STEPHEN YABLO

TRUTH AND REFLECTION

> My reasoning wants to be faithful to the evidence that aroused it. That evidence is the absurd.
>
> Camus, *The Myth of Sisyphus*, "An Absurd Reasoning"

I. INTRODUCTION

Out of their anxiety about semantical paradox philosophers have devised a variety of formal theories of truth. What are these theories supposed to accomplish? The quick answer is that they're supposed to solve, or help solve, the problem the paradoxes raise for truth. But what problem is that? In recent years it has become apparent that the paradoxes raise a *number* of problems for truth. These need distinguishing before they can be effectively tackled.

Darkest and deepest of all is what Charles Chihara has called the "diagnostic problem":

> Alfred Tarski once remarked: "The appearance of an antimony is for me a symptom of disease." But what disease? (SP, p. 590).

More neutrally put, the paradoxes seem to show that there is something somehow "wrong" with the way we evaluate sentences. What could that something be? If Tarski's way of seeing the matter is correct, the condition the paradoxes betoken is, like a disease, unnatural, exceptional, and on balance harmful. But other perspectives are possible, and indeed it cannot be ruled out that like blind spots, the paradoxes fall naturally out of the very principles responsible for the patient's health.[2] Decide this further issue how we will, the basic problem is simply this, to say what in our semantical procedures makes the paradoxes happen.

According to Professor Chihara, the diagnostic problem has got to be kept separate from the familiar "preventative problem"

Journal of Philosophical Logic **14** (1985) 297–349. 0022–3611/85.10.

of devising languages or logical systems which capture certain essential or useful features of the relevant semantical concepts, but within which the paradoxes cannot arise (SP, pp. 590–1).

Not everyone finds the distinction compelling; in particular, a certain philosophical cast of mind is strongly disposed to construe favoured preventative measures as implying diagnoses. Thus it is sometimes said that the problem with natural languages is that they contain their own metalanguages (a thing which Tarski has shown to lead to all sorts of trouble). This is rather like blaming headaches on an insufficient supply of aspirin in the brain.[3] The point is that there is a difference between the presence of that which produces the symptoms and the absence of that which would prevent them. To prevent the paradoxes is not necessarily to have divined their source.

But suppose we *have* divined their source; what then? This is Professor Chihara's "treatment problem":

Should natural language be altered in order to remove the causes of the paradoxes? In particular, should a new concept of truth be constructed to replace the present one? And if so, how? (SP, p. 616).

In a similar vein, Anil Gupta speaks of the "normative problem"

of discovering the changes (if any) that the paradox dictates in our conception and use of "true" (TP, p. 2).

Deliberations about treatment must naturally be informed by the very latest in the technology of prevention, but the problems are fundamentally distinct, the treatment problem requiring us to *balance* the attractions of proposed preventative measures against the ineffable comforts of fidelity to the semantical habits into which we were reared.

People have problems with the word "true", but on the whole it must be said that we do rather well with it. By and large, parties in agreement on substantive matters agree also about which sentences are true and which false. To be sure, there are sentences to which we have difficulty assigning any semantical status whatsoever, but even here there is considerable agreement as to which these are. The moral is that our inability to deal satisfactorily with certain troublesome cases should not draw us into overly pessimistic conclusions about the integrity of our semantical procedures. If the paradoxes show that something is wrong with these procedures, our general success in dealing with truth shows that there must be something

very right about them too. Certainly there is no ground for claiming that our procedures are *vitiated* by the paradoxes, unless one is prepared to accept that we find their vitiation very little handicap in practice.

