VANN McGEE

# HOW TRUTHLIKE CAN A PREDICATE BE? A NEGATIVE RESULT[1]

Tarski [10, Section 7] showed that, if our theory of truth Γ includes among its consequences all instances of the schema

(T) $\qquad T\ulcorner\varphi\urcorner \leftrightarrow \varphi$

(where $\ulcorner\varphi\urcorner$ is the Gödel number of $\varphi$ under some reasonable coding), and if, moreover, Γ entails certain very basic truths of arithmetic, then Γ must be inconsistent. The result has been strengthened, notably by Montague [8], who showed that, if we require Γ to be closed under first-order consequence and we require it to contain $T\ulcorner\varphi\urcorner$ whenever it contains $\varphi$, then the left-to-right direction of schema (T) will be enough to give a contradiction. The present note gives a further result along the same lines, showing that, if we require Γ to be closed under certain further natural logical operations, we can get an ω-inconsistency without assuming either direction of schema (T).

Let $\mathscr{L}$ be a countable first-order language which includes the language of arithmetic; there will be a predicate "$N$" whose intended extension is the set of natural numbers. Let $\mathscr{L}^+$ be the language obtained from $\mathscr{L}$ by adjoining the new predicate "$T$".

THEOREM. Let Γ be a set of sentences of $\mathscr{L}^+$ which:

(1) contains axioms for the theory Q obtained from Robinson's arithmetic by relativizing the quantifiers to "N" (see [11]);

(2) is closed under first-order consequence;

(3) contains $T\ulcorner\varphi\urcorner$ whenever it contains $\varphi$; and

(4) contains all instances of the following schemata:

(a) $\qquad T\ulcorner\varphi \rightarrow \psi\urcorner \rightarrow (T\ulcorner\varphi\urcorner \rightarrow T\ulcorner\psi\urcorner)$;

(γ) $\qquad T\ulcorner\neg\varphi\urcorner \rightarrow \neg T\ulcorner\varphi\urcorner$;

(c) $\qquad (\forall x)(Nx \rightarrow T\ulcorner\varphi(\dot{x})\urcorner) \rightarrow T\ulcorner(\forall x)(Nx \rightarrow \varphi(x))\urcorner$.[2]

Then Γ is ω-inconsistent.

*Journal of Philosophical Logic* 14 (1985) 399–410. 0022–3611/85.10.

*Proof.* Using the usual diagonal construction (all such constructions are modeled after Gödel [3]), find a ternary predicate $F$ of the language of arithmetic such that

$$\text{(i)} \qquad \vdash_Q (\forall y)(\forall z)(F(0,y,z) \leftrightarrow y = z);$$

$$\text{(ii)} \qquad \vdash_Q (\forall x)(\forall y)(\forall z)(Nx \to (F(s(x),y,z) \leftrightarrow y = \ulcorner(\forall y)(F(\dot{x},y,\dot{z}) \to Ty)\urcorner)).$$

($s$ is the successor function). Now find a sentence $\sigma$ of $\mathscr{L}^+$ such that

$$\text{(iii)} \qquad \vdash_Q \sigma \leftrightarrow \neg(\forall x)(Nx \to (\forall y)(F(x,y,\ulcorner\sigma\urcorner) \to Ty)).$$

$\sigma$ says that not every result of prefixing $T$'s to $\sigma$ is true. The following sentences are in $\Gamma$:

$$\text{(iv)} \qquad \neg\sigma \to (\forall x)(Nx \to (\forall y)(F(x,y,\ulcorner\sigma\urcorner) \to Ty)) \qquad \text{(from (iii))}$$

$$\text{(v)} \qquad \neg\sigma \to (\forall y)(F(0,y,\ulcorner\sigma\urcorner) \to Ty) \qquad \text{(from (iv))}$$

$$\text{(vi)} \qquad (\forall y)(F(0,y,\ulcorner\sigma\urcorner) \leftrightarrow y = \ulcorner\sigma\urcorner) \qquad \text{(from (i))}$$

$$\text{(vii)} \qquad \neg\sigma \to T\ulcorner\sigma\urcorner \qquad \text{(from (v) and (vi))}$$

$$\text{(viii)} \qquad \sigma \to \neg(\forall x)(Nx \to (\forall y)(F(x,y,\ulcorner\sigma\urcorner) \to Ty)) \qquad \text{(from (iii))}$$

$$\text{(ix)} \qquad T\ulcorner\sigma \to \neg(\forall x)(Nx \to (\forall y)(F(x,y,\overline{\ulcorner\sigma\urcorner}) \to Ty))\urcorner \qquad \text{(from (viii) by (3))}$$

$$\text{(x)} \qquad T\ulcorner\sigma\urcorner \to T\ulcorner\neg(\forall x)(Nx \to (\forall y)(F(x,y,\overline{\ulcorner\sigma\urcorner}) \to Ty))\urcorner \qquad \text{(from (ix) by (4a))}$$

$$\text{(xi)} \qquad \neg\sigma \to T\ulcorner\neg(\forall x)(Nx \to (\forall y)(F(x,y,\overline{\ulcorner\sigma\urcorner}) \to Ty))\urcorner \qquad \text{(from (vii) and (x))}$$

$$\text{(xii)} \qquad \neg\sigma \to \neg\, T\ulcorner(\forall x)(Nx \to (\forall y)(F(x,y,\overline{\ulcorner\sigma\urcorner}) \to Ty)) \qquad \text{(from (xi) by (4b))}$$

(xiii) $\neg\sigma \rightarrow \neg(\forall x)(Nx \rightarrow T\ulcorner(\forall y)(F(\dot{x}, y, \overline{\ulcorner\sigma\urcorner}) \rightarrow Ty)\urcorner)$

(from (xii) by (4c))

(xiv) $(\forall x)(Nx \rightarrow (\forall y)(F(s(x), y, \ulcorner\sigma\urcorner) \leftrightarrow y = \ulcorner(\forall y)(F(\dot{x}, y, \overline{\ulcorner\sigma\urcorner}) \rightarrow Ty)\urcorner))$

(from (ii))

(xv) $\neg\sigma \rightarrow \neg(\forall x)(Nx \rightarrow (\forall y)(F(s(x), y, \ulcorner\sigma\urcorner) \rightarrow Ty))$

(from (xiii) and (xiv))

(xvi) $\neg\sigma \rightarrow \neg(\forall x)(Nx \rightarrow (\forall y)(F(x, y, \ulcorner\sigma\urcorner) \rightarrow Ty))$

(from (xv))

(xvii) $\neg\sigma \rightarrow \sigma$ (from (iii) and (xvi))

(xviii) $\sigma$ (from (xvii))

(xix) $T\ulcorner\sigma\urcorner$ (from (xviii) by (3))

(xx) $(\forall y)(F(0, y, \ulcorner\sigma\urcorner) \leftrightarrow y = \ulcorner\sigma\urcorner)$ (from (i))

(xxi) $(\forall y)(F(0, y, \ulcorner\sigma\urcorner) \rightarrow Ty)$ (from (xix) and (xx))

We intend to show that, for each $n$, the sentence "$(\forall y)(F(n, y, \ulcorner\sigma\urcorner) \rightarrow Ty)$" is in $\Gamma$. We have just shown this for $n = 0$. Now suppose that

(xxii) $(\forall y)(F(k, y, \ulcorner\sigma\urcorner) \rightarrow Ty)$

is in $\Gamma$. Then so are these sentences:

(xxiii) $T\ulcorner(\forall y)(F(\bar{k}, y, \overline{\ulcorner\sigma\urcorner}) \rightarrow Ty)\urcorner$ (from (xxii) by (3))

(xxiv) $(\forall y)(F(k+1, y, \ulcorner\sigma\urcorner) \leftrightarrow y = \ulcorner(\forall y)(F(\bar{k}, y, \overline{\ulcorner\sigma\urcorner}) \rightarrow Ty)\urcorner)$

(from (ii))

(xxv) $(\forall y)(F(k+1, y, \ulcorner\sigma\urcorner) \rightarrow Ty)$ (from (xxiii) and (xxiv))

It follows by mathematical induction that, for each $n$, the sentence

(xxvi) $(\forall y)(F(n, y, \ulcorner\sigma\urcorner) \rightarrow Ty)$

is in $\Gamma$. But also

(xxvii) $\neg(\forall x)(Nx \rightarrow (\forall y)(F(x, y, \ulcorner\sigma\urcorner) \rightarrow Ty))$

(from (iii) and (xviii))

is in $\Gamma$. So $\Gamma$ is $\omega$-inconsistent.

This theorem is a variant of Montague's result [8] that we get an outright inconsistency if we replace condition (4) by the requirement that all instances of the schema

$$T\ulcorner\varphi\urcorner \rightarrow \varphi$$

be in $\Gamma$. The present result is interesting, I think, because its hypotheses do not include either direction of schema (T) and they do not include the principle of bivalence. (The converse of (4b) requires that each sentence be either true or false; (4b) only requires that no sentence be both true and false.)

Montague was particularly interested in cases where the theory $\Gamma$ was not a theory of truth, but a theory of provability. If we take "$T$" to be an abbreviation for the natural arithmetical predicate expressing provability in Peano arithmetic and we take $\Gamma$ to be the set of theorems of Peano arithmetic, $\Gamma$ will satisfy conditions (1), (2), (3), and (4a). Condition (4b) will fail, by Gödel's second incompleteness theorem [3], and condition (4c) will fail, by the proof of Gödel's first incompleteness theorem.

Various recent authors have come up with theories of truth which attempt to circumvent the difficulty raised by Tarski, although in admittedly imperfect ways. The consequences of these various theories do not satisfy conditions (1) through (4c), as we can see by noting that the theories have models with standard integers. Yet (1) through (4c) are structural conditions we might naturally have expected the set of consequences of a theory of truth to satisfy. Indeed they are structural conditions we might have continued to look for even after giving up hope of getting one or both directions of schema (T). It will be interesting to see, for the most prominent of the current theories of truth, which of conditions (1) through (4c) fail. To simplify our notation, let us assume that $\mathscr{L}$ is just the language of arithmetic. Let $\mathscr{N}$ be the standard model of $\mathscr{L}$, and, for $U$ a set of natural numbers, let $(\mathscr{N}, U)$ be the model of $\mathscr{L}^+$ obtained by setting the extension of "$T$" equal to $U$.

The best known alternative theory of truth is that developed by Kripke

[7], who talks about partial interpretations of $\mathscr{L}^+$. A partial model of $\mathscr{L}^+$ is a structure $(\mathscr{N}, (U, V))$, where $U$ and $V$ are disjoint sets of sequences. $U$, the extension of "$T$", consists of those sentences that are definitely true, whereas the antiextension $V$ consists of those sentences that are definitely untrue. Truth in a partial model is specified according to Kleene's rules for strong 3-valued logic [6, Section 64]. If we take $(\mathscr{N}, (U, V))$ to be the minimal fixed point, that is, we take $(U, V)$ to be the smallest pair such that

$U$ = the set of sentences true in the partial model $(\mathscr{N}, (U, V))$, and

$V$ = the set of sentences false in the partial model $(\mathscr{N}, (U, V))$,

we find that the set of sentences true in the partial model $(\mathscr{N}, (U, V))$ meets only conditions (1) and (3). Thus for Kripke's theory the constraints imposed by our theorem aren't the least bit confining. This is, I think, unsurprising. We normally require that, when someone proposes a theory, he be willing to accept all the logical consequences of that theory. This puts some rather strong restraints on what his theory can look like; it cannot have impossible or incredible logical consequences. If semantic theory is exempted from this requirement, so that, in doing semantics, one is only bound to accept those consequences of one's theory that are prescribed by some very weak logical calculus, then one has a free hand to develop the semantic theory any way one likes. We will be unlikely to find any interesting formal constraints on what such a theory can look like.

Thus our theorem has nothing interesting to say about those theories that resort to nonclassical logics to solve the semantic antinomies. It only has bearing on those theories of truth that allow us the full range of our customary modes of inference.

Feferman [2] has shown how to develop an interesting version of Kripke's theory entirely within a classical logical context. He gives a simple set of axioms of the classical predicate calculus that are satisfied by a classical model $(\mathscr{N}, U)$ if and only if, for some $V$, the partial model $(\mathscr{N}, (U, V))$ is a fixed point of the Kripke construction. The set of consequences of Feferman's theory will satisfy conditions (1), (2), (4a), (4b), and (4c). Condition (3) will not be met, since it is possible to prove things in Feferman's theory that are not, according to Feferman's theory, true.[3]

While the construction using the 3-valued logic is the best-known version of Kripke's theory, Kripke [7, pp. 711–712] mentions alternative formulations that are, from our present point of view, more interesting. These variants use some form of van Fraassen's method of supervaluations [12] to evaluate truth in a partial model. If we stipulate

> A sentence is true (false) in the partial model $(\mathscr{N}, (U, V))$ iff it is true (false) in every classical model $(\mathscr{N}, W)$ in which $U \subseteq W \subseteq \text{Sent} - V$

(where Sent = the set of sentences of $\mathscr{L}^+$), the set of sentences true at the minimal fixed point will satisfy conditions (1), (2), and (3). If we restrict what classical models we will look at, saying

> A sentence is true (false) in the partial model $(\mathscr{N}, (U, V))$ iff it is true (false) in every classical model $(\mathscr{N}, W)$ in which $W$ is a consistent set of sentences with $U \subseteq W \subseteq \text{Sent} - V$,

the set of sentences true at the minimal fixed point will satisfy condition (4b) as well. Further restriction, saying

> A sentence is true (false) in the partial model $(\mathscr{N}, (U, V))$ iff it is true (false) in every classical model $(\mathscr{N}, (U, V))$ in which $W$ is a maximal consistent set of sentences with $U \subseteq W \subseteq \text{Sent} - V$,

will yield a set of sentences that satisfies all the conditions except (4c).

If we restrict the class of classical models we look at still further, stipulating

> A sentence is true (false) in the partial model $(\mathscr{N}, (U, V))$ iff it is true (false) in every classical model $(\mathscr{N}, W)$ in which $W$ is a maximal $\omega$-consistent set of sentences with $U \subseteq W \subseteq \text{Sent} - V$,

we find that, for any fixed point, the set of sentences true at the fixed point would have to satisfy conditions (1) through (4c). Yet the set of sentences true at a fixed point would also have to be $\omega$-consistent. Consequently, there is no fixed point. What has gone wrong? Kripke proves the existence of fixed points by an inductive construction, setting

$U_0 = V_0 =$ the empty set $\emptyset$,

$U_{\alpha+1}$ (resp., $V_{\alpha+1}$) = the set of sentences true (resp., false) in the partial model $(\mathcal{N}, (U_\alpha, V_\alpha))$, and

$U_\lambda = \cup_{\alpha<\lambda} U_\alpha$ and $V_\lambda = \cup_{\alpha<\lambda} V_\alpha$, for $\lambda$ a limit.

The minimal fixed point is the model $(\mathcal{N}, (U_\alpha, V_\alpha))$ for which we have $U_\alpha = U_{\alpha+1}$ and $V_\alpha = V_{\alpha+1}$. What happens in the present case is that $U_\omega$ is an $\omega$-inconsistent set of sentences, so that at the $\omega + 1$st stage we get the degenerate model $(\mathcal{N}$(Sent, Sent)).

We know from the general theory of inductive definitions (see, e.g., [9, p. 7]) that a monotone operator, acting on a set of objects of any sort, will have a fixed point. We now see, however, that monotonicity alone is not enough to guarantee success in the present context, since it will not rule out the possibility that, at a certain stage, a sentence will be declared both definitely true and definitely untrue.[4] We need some further condition, such as the following: Let's say that a partial model $(\mathcal{N}, (U, V))$ is *consistent* (with respect to a scheme for evaluating truth in a partial model) iff no sentence is both true and false in $(\mathcal{N}, (U, V))$. Thus with respect to the 3-valued logic and with respect to the first supervaluational scheme we looked at, a model $(\mathcal{N}, (U, V))$ is consistent iff $U \cap V = \emptyset$. With respect to the next two supervaluational schemes we examined, $(\mathcal{N}, (U, V))$ is consistent iff $U$ is first-order consistent and no member of $V$ is a first-order consequence of $U$. With respect to the last, ill-fated supervaluational scheme we looked at, $(\mathcal{N}, (U, V))$ is consistent just in case the union of $U$ with the set of negations of members of $V$ is consistent by $\omega$-logic. To insure that we get a consistent fixed point, we require:

> $(\mathcal{N}, (\emptyset, \emptyset))$ is consistent. If $(\mathcal{N}, (U, V))$ is consistent, so is $(\mathcal{N}, (\{$sentences true at $(\mathcal{N}, (U, V))\}, \{$sentences false at $(\mathcal{N}, (U, V))\}))$. The union of a chain of consistent partial models is consistent.[5]

Let us now turn to the so-called naive semantics of Hans Herzberger [5]. For $U$ a set of sentences, Herzberger defines:

$h(U, 0) = U$;

$h(U, \alpha + 1) = \{$sentences true in the classical model $(\mathcal{N}, h(U, \alpha))\}$.

$$h(U, \lambda) = \{\text{sentences eventually in the sequence } \langle h(U, \alpha) \colon \alpha < \lambda \rangle\}$$

$$= \bigcup_{\alpha < \lambda} \bigcap_{\alpha \leqslant \beta < \lambda} h(U, \beta).$$

$$h(U, \infty) = \{\text{sentences eventually in } \langle h(U, \alpha) \colon \alpha \in OR \rangle\}.$$

$\varphi$ is *everywhere stably true* iff, for every $U$, $\varphi$ is in $h(U, \infty)$.

The set of everywhere stably true sentences satisfies conditions (1), (2), (3), (4a), and (4b), but not (4c). So does the set $h(U, \infty)$, for each $U$. Gupta [4] and Belnap [1] propose systems closely similar to Herzberger's which do something different at limit stages. In their systems (4a)–(4c) all fail.

The three authors do different things at limit stages, because what happens at limit stages is somewhat arbitrary. What is crucial is what is done at successor stages, namely, application of the function taking $U$ to $h(U, 1)$, which Gupta calls the "rule of revision". Yablo (in conversation) has defined the *degree of classicality* of a set of sentences $U$ to be the supremum of the set of ordinals $n < \omega$ such that there is a set $V$ with $U = h(V, n)$; this is the number of times that one can apply the inverse of the rule of revision to $U$. Our theorem shows that every set has a finite degree of classicality. There is no infinite sequence $U_0, U_1, U_2, \ldots$ with $U_n = \{\text{sentences true in the model } (\mathcal{N}, U_{n+1})\}$, since, if there were such a sequence, $\bigcap_{n<\omega} U_n$ would satisfy conditions (1) through (4c), even though $(\mathcal{N}, U_1)$ would be a standard model of $\bigcap_{n<\omega} U_n$.

Up to now, our attention has been focused on condition (4c), which says that our theory of truth should require that a generalization be true if each of its instances are true. Both in the various versions of the Kripke semantics we looked at and in the different systems of naive semantics, condition (4c) fails, while in some of the systems all the conditions other than (4c) are met. Yet there do not appear to be any convincing *philosophical* reasons for giving up (4c). Thus it might be worth our while to look at the prospects for keeping (4c) and giving up one of the other conditions.

We have already mentioned the work of Feferman [2], who investigates (among other things) the consequences of abandoning condition (3), the analogue of the rule of necessitation in modal logic. In Feferman's system, there are sentences $\theta$ such that it is possible to prove $\theta$ and also possible to prove $\neg T\ulcorner\theta\urcorner$. The system is certainly elegant, yet one is uncomfortable in accepting it, for doing so seems to obscure the connection between truth and proof. Why do we prove things, if proving something gives us no guarantee of its truth? Why do we cherish truth, if we are willing to accept as proven statements that we admit to be untrue?

The consistency principle (4b) naturally falls under suspicion, not only because of the second incompleteness theorem, but also because of the prominent role such consistency principles play in paradoxes about knowledge and belief, such as the surprise examination paradox. It turns out that we get a rather elegant system if we give up (4b) and keep the natural rules for the connectives other than negation and for the quantifiers. Consider the following axioms:

$$\varphi \rightarrow T^k \ulcorner\varphi\urcorner \text{ for } \varphi \text{ a sentence of } \mathscr{L};$$

$$T^k \ulcorner\varphi \vee \psi\urcorner \leftrightarrow (T^k \ulcorner\varphi\urcorner \vee T^k \ulcorner\psi\urcorner);$$

$$T^k \ulcorner\varphi \,\&\, \psi\urcorner \leftrightarrow (T^k \ulcorner\varphi\urcorner \,\&\, T^k \ulcorner\psi\urcorner);$$

$$T^k \ulcorner\varphi \rightarrow \psi\urcorner \leftrightarrow (T^k \ulcorner\varphi\urcorner \rightarrow T^k \ulcorner\varphi\urcorner);$$

$$T^k \ulcorner\varphi\urcorner \vee (T^k \ulcorner\neg\varphi\urcorner \leftrightarrow \neg\, T^k \ulcorner\varphi\urcorner);$$

$$T^k \ulcorner(\forall x)\, \varphi(x)\urcorner \leftrightarrow (\forall x) T^k \ulcorner\varphi(\dot{x})\urcorner;$$

$$T^k \ulcorner(\exists x)\varphi(x)\urcorner \leftrightarrow (\exists x) T^k \ulcorner\varphi(\dot{x})\urcorner;$$

(Here $k$ varies over integers $\geqslant 1$, and $T^k \ulcorner\varphi\urcorner$ is the result of prefixing $k$ T's to the Gödel number of $\varphi$.) The set of sentences true in all standard models of this axiom system satisfies conditions (1), (2), (3), (4a), and (4c). The standard models are precisely the models $(\mathscr{N}, h(\text{Sent}, n))$, where $n$ is a natural number. They have the following pleasant property;

> A sentence $\varphi$ is true in every standard model of the axioms iff $T\ulcorner\varphi\urcorner$ is true in every such model.

Let me conclude by showing that we cannot strengthen the main theorem to say that any set of sentences that satisfies conditions (1) through (4c) is simply inconsistent; indeed we do not give a simple inconsistency even if we strengthen condition (1) to require that $\Gamma$ contain all of true arithmetic. Consider the axiom system, which I'll call $\Delta$, consisting of all sentences obtained by prefixing $T$'s and universal quantifiers to instances of the following schemata:

$$\varphi \leftrightarrow T\ulcorner\varphi\urcorner \text{ for } \varphi \text{ an atomic formula of } \mathscr{L};$$

$$T\ulcorner\varphi \vee \psi\urcorner \leftrightarrow (T\ulcorner\varphi\urcorner \vee T\ulcorner\psi\urcorner);$$

$$T\ulcorner\varphi \,\&\, \psi\urcorner \leftrightarrow (T\ulcorner\varphi\urcorner \,\&\, T\ulcorner\psi\urcorner);$$

$$T\ulcorner\varphi \rightarrow \psi\urcorner \leftrightarrow (T\ulcorner\varphi\urcorner \rightarrow T\ulcorner\psi\urcorner);$$

$$T\ulcorner\neg\varphi\urcorner \leftrightarrow \neg T\ulcorner\varphi\urcorner;$$

$$T\ulcorner(\forall x)\varphi\urcorner \leftrightarrow (\forall x)T\ulcorner\varphi\urcorner.$$

$$T\ulcorner(\exists x)\varphi\urcorner \leftrightarrow (\exists x)T\ulcorner\varphi\urcorner.$$

(Here "$T\ulcorner\varphi\urcorner$" is to be understood in such a way that all variables that occur free in $\varphi$ occur free in $T\ulcorner\varphi\urcorner$. If $\varphi$ is $\varphi(v_1, v_2, \ldots, v_n)$, $T\ulcorner\varphi\urcorner$ is $T\ulcorner\varphi(\dot{v}_1, \dot{v}_2, \ldots, \dot{v}_n)\urcorner$.) The set of first-order consequences of true arithmetic together with $\Delta$ satisfies conditions (1) through (4c). Yet $\Delta$ is first-order consistent with true arithmetic, as we can see by observing that the model $(\mathscr{N}, h(\text{Sent}, k))$ satisfies all axioms of $\Delta$ in which fewer than $k$ $T$'s have been prefixed to instances of the given schemata.

Theory $\Delta$ is motivated, first of all, by a belief that the laws of semantics ought to themselves be true and, second, by the idea that there is nothing wrong with the truth conditions for quantified and molecular sentences that we get from Tarski's analysis [10] and nothing wrong with the truth conditions for atomic sentences of $\mathscr{L}$ that we get from schema (T); problems only arise when we attempt to use schema (T) to get truth conditions for atomic sentences of the form $T\ulcorner\psi\urcorner$. At one time, $\Delta$ seemed to me to be a rather attractive theory of truth. That was before I realized it didn't have any models with standard integers.

## NOTES

[1] In preparing this note, I profited from valuable discussions with a number of persons. Let me mention Nuel Belnap, Charles Chihara, Michael Lavine, William Reinhardt, Jack Silver, Albert Visser, Peter Woodruff, and Steven Yablo. Work along the same lines by Harvey Friedman was alluded to by Kripke [7, p. 712n].

[2] $\ulcorner\varphi(\bar{n})\urcorner$ is the Gödel number of the result of substituting the numeral for n for free occurrences of "$x$" in the formula $\varphi(x)$. $\ulcorner\varphi(\dot{x})\urcorner$ is the function which, for $n$ as argument, takes $\ulcorner\varphi(\bar{n})\urcorner$ as value.

[3] Woodruff [13] has investigated the effects of dropping the requirement that the extension and the antiextension be disjoint. This gets a larger class of fixed points. We can modify Feferman's axioms to get an axiom system characterizing the fixed points of Woodruff's system simply by dropping the axiom

$$T\neg x \to \neg Tx.$$

The consequences of this weakened system will satisfy (1), (2), and (4c).

[4] If, following Woodruff [13], we regard the prosciption of truth-value gluts as artificial and unnecessary, we will be automatically guaranteed that there is a fixed point. Even so, we will need to take pains to insure that the fixed point we get is not the degenerate model $(\mathscr{N}, (\text{Sent}, \text{Sent}))$.

[5] More precisely, the last requirement is this: Given a set $\{(\mathscr{N}, (U_i, V_i)): i \in I\}$ of consistent partial models, where, for each $i$ and $j$ in $I$, either $U_i \subseteq U_j$ and $V_i \subseteq V_j$ or $U_j \subseteq U_i$ and $V_j \subseteq V_i$, the partial model $(\mathscr{N}, (\cup_{i\in I} U_i, \cup_{i \in I} V_i))$ will be consistent.

## REFERENCES


[1] Belnap, Neul D., Jr., 'Gupta's rule of revision theory of truth', *Journal of Philosophical Logic* 11 (1982), 103–116.

[2] Feferman, Solomon, 'Toward useful type-free theories. I', *Journal of Symbolic Logic* 49 (1984), 75–111.

[3] Gödel, Kurt, 'Über formal unentscheidbare Sätze der *Principia mathematica* und verwandter Systeme I', *Monatshefte für Mathematik und Physik* 38 (1931) 173–198. English translation by Stefan Bauer-Mengelberg in Jean van Heienoort, ed., *From Frege to Godel*, pp. 596–616. Harvard University Press, Cambridge, Mass.: 1967.

[4] Gupta, Anil, 'Truth and paradox', *Journal of Philosophical Logic* 11 (1982) 1–60.

[5] Herzberger, Hans G., 'Notes on naive semantics', *Journal of Philosophical Logic* 11 (1982), 61–102.

[6] Kleene, Stephen Cole, *Introduction to Metamathematics*, American Elsevier, New York, 1952.

[7] Kripke, Saul A., 'Outline of a theory of truth', *Journal of Philosophy* 72 (1975), 690–716.

[8] Montague, Richard, 'Syntactic treatments of modality, with corollaries on reflexion principles and finite axiomatizability', *Acta Philosophica Fennica* 16 (1966), 154–167. Reprinted in Montague, *Formal Philosophy*, pp. 286–302. New Haven: Yale University Press, 1974.

[9] Moschovakis, Yiannis N., *Elementary Induction on Abstract Structures*, American Elsevier, New York, 1974.

[10] Tarski, Alfred, 'Der Wahrheitsbegriff in den formalisierten Sprachen', *Studia Philosophica* 1 (1935) 261–405. English translation by Joseph H. Woodger, in Tarski, *Logic, Semantics, Metamathematics*, pp. 152–278, Oxford University Press, Oxford 1956.

[11] Tarski, Alfred, Mostowski Andrzej, and Robinson, Raphael M., *Undecidable Theories*, North-Holland, Amsterdam, 1953.

[12] Van Fraassen, Bas C., 'Singular terms, truth-value gaps, and free logic', *Journal of Philosophy* **63** (1966), 481–495.
[13] Woodruff, Peter W., 'Paradox, truth and logic, Part I', *Journal of Philosophical Logic* **13** (1984), 213–232.

*Group in Logic and Methodology of Science,*
*University of California,*
*Berkeley, CA 49720,*
*U.S.A.*