finite number n, the result of applying n occurrences of "true" to B is not true_0. We would further interpret each sentence "$\text{true}^n B$" as applying the property of truth_1 to B, n times. We would then conclude that B is not true_0, but that it is true_1 (and $\text{truly}_1\ \text{true}_1$, $\text{truly}_1\ \text{truly}_1\ \text{true}_1$, etc.). In other words, not every result of applying 0-or-more predications of "true" to B is true_0 (indeed, none of the resulting propositions is true_0), but every result of applying any number of predications of "true" to B is true_1.

Gupta and Belnap have made the following reply to this objection, which I will quote at length:

> McGee's theorem does not, it seems to us, provide any reason for embracing the hierarchy view. Insofar as the preservation of the semantic laws $T\&$, etc. is a desideratum for a theory of truth, the theorem provides a reason for rejecting all theories that do not make truth ω-inconsistent. Insofar as this is not a desideratum, the theorem is neutral on competing theories. Put another way: the hierarchy theories do not preserve the *unrestricted* [emphasis theirs] versions of the semantic laws that are shown to lead to ω-inconsistency in McGee's theorem. The hierarchy theories preserve only restricted versions of these laws, and these restricted versions are validated by nonhierarchical theories (*Ibid.*).

Gupta and Belnap overlook the fact that, on Burge's account, the ω-inconsistency demonstrated by McGee's theorem is rendered innocuous, just as is the inconsistency of the Tarski biconditionals. The ω-inconsistency is understood to be only apparent: consistency is restored when the implicit indices on "true" are taken into account. We affirm, of each member of an infinite class $\mathcal{A}$, that it is not true_0, while simultaneously asserting the generalization that some member of $\mathcal{A}$ is true_1. Thus, it is possible for a Burgean theory to preserve *unrestricted* versions of the semantic laws (so long as $T\neg W$ is substituted for $T\neg$, as in McGee's original theorem).

One explanation of the fact that Gupta and Belnap overlook this possibility is that they assume that we must employ a *weak* notion of truth, in the sense that "A is true" should have a truth-value gap whenever A does. A Burgean account incorporates elements of both the weak and the strong notions of truth. For example, if A has a truth_0-value gap, then so does "$\neg true_0(\text{`}A\text{'})$." However, the fact that A has a truth_0-value gap does not necessitate that either A or "$\neg true_0(A)$" have truth_1-value gaps. Truth_1 involves the closing off of the extension of truth_0. Consequently, the truth_0-Liar L—viz., "$\neg\text{true}_0(L)$"—is true_1, since L does not in fact fall in the extension of "true_0" (which is just what L says). In the case of Burge's analysis of McGee's infinitary liar B—viz., "no result of appending 'true' finitely often to M is true_0"—we can similarly affirm that B is true_1, albeit suffering a truth_0-value gap.

Gupta and Belnap claim that none of their objections to the Inconsistency View apply to the ω-Inconsistency View involved in embracing $V^{\#}$. However, it is hard to see how this claim can be sustained. Consider the Curry-like paradox C:

(C) If every finite number of applications of "true" to C is true, then God exists.

By a proof analogous to McGee's, we can show that if Γ contains (TL) $(T\neg W)$, $(T\&)$, and $(T\forall)$ and the truths of $\mathcal{L}$, and if Γ contains "true φ" whenever it contains φ, then Γ must contain both (C) and "$\text{true}^n(C)$," for every n. If we require that Γ be closed under Rosser's ω-rule, then, even if we substantially weaken the remaining logic (say, to relevance logic), we will still be forced to conclude that Γ contains every sentence whatsoever (under the stipulated conditions). Thus, the usefulness of the ordinary conventions of truth turns on the unavailability of infinitary inference rules, like Rosser's ω-rule. But surely there is nothing unreasonable about Rosser's rule. When I am aware that it applies, I would be remiss in not using it. Consequently, Gupta and Belnap face a dilemma. They must either (1) provide some defense for the claim that it is reasonable to neglect such infinitary inference rules, even when they are known to apply, or (2) provide us a substantial account of "reasonable inference" (of exactly the kind they demand of Chihara) to block such Curry-like inferences to arbitrary conclusions.

I have two original objections to urge against Gupta and Belnap's approach. In the first place, their account overlooks the fact that every act of definition presupposes the concept of truth, and in the second place, they are unable to account for the asymmetry of truth and falsity. Gupta and Belnap wish to convince us, not only that the Tarski biconditionals fix the signification of truth in each possible world, but that truth itself is a "circular concept." They propose that the Tarskian Convention T constitutes an infinitary definition of truth. We can use the Tarski biconditionals to define a new notion, Tarski-wahrheit, which is certainly a circular concept, since the Tarski biconditionals are circular. Gupta and Belnap argue that truth and Tarski-wahrheit

are intensionally equivalent. However, is that enough to enable us to conclude that truth itself is a circular concept? I would argue that at least two additional premises are required: (1) it must be the case that the only way to construct an artificial concept intensionally equivalent to truth involves the use of a circular definition, and (2) it must be the case that the concept of truth could be introduced for the first time to a cognitive agent by means of the deployment of such a definition. Whatever may be the status of the first premise, the second premise is clearly false.

The authors claim that Tarski-wahrheit is not only intensionally equivalent to truth but also the correct definition (in the philosophic, descriptive sense) of truth. What makes a proposed definition or analysis adequate from a philosophical point of view is of course a problem dating back to Plato. As the authors state, perfect synonymy (in the sense of cognitive equivalence) is too strong a condition. Even definitions like "a bachelor is an unmarried male adult" or "a circle is a set of points on a plane equidistant from some point" could not meet so stringent a condition. Indeed, no definition that is in any way informative or enlightening could meet this condition, since the very fact that it was enlightening would reflect some cognitive distinction between definiendum and definiens. At the same time, it is clear that intensional equivalence alone is not an adequate criterion. The condition "either a bachelor or such that $0 = 1$" is intensionally equivalent to bachelorhood, but this fact does not show either that bachelorhood involves such concepts as zero, one, and identity, nor that bachelorhood is inherently circular (despite the fact that "bachelor" occurs in this purported definition). In fact, the authors themselves distinguish between extensional equivalence and extensional adequacy, and this sort of distinction can be extended to a distinction between logical or analytic equivalence and logical or analytic adequacy. According to the authors' distinction, the definition "either a bachelor or an unmarried male adult" (let's call this $B1$) is extensionally equivalent to "bachelor," but not extensionally adequate for it. If the meaning of bachelor is treated as fixed, then the circular definition $B1$ is true of exactly the bachelors, but if the meaning of bachelor is not taken as antecedently fixed but rather as determined *ab initio* by the proposed definition, then the signification of the definition does not coincide with that of "bachelor." Any bachelor is definitely in the extension of $B1$ (since he satisfies $B1$'s first disjunct), but any nonbachelor is neither definitely in nor definitely out of the extension of $B1$ (treating the meaning of "bachelor" as a variable to be fixed by the definition). If I assume that my chair is a bachelor, then it is (relative to that hypothesis) in the extension of $B1$; if I assume that my chair is not a bachelor, then relative to that hypothesis, it is in the antiextension of $B1$.

Why must a satisfactory philosophical definition be adequate and not just equivalent to the definiendum? Presumably because the definition is supposed to provide a hypothetical origin for the concept. A definition is philosophically adequate just in case, were someone lacking the concept of the definiendum to acquire a new concept through acceptance of the definition as a stipulation, the new concept so acquired would be intensionally equivalent (coextensive in all possible worlds) to the definiendum. The philosophical analyst should not be concerned with how normal human beings in fact acquire the concept of the definiendum—whether, for example, the concept is innate or acquired—since this is the province of the developmental psychologist, not the philosopher. The philosopher investigates how the concept in question

(or one intensionally equivalent to it) could be acquired by an otherwise normal human being who is lacking it. This investigation reveals something about the essence of the concept, abstracting from accidental features of the concept's etiology.

Suppose that Alfred is an otherwise normal human being lacking the concept of truth (and other closely related semantic concepts, like those of denotation and veridical representation). Could Alfred come to acquire the concept of Tarski-wahrheit through acceptance of the Tarski biconditionals as an infinitistic, stipulative definition of this notion? Clearly, he could not, and this for several reasons. Firstly, no one can understand any indicative sentence without grasping (at least implicitly) the concept of truth. Consequently, Alfred cannot understand any part of the definiens. Secondly, no one can understand truth-functional connectives without grasping the concept of truth (and its complement, falsity). Alfred can therefore not understand the infinitistic definition as a whole, since conjunction and disjunction play an ineliminable role in it. Finally, no one can understand a speech act as a stipulative definition without quite explicitly comprehending the concept of truth. A stipulation of the meaning of φ is a speech act whose content is φ: so understand φ so as to make the following statement true. Alfred cannot participate in such a sophisticated linguistic practice without prior understanding of semantic notions.

Thus, no hypothetical origin for the concept of truth relies on such sophisticated linguistic practices as stipulative definition. This means that truth is essentially indefinable. Hence, it is not definable in a circular fashion and it is not (in Gupta and Belnap's sense) a circular concept.

In response to this objection, Gupta and Belnap insist on distinguishing between a definition's being intensionally adequate (extensionally adequate in all possible worlds) and its being logically or analytically adequate. They argue that they are committed only to the weaker of the two claims: that the Tarski biconditionals are intensionally adequate for the concept of truth. They claim that intensionally adequate formulas deserve to be considered definitions, of a sort, and they urge that their demonstration that an infinitary circular formula is intensionally adequate for truth proves that truth is, in an important sense, a circular concept (*Ibid.*).

How are Gupta and Belnap's notion of intensional adequacy and my notion of logical or analytic adequacy related? In both cases, the intuition behind the requirement appeals to a kind of hypothetical origin or introduction: mine to a hypothetical introduction of a cognitively novel concept, and theirs to a hypothetical introduction of a linguistically novel expression. Gupta and Belnap insist that they are interested here only in the narrower linguistic question: how might the intension of "true" be fixed for someone who is *ex hypothesi* unaware of its meaning? Let's say that truth is linguistically circular iff there is an intensionally adequate circular definition of it, and that it is cognitively circular iff there is a logically adequate circular definition. I insist that truth is not cognitively circular, and Gupta and Belnap claim that it is linguistically circular. As they point out, these two claims are compatible.

What then is the relationship between circularity and the revision theory of the signification of truth? Clearly, if the revision theory is the correct account of the signification of truth, then truth is linguistically circular. Moreover, the converse is also true: if truth is linguistically circular, then the Tarski biconditionals must constitute an intensionally adequate definition of truth, and only revision theories of the signi-

fication of truth respect this requirement. On a natural-hierarchy approach, it is the Tarski biconditionals, *together with* certain kinds of facts about the semantic interconnections within the network of sentences or propositions, that determine the interpretations of occurrences of "true."

What evidence is there for the claim that truth is linguistically circular? It is certainly plausible to claim that Tarski-wahrheit is intensionally or necessarily equivalent to truth. From this, however, it does not follow that Tarski-wahrheit is intensionally adequate to truth. As Gupta and Belnap admit, it is possible to give a natural-hierarchical account of the intension of such a circularly-defined expression. Given that an expression is introduced by a circular definition, how do we know whether to use revision theory or hierarchy theory in giving its semantics? The answer seems clear to me: revision-theoretic semantics is appropriate if *and only if* the expression is *cognitively* circular.

If an expression is not cognitively circular, then a circular "definition" (in Gupta and Belnap's weak sense) cannot be used to determine the expression's intension. Instead, one must look to the intension determined by the original, possibly indefinable concept. The circular "definition" merely agrees with and thereby picks out this pre-existing intension; it does not explain it. The Signification Thesis is plausible only for cognitively novel expressions. Since truth is not cognitively circular, we must not (contra Gupta and Belnap) attempt to complete the semantics of "true" without first settling on a substantive theory of truth.

In my last and final objection, I present such a substantive theory of truth and use it to motivate the hierarchical account of the semantics of "true." This final objection concerns the asymmetry of truth and falsity. These are not symmetrical complementary qualities, like red and green or left and right. Truth is more fundamental than falsity; falsity is the absence of truth, and not vice versa. In contrast, Gupta and Belnap's theory treats truth and falsity as exactly on a par. Truth can be defined circularly, and so can falsity:

$$x \text{ is true} =_{Df} [(x = \ulcorner A_1 \urcorner \& A_1) \vee (x = \ulcorner A_2 \urcorner \& A_2) \vee \dots]$$
$$x \text{ is false} =_{Df} [(x = \ulcorner A_1 \urcorner \& \neg A_1) \vee (x = \ulcorner A_2 \urcorner \& \neg A_2) \vee \dots].$$

If the two are definable in symmetrical ways, why do we all strive to achieve truth and avoid falsehood in our beliefs? Why is truth-telling approved and encouraged, and falsehood-telling penalized? One might try to respond by claiming that it is merely a convention that we all aim at truth rather than falsity in our public pronouncements, but this would mean that the intensions of truth and falsity would be radically underdetermined by our linguistic practices. Imagine two linguistic communities *A* and *B*, speaking morphologically and syntactically identical languages and displaying exactly isomorphic linguistic behavior. Suppose that in community *A*, the convention of truth-telling holds, while in community *B* it is the convention of falsehood-telling that is practiced. Complementary semantics can be given for the two languages: the extension of a predicate, say "red," in language *A* would be the anti-extension of its morphological counterpart in language *B*. The conjunction sign of A would be morphologically identical to the disjunction sign of *B*, and so on for the other connectives and quantifiers. Thus, two radically different semantic theories could be fit to exactly the same behavior. Of course, community *B* is being

misdescribed in the preceding thought-experiment. It is not possible that a community adopt falsehood-telling as its governing linguistic convention, but nothing in Gupta and Belnap's account of truth explains why this is so.

What would an explanation of this fact look like? To begin with, we must take into account the fact that the word "true" has nonsemantic applications, and it is not obvious that we are dealing with mere homonymy here. For example, we can speak of a "true" victory or a "true" friend. A true φ is something that exemplifies fully the essential features or constitutive function of a φ. Thus, a true sentence or statement is one that fulfills the function of a statement, namely, the function of carrying information about the world. (To be more precise, each sentence of a language has its own specialized function: to convey some particular bit of information.) A false statement is something that is purported or intended to be a statement but which fails to fulfill this function. A true statement is just a statement of the facts, while a false statement is something that resembles a true statement but somehow falls short of stating the facts. Truth is fundamental, and falsity derivative. (I would contend, analogously, that affirmation is fundamental and that denial and negation are secondary. For instance, it would be possible to imagine a community *C*, speaking a behaviorally indistinguishable language from *A* and *B*, in which everyone aims at truthfully denying each spoken sentence. However, such a community is not really possible, since denial is conceptually parasitic on affirmation.)

The Liar sentence can be paraphrased as saying "I do not fulfill my function." Since the function of a statement is to carry some information about the world, a viciously circular statement like the Liar clearly fails to fulfill this function. But now we have said exactly what the Liar itself says, namely, that it does not fulfill its function. What has happened is that the Liar sentence has acquired a new function, that of conveying the information that it has failed to fulfill its original function. Thus, we must distinguish two functions of the same sentence: the first, the function of simply and straightforwardly conveying information about the world (a function that only grounded sentences can perform), and the second, the function of conveying information about which sentences do and do not fulfill their primary functions. Consequently, a sentence can be true in either of two ways, and the word "true" must be recognized as having multiple, related extensions. A statement that fulfills the primary function of its sentence is true_0, and one that fulfills the secondary function is true_1.

What about a sentence of the form "I do not fulfill any of the functions of this sentence?" This sentence too can acquire a variety of functions, depending on the domain of quantification about which it is said. The diagonal argument demonstrates that there is no single domain that includes all possible functions. Given any domain about which a given sentence can be interpreted as speaking, there are functions for that very sentence that do not belong to the domain, including, for instance, the function of stating that that very sentence cannot be used to fulfill any of the functions in that domain. The impossibility of unsurpassable generality is the ultimate lesson the Liar teaches us.

Acknowledgments I would like to thank Professor T. K. Seung, my colleague at the University of Texas at Austin, for his invaluable help to me in developing the objections in the

last section of the paper. I would also like to thank Professors Belnap and Gutpa for their very stimulating and thoughtful comments on an earlier draft of this paper.

REFERENCES

[1] Antonelli, A., "The Complexity of Revision," *Notre Dame Journal of Formal Logic*, vol. 35 (1994), pp. 67–72. Zbl 0801.03021 MR 95d:03042 3, 3

[2] Bar-Hillel, Y., "Do natural languages contain paradoxes?," *Studium Generala*, vol. 19 (1966), pp. 391–397. Zbl 0147.24612 2

[3] Barwise, J., and J. Etchemendy, *The Liar*, Oxford University Press, New York, 1987. Zbl 0678.03001 MR 88k:03009 1, 2

[4] Burge, T., "Semantical Paradox," *Journal of Philosophy*, vol. 81 (1979), pp. 5–28. 1, 2, 2

[5] Chihara, C., "The semantic paradoxes: A diagnostic investigation," *Philosophical Review*, vol. 88 (1979), pp. 590–618. 1, 2

[6] Gaifman, H., "Operational Pointer Semantics: Solution to the Self-Referential Puzzles, I," pp. 275–292 in *Proceedings of the Second Conference on Theoretical Aspects of Reasoning about Knowledge*, edited by M. Vardi, Kaufmann, Los Altos, 1988. Zbl 0705.03005 2, 2

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

[8] Kaplan, D., and R. Montague, "The Paradox of the Knower," *Notre Dame Journal of Formal Logic*, vol. 1 (1960), pp. 79–90. 3

[9] Koons, R., *Paradoxes of Belief and Strategic Rationality*, Cambridge University Press, New York, 1992. MR 93d:03029 1, 2, 2

[10] Kremer, P., "The Gupta-Belnap Systems $S^{\#}$ and S^{*} are not Axiomatisable," *Notre Dame Journal of Formal Logic*, vol. 34 (1993), pp. 583–596. Zbl 0795.03032 MR 94m:03046 3, 3

[11] Kripke, S., "Outline of a Theory of Truth," *Journal of Philosophy*, vol. 72 (1975), pp. 690–716. Zbl 0952.03513 1, 2, 2

[12] Mackie, J., *Truth, Probability, and Paradox*, Oxford University Press, London, 1973. Zbl 0301.02003 MR 57:9463 2

[13] Martin, R., and P. Woodruff, "On representing 'true-in-L' in L," *Philosophia*, vol. 5 (1975), pp. 217–221. Zbl 0386.03001 2

[14] McGee, V., *Truth, Vagueness and Paradox: an Essay on the Logic of Truth*, Hackett, Indianapolis, 1991. Zbl 0734.03001 MR 92k:03004 2, 3, 4

[15] Montague, R., "Syntactical treatments of modality, with corollaries on reflexion principles and finite axiomatizability," *Acta Philosophica Fennica*, vol. 16 (1963), pp. 153–167. Zbl 0117.01302 MR 29:1140 3

[16] Parsons, C., "The Liar Paradox," *Journal of Philosophical Logic*, vol. 3 (1974), pp. 381–412. Zbl 0296.02001 MR 58:21402 1, 2

[17] Russell, B., "Mathematical logic as based on the theory of types," *American Journal of Mathematics*, vol. 30 (1908), pp. 222–262. 2

[18] Simmons, K., "The diagonal argument and the Liar," *Journal of Philosophical Logic*, vol. 19 (1990), pp. 277–303. Zbl 0702.03001 MR 91k:03013 2, 4

[19] Sobel, J., "Lies, lies and more lies: A plea for propositions," *Philosophical Studies*, vol. 67 (1992), pp. 51–69. MR 93i:03007 2

Robert C. Koons
Department of Philosophy
University of Texas at Austin
Waggener Hall 316
Austin, TX 78712-1180
email: koons@la.utexas.edu