Two new problems now suggest themselves. The first is the broadly empirical one of discovering how people actually come up with their semantical assessments. What rules are consciously followed? What rules are unconsciously followed? What are the principles of operation of such subpersonal systems as may be involved? The second is the philosophical problem of devising systematic semantical procedures which yield the (intuitively) correct results, and in a manner as instructive as possible about the nature of truth. After all, we are ourselves somehow able to come up with the intuitively "right" evaluations; so there must be a way, in a suitably broad sense of "way", in which we do it; so there must be a way in which *to* do it; so let's try to figure out such a way, and while we're at it let's try to make it as instructive about truth as we can.[4]

To understand how the last-mentioned problems differ, consider an analogy. Turing, Gödel, and others have provided extremely neat and instructive characterizations of the computable functions. Yet no one actually decides whether proffered functions are computable by, e.g., trying to dream up Turing machines which compute them! Conversely, the true story of how these decisions are actually reached need not shed much light on what *makes* functions computable. In the same way, the most illuminating semantical procedures are probably not the ones that people actually employ, and even the best account of our actual procedures need not be particularly revealing about the nature of truth. The empirical problem of ascertaining our practical procedures of evaluation has only so much to do with the philosophical problem of devising instructive procedures which yield the correct results.

Call these problems the "psychological" and the "descriptive".[5] Failure to keep them separate can lead to a lot of needless and debilitating psychologistic hand-wringing. Solutions to the psychological problem are obviously limited by what is psychologically possible; if a proposed solution would require us to do things which we are incapable of doing, then it must be wrong. But solutions to the descriptive problem are not similarly constrained, and where the latter is concerned it is no objection to a system of evaluation that the procedures it invokes are, for one reason or another, incapable of psychological realization.

To return to the question that started us off, what are formal theories of truth supposed to accomplish? Relevance though they may have for the problems of diagnosis, prevention, treatment, or possibly even psychology, such theories are first and foremost attempts to solve the *descriptive* problem of truth. If one of them is to succeed, it will be by doing for truth what the theories of Turing and Gödel did for computability.[6]

II. INTUITIVE BACKGROUND

Three basic intuitions underlie the present proposal: truth is strong, truth is grounded, and the paradoxes are genuine. Much of what follows is essentially a sustained attempt to hammer these intuitions home, but they should be briefly explained at the outset.

The grounding intuition is best reached in stages. To say that truth is *supported* is to say that nothing is true unless something *makes* it true. Nonsemantical atomic sentences can only be made true by the nonsemantical circumstances; semantical atomic sentences like "ϕ is true" can only be made true by the truth of ϕ; negations are made true by the falsity of the sentence negated, disjunctions by the truth of one of their disjuncts, and universal generalizations by the truth of each of their instances.

This is all right as far as it goes, but that isn't as far as we might like. For consider the sentence K = "K is true". Evidently on the *assumption* that K is true, it functions to *make* itself true. But surely K is *not*, in and of itself, true, and this shows that supportedness isn't enough: it leaves the way open for sentences to be made true by their own truth, or, more generally, for numbers of sentences, each lacking independent means of support, to pass truth around in circles.

A natural reaction to the foregoing is to insist that what *makes* a sentence true must somehow obtain *prior* to its so doing; to insist, that is, that truth is *forced*. Thus the truth of "ϕ is true" really requires the *prior* truth of ϕ, that of $-\phi$ the *prior* falsity of ϕ, and so on. This closes the door on the unintuitive possibilities noted above, but others remain. Imagine an infinite sequence of sentences, each of which describes its successor as true. Could all of them be true? On the *assumption* that they are, each inherits its truth from the one following, as supportedness requires. And evidently there is nothing to prevent us from seeing all their truth-values as having been passed backward "from infinity", so the requirement of forcing seems

to be met too. From an intuitive standpoint, though, something is very wrong. If the chain of priority goes back forever, how did any of the sentences ever *get* to be true in the first place? To say that truth is *grounded* is to say that every chain of priority must terminate, eventually, in the nonsemantical circumstances.

Summing up, sentences are only true if something, typically the truth or falsity of other sentences, makes them true; whatever makes a sentence true must obtain prior to its so doing; and the chain of priority cannot go back forever. All of this might seem too obvious to mention, but as we shall see it has important, and in some instances even controversial, consequences.

To call a true sentence true is to say something true, and to call an untrue sentence true is to say something false; this is what is meant by the assertion that truth is *strong*. On a competing conception of truth, which may be dubbed the *weak* conception, the statement that ϕ is true simply inherits ϕ's truth-status, whatever it may be. Thus if ϕ is neither true nor false, then to call it true is, on the weak conception, to say something neither true nor false; and if ϕ is for some reason both true and false, then to call it true is to say something itself both true and false. How will the strong conception deal with these cases? If ϕ is neither true nor false, then it is at any rate not true, so to call it true is to say something uniquely false (rather than valueless). If ϕ is both true and false, then it is at any rate true, so to call it true is to say something uniquely true (rather than both true and false).[7]

The argument for the strong conception is straightforward. We intend that an English sentence of the form "α is P" should be true if and only if the object denoted by "α" has the property expressed by "P", and false if and only if the object denoted by "α" lacks that property. If "P" expresses the property of being blue, for example, then we intend "α is P" to be true if and only if α is blue, and false if and only if it is not blue. Applying the same principle to the case where "P" expresses the property of being true, we get the result that "α is P" should be true just in case α is true, and false just in case it is not true. And this is precisely the strong conception of truth.

On the weak conception of truth, the Liar sentence "I am not true" may safely be regarded as either (i) neither true nor false, or (ii) both true and false.[8] On the strong conception, though, the former assumption leads quickly to the conclusion that it is uniquely true, and the latter to the

conclusion that it is uniquely false. (If it is neither true nor false, then to say it is true is to say something uniquely false, so to say it is not true is to say something uniquely true; but that is exactly what is says. If it is true and false, then to say it is true is to say something uniquely true, so to say that it is not true is to say something uniquely false; again, that is exactly what it says.) Thus on the strong conception of truth, truth-value gaps (gluts) do not function as safe houses for paradoxes, and in fact, that conception makes *any* assumption about the truth-status of the Liar self-defeating.

The rules outlined below seem to yield most of the intuitively correct results. But the attempt to apply them to certain semantical anomalies leads inevitably to contradictions. In my view, this is itself an intuitively correct result, and constitutes one of the best arguments in favour of the rules. But it also suggests that the rules are in some sense inconsistent. If this is right, it supports something like Professor Chihara's solution to the diagnostic problem: semantical rules which accurately reflect our intentions about the use of "true", and which consequently strike us as *obviously* correct, are nevertheless inconsistent in an unanticipated way.[8] The treatment problem is also shown in a new light. That the same rules which cause so much trouble in certain of their applications yield unequivocal and intuitive results in all the others seems to show that the paradoxes can exist in, and indeed issue out of, a substantially healthy nature. This creates strong pressure for leaving things as they are, on pain of undermining the very principles which guide the enterprise. The treatment of choice might well be benign neglect.

III. INDUCTIVE AND ANTIINDUCTIVE SPACES

The present theory takes Kripke's 1975 Theory of Truth as its starting point. On Kripke's approach (see the next section for details), sentences derive their truth-values from the truth-values of other sentences, so that every expansion of the class of sentences already assigned truth-values means an expansion of the class of sentences to which truth-values are due. This leads naturally to the study of *monotonic* operators – those obeying the rule: the bigger the input, the bigger the output – and the *inductive spaces* which they inhabit.

The approach developed here takes a somewhat more liberal view of the

inheritability of truth-value, allowing sentences to derive their truth-values not only from the truth-values which other sentences *have*, but also from those which they *lack*. Thus not only can "ϕ is true" inherit truth from the truth of ϕ, it can also inherit falsity from ϕ's *un*truth. This departure from Kripke leads us into the slightly more involved study of *antimonotonic* operators – those obeying the rule: the *smaller* the input, the bigger the output – and their associated *antiinductive spaces*.

Inductive and antiinductive spaces are abstract set-theoretic objects whose relevance to semantics may not be immediately apparent. The reader is advised to tour quickly through the definitions and main results now, and to look back as required when considering their semantical applications later on. It is only on account of their applications that inductive and antiinductive spaces are mentioned at all, but to appreciate those applications one needs a preliminary feeling for the spaces themselves.

Let U be a set, called the **universe**, and let J be a **monotonic operator** on th power set of U, i.e., $J: P(U) \to P(U)$ and $A \subseteq B \Rightarrow J(A) \subseteq J(B)$. Then J is called a **jump operator** on U, and the ordered pair $\langle U, J\rangle$ is an **inductive space**. Given a subset S of U, the sequence $\langle J^\alpha(S) \mid \alpha \in OR\rangle$ is defined thus:

$$(1) \qquad J^0(S) = S;$$

$$(2) \qquad \forall \alpha > 0, J^\alpha(S) = J(J^{\alpha-1}(S)),$$

where $J^{\alpha-1}(S)$ is understood to be $\cup_{\beta<\alpha} J(S)$ when α is a limit ordinal. A subset S of U is **sound** with respect to J if $S \subseteq J(S)$. Since J is monotonic, successive applications of J to a sound initial set may be relied on to preserve the containment; and this observation is easily converted into a proof of the following proposition.

PROPOSITION 1. If S is sound, then $\langle J^\alpha(S) \mid \alpha \in OR\rangle$ is an increasing sequence.[10]

Proof. Given our reading of $J^{\alpha-1}(S)$ for limit ordinals α, it suffices to show that $\forall \alpha > 0, J^{\alpha-1}(S) \subseteq J^\alpha(S)$. The proof is an easy induction on α. □

The **closure** S^* of S is defined to be $\cup_\alpha J^\alpha(S)$. Do the $J^\alpha(S)$'s approach their limit asymptotically, or do they finally attain it? The next proposition provides an answer.

PROPOSITION 2. Let S be sound. Then there is a β such that $\forall\gamma \geqslant \beta$ $J^\gamma(S) = S^*$.

Proof. First we show that there is a β such that $J^\beta(S)$ is a **fixed point** of J, i.e., a set X such that $X = J(X)$. Let κ be the cardinality of U, and let κ^+ be the least cardinal greater than κ. If $\langle J^\alpha(S) \mid \alpha \in OR\rangle$ were strictly increasing, then $J^{(\kappa^+)}(S)$ would have at least κ^+ members, impossible for a subset of U. So there must be a β such that $J^{\beta+1}(S) = J^\beta(S)$. It is routine to check that $\forall\gamma > \beta, J^\gamma(S) = J^\beta(S)$. Since $\forall\alpha < \beta, J^\alpha(S) \subseteq J^\beta(S)$ by Proposition 1, $J^\beta(S) = \cup_\alpha J^\alpha(S) = S^*$. □

Clearly S^* is the smallest fixed point extending S, and because of this it is sometimes referred to as the **fixed point generated by S.**[11]

We turn now to antiinductive spaces. Let U be as before, and let L be an **antimonotonic operator** on the power set of U, i.e., $\mathsf{L}: \mathsf{P}(U) \to \mathsf{P}(U)$ and $A \subseteq B \Rightarrow \mathsf{L}(B) \subseteq \mathsf{L}(A)$. Then L is called a **leap operator** on U, and $\langle U, \mathsf{L}\rangle$ is an **antiinductive space.** Given a subset K of U the sequence $\langle \mathsf{L}^\alpha(K) \mid \alpha \in OR\rangle$ is defined as follows:

(1) $\quad \mathsf{L}^0(K) = K;$

(2) $\quad \forall\alpha > 0, \mathsf{L}^\alpha(K) = \mathsf{L}(\mathsf{L}^{\alpha-1}(K)),$

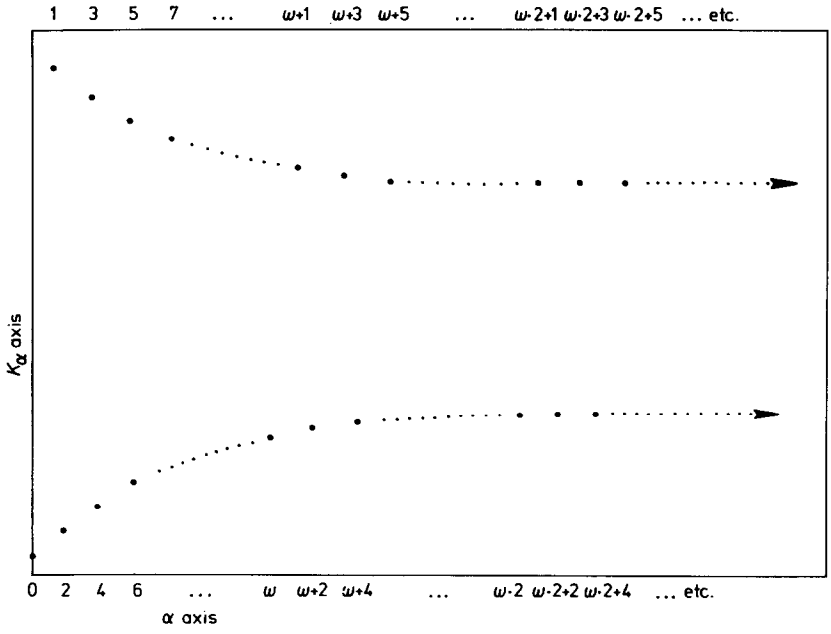
where $\mathsf{L}^{\alpha-1}(K)$ is understood to be $\text{limsup}_{\beta<\alpha}\mathsf{L}^\beta(K)$ when α is a limit ordinal.[12] As before, we want to find conditions on K which will ensure a modicum of good behaviour on the part of the sequence it generates. That $\langle \mathsf{L}^\alpha(K) \mid \alpha \in OR\rangle$ should be increasing (except trivially) is too much to hope for, but we can arrange for something almost as nice: its decomposability into upper and lower subsequences which gradually grow towards each other.

Let $\mathcal{J}$ be either OR or an initial segment of OR. An element K_α of the sequence $\langle K_\alpha \mid \alpha \in \mathcal{J}\rangle$ is *inferior* (*superior*) therein if and only if it is a subset (superset) of every subsequent K_α. The sequence $\langle K_\alpha \mid \alpha \in \mathcal{J}\rangle$ **telescopes** if and only if:

(a) $\quad \forall\beta \in \mathcal{J}, K_\beta$ is either inferior or superior in $\langle K_\alpha \mid \alpha \in \mathcal{J}\rangle$, and

(b) $\quad \forall\beta, \beta + 1 \in \mathcal{J}$ [K_β superior (inferior) in $\langle K_\alpha \mid \alpha \in \mathcal{J}\rangle$ $\Rightarrow K_{\beta+1}$ inferior (superior) in $\langle K_\alpha \mid \alpha \in \mathcal{J}\rangle$].

Note that $\langle K_\alpha \mid \alpha \in \mathcal{J}\rangle$ telescopes if and only if consecutive K_α's always



"frame" the remainder of the sequence, i.e., iff $\forall\alpha\,\forall\beta > \alpha, K_\beta$ lies between K_α and $K_{\alpha+1}$. A typical telescoping sequence is shown in Figure 1.

According to Proposition 1, a sound starting set S generates an increasing jump-sequence. What should K be like to generate a telescoping leap-sequence? The answer is that K should be **supersound,** in the sense of being a subset not only of $L(K)$ but also of $L(L(K))$.

PROPOSITION 3. If K is supersound, then $\langle L^\alpha(K) \mid \alpha \in OR\rangle$ telescopes.

Proof. Since L is antimonotonic, L^2 is monotonic. From this it is not hard to prove that

(A) $\quad X$ supersound $\Rightarrow \langle L^n(X) \mid n \in \omega\rangle$ telescopes.

To show that

(B) $\quad \langle L^\alpha(X) \mid \alpha < \lambda\rangle$ telescopes $\Rightarrow \langle L^\alpha(X) \mid \alpha \leqslant \lambda + 2\rangle$ telescopes,

use the fact that if the antecedent holds, $\limsup_{\alpha<\lambda} L^\alpha(X)$ is the intersection of all $L^\alpha(X)$'s superior in $\langle L^\alpha(X) \mid \alpha < \lambda\rangle$. Then use (A) and (B) to prove by induction on α that

(C) $\quad \forall\alpha\,\langle L^\beta(X) \mid \beta < \omega \cdot \alpha\rangle$ telescopes.

The proposition follows. □

If $\langle L^\alpha(K) \mid \alpha \in OR\rangle$ telescopes, then the $L^\alpha(K)$'s alternate between the inferior and the superior. That in itself doesn't say anything about the status of particular $L^\alpha(K)$'s, but there is a simple way of telling the two kinds apart. Each ordinal α is uniquely representable as the sum of λ_α – either 0 or a limit ordinal – and n_α – a natural number. An ordinal α is called **even** if n_α is even, and **odd** if n_α is odd. For example, ω, $\omega^2 + 26$, and ω_ω are even, whereas 7 and $\omega_1 + 19$ are odd. It emerges from the proof of Proposition 3, when done out in full, that $L^\alpha(K)$ is inferior iff α is even, and (consequently) superior iff α is odd.

Let $S^\alpha = J^\alpha(S)$, and let $K_\beta = L^\beta(K)$. The sequence $\langle S^\alpha \mid \alpha \in OR\rangle$ generated by a sound set S under the operation of J is increasing, and converges to a single limit, namely the union of its members. The sequence $\langle K_\beta \mid \beta \in OR\rangle$ generated by a supersound set K under the operation of L is telescoping, and leads instead to a pair of limits: the union of its inferior entries, and the intersection of its superior entries. Let B be $\{\beta \mid K_\beta$ is superior$\} = \{\beta \mid \beta$ is odd$\}$, and let I be $\{\beta \mid K_\beta$ is inferior$\} = \{\beta \mid \beta$ is even$\}$. Let the variables σ and ι range over Σ and I. Then K's **lower closure** $\underline{K}$ can be defined as $\cup_\iota K_\iota$, and its **upper closure** $\overline{K}$ as $\cap_\sigma K_\sigma$. The next proposition shows that like the S^α's, the K_β's eventually attain their limit(s).

PROPOSITION 4. If K is supersound, then $\exists\beta[\forall\iota \geqslant \beta K_\iota = \underline{K} \;\&\; \forall\sigma \geqslant \beta K_\sigma = \overline{K}]$.

Proof. K is supersound $\Rightarrow \langle K_\iota \mid \iota \in \mathrm{I}\rangle$ is increasing $\Rightarrow \exists\beta \in \mathrm{I}\ K_\beta = \underline{K}$ (by the argument of Proposition 2) $\Rightarrow \forall\iota \geqslant \beta K_\iota = \underline{K} \Rightarrow \forall\sigma \geqslant \beta K_\sigma = L(\underline{K}) \Rightarrow \forall\sigma \geqslant \beta K_\sigma = \overline{K}$. □

If K is supersound, then the sequence of even-numbered K_β's converges to $\underline{K}$, and the sequence of odd-numbered K_β's converges to $\overline{K}$. But the sequence $\langle K_\beta \mid \beta \in OR\rangle$ of *all* K_β's need not converge, or, equivalently, $\underline{K}$ need not equal $\overline{K}$. The members of $\overline{K}$–$\underline{K}$ can be thought of as the things which $\langle K_\beta \mid \beta \in OR\rangle$ is unable to decide about, which suggests that we call them the ***K*-undecidables.** This interpretation of $\overline{K}$–$\underline{K}$ is supported by the next proposition.

PROPOSITION 5. $(\forall x)$ (x is K-undecidable $\Leftrightarrow \forall\iota\, x \notin K_\iota \;\&\; \forall\sigma\, x \in K_\sigma$).

Proof. $x \in \overline{K}$–$\underline{K} \Leftrightarrow x \in \cap_\sigma K_\sigma - \cup_\iota K_\iota \Leftrightarrow x \in \cap_\sigma K_\sigma \;\&\; x \notin \cup_\iota K_\iota \Leftrightarrow \forall\sigma\, x \in K_\sigma \;\&\; \forall\iota\, x \notin K_\iota$. □

In other words, as β increases the K-undecidables revolve, with periodicity two, in and out of K_β.

To sum up the essential facts about inductive and antiinductive spaces: if S is sound, then $\langle S^\alpha \mid \alpha \in OR\rangle$ increases to the constant value S^*; if K is supersound, then $\langle K_\beta \mid \beta \in OR\rangle$ telescopes its way into a neverending oscillation between $\underline{K}$ and $\overline{K}$. For the semantical import of inductive and antiinductive spaces see Sections IV and IX; Section V attempts to motivate the move from inductive to antiinductive methods.[13]

IV. FIXED POINT SEMANTICS[14]

Let L_T be an ordinary first-order language with distinguished predicate T, for truth, and enough individual constants to name all its own sentences. An ordered pair $M = \langle D, I\rangle$ is a **(general) ground model** of L_T if D is a set containing (among other things) all of L_T's sentences, and I is a function with the following properties. First, I's domain is the set of names and predicates of L_T. Second, if c is a name of L_T, then $I(c) \in D$. Third, for all x in D, there is a name c of L_T such that $I(c) = x$ (this is just for convenience, so that we can interpret the quantifiers substitutionally). Finally, if P is an n-place predicate of L_T (other than T), then $I(P) = \langle I^t(P), I^f(P)\rangle$, where $I^t(P)$ and $I^f(P)$ – P's **extension** and **antiextension** – are *arbitrary* subsets of D^n. Note that I assigns nothing to the truth-predicate T.[15]

Fix a ground model M of L_T. M will be our formal stand-in for "the way the world is", semantical facts aside. A **(general) valuation** of L_T is an arbitrary subset of $\{\langle \phi, v\rangle \mid \phi \in \mathrm{Sent}(L_T) \,\&\, v = t \text{ or } f\}$ – hereafter the set of **(positive) facts.** The problem is simple: given the presemantical circumstances, to find the correct valuation. We can break it up into two parts. First, how should a correct valuation deal with the language's nonsemantical vocabulary? Second, how should a correct valuation deal with the truth-predicate? For the purposes of the present discussion, the first problem will be "logical", the second "semantical".

Let v be a valuation. v is **logically closed** if it satisfies conditions (A.1)–(∀.1), and **semantically closed** (note the unorthodox usage) if it satisfies (T.1). If v satisfies the corresponding converse conditions – (A.2)–(∀.2) or (T.2) – it is **logically** or **semantically supported.** Logically (semantically) closed *and* supported valuations are **logically (semantically) balanced.** Some-

times we will abbreviate to, e.g., "l-closed", "s-balanced". Observe that logical closure, supportedness, and balance are all relative to the ground model M, though for reasons of brevity this will usually be left unremarked.

(A.1) $\vec{I}(\vec{d}) \in I^t(P) \Rightarrow \langle P\vec{d}, t\rangle \in \nu$
$\vec{I}(\vec{d}) \in I^f(P) \Rightarrow \langle P\vec{d}, f\rangle \in \nu$;

(−.1) $\langle \phi, t\rangle \in \nu \Rightarrow \langle -\phi, f\rangle \in \nu$
$\langle \phi, f\rangle \in \nu \Rightarrow \langle -\phi, t\rangle \in \nu$;

(∨.1) $\langle \phi, t\rangle \in \nu$ or $\langle \psi, t\rangle \in \nu \Rightarrow \langle \phi \vee \psi, t\rangle \in \nu$
$\langle \phi, f\rangle \in \nu$ and $\langle \psi, f\rangle \in \nu \Rightarrow \langle \phi \vee \psi, f\rangle \in \nu$;

(∀.1) $\forall c\, \langle \phi(c), t\rangle \in \nu \Rightarrow \langle (\forall x)\phi(x), t\rangle \in \nu$
$\exists c\, \langle \phi(c), f\rangle \in \nu \Rightarrow \langle (\forall x)\phi(x), f\rangle \in \nu$;

(T.1) $\langle \phi, t\rangle \in \nu \Rightarrow \langle T\ulcorner\phi\urcorner, t\rangle \in \nu$
$\langle \phi, f\rangle \in \nu \Rightarrow \langle T\ulcorner\phi\urcorner, f\rangle \in \nu$

(A.2) $\langle Pa, t\rangle \in \nu \Rightarrow I(a) \in I^t(P)$
$\langle Pa, f\rangle \in \nu \Rightarrow I(a) \in I^f(P)$,

(−.2) $\langle -\phi, f\rangle \in \nu \Rightarrow \langle \phi, t\rangle \in \nu$
$\langle -\phi, t\rangle \in \nu \Rightarrow \langle \phi, f\rangle \in \nu$;

(∨.2) $\langle \phi \vee \psi, t\rangle \in \nu \Rightarrow \langle \phi, t\rangle \in \nu$ or $\langle \psi, t\rangle \in \nu$
$\langle \phi \vee \psi, f\rangle \in \nu \Rightarrow \langle \phi, f\rangle \in \nu$ and $\langle \psi, f\rangle \in \nu$;

(∀.2) $\langle (\forall x)\phi(x), t\rangle \in \nu \Rightarrow \forall c\, \langle \phi(c), t\rangle \in \nu$
$\langle (\forall x)\phi(x), f\rangle \in \nu \Rightarrow \exists c\, \langle \phi(c), f\rangle \in \nu$;

(T.2) $\langle T\ulcorner\phi\urcorner, t\rangle \in \nu \Rightarrow \langle \phi, t\rangle \in \nu$
$\langle T\ulcorner\phi\urcorner, f\rangle \in \nu \Rightarrow \langle \phi, f\rangle \in \nu$.

(Here and throughout "∀" and "∃" do double-duty for intra- and meta-linguistic quantification.) Call a valuation **closed (supported, balanced)** if it is logically and semantically closed (supported, balanced); then ν is balanced if and only if it is both closed and supported.[16]

Are there any balanced valuations? Here is an informal argument to show that there are. Note that (A.1)–(T.1) can be seen not just as requirements on existing valuations, but also as rules for valuations' construction, e.g., "if $\vec{I}(\vec{d}) \in I^t(P)$, throw in $\langle P\vec{d}, t\rangle$", "if $\langle \phi, t\rangle$ is in, throw in $\langle -\phi, f\rangle$, and so

on. Given any supported valuation μ, the result of applying these rules to μ is again a supported valuation, and the union of an increasing sequence of supported valuations is supported too; so μ's closure under (A.1)–(T.1) is supported. On the other hand, μ's closure under (A.1)–(T.1) is certainly closed, and since both supported and closed, balanced. Thus any supported valuation μ can be developed, by application of (A.1)–(T.1), into one that is balanced.[17]

A rigorous development of the foregoing would appeal to Proposition 2, according to which repeated application of a monotonic operator to a sound set eventually produces a fixed point. What the application primarily requires is a monotonic operator J such that (i) every supported valuation is J-sound, and (ii) all of J's fixed points are balanced. First define the subsidiary jump operators J_l and J_s: J_l maps μ into the sets of all facts *obtainable from* μ by application of (A.1)–($\forall$.1), i.e., $\{\langle P\vec{a}, v\rangle \mid \vec{I}(\vec{a}) \in I^v(P)\} \cup \{\langle \neg \phi, \neg v\rangle \mid \langle \phi, v\rangle \in \mu\} \cup$ etcetera, and J_s takes μ to the set of facts obtainable from μ by application of (T.1), i.e., $\{\langle T^{\ulcorner}\phi^{\urcorner}, v\rangle \mid \langle \phi, v\rangle \in \mu\}$. A variety of serviceable jump operators, differing mainly in the relative velocities with which they prosecute different aspects of the induction, can be defined from J_l and J_s. (Perhaps the simplest is the operator taking μ to the union of $J_l(\mu)$ and $J_s(\mu)$, which essentially mimics the application of (A.1)–(T.1) to its argument.) The one considered here, due to Kripke, has among others the advantage of prolonging the process of semantical closure (which interests us) relative to that of logical closure (with which we are wearily familiar). Given a valuation μ, let $\mu^{*(l)}$ be the least logically closed extension of μ, or more precisely, the least ν such that $\mu \subseteq \nu$ and $J_l(\nu) \subseteq \nu$ (note that $\mu^{*(l)}$ is the fixed point generated by μ under the operation not of J_l but of $J_l \cup J_i$, where J_i is the identity operator.) Then Kripke's jump operator J can be defined thus: $J(\mu) = [J_s(\mu)]^{*(l)}$. It is easy to check that supported valuations are J-sound, and that J's fixed points are balanced, so Proposition 2 shows that every supported valuation has a balanced extension. Which fixed point, i.e., balanced valuation, we should ultimately opt for is a matter for further discussion, which we need not go into here;[18] the important thing for present purposes is that it is possible to construct logically impeccable valuations in which the truth-predicate applies truly to the truths, and falsely to the falsehoods. Whether this is really the result we *want* is another question, one which will be taken up in a moment.

V. TRUTH AS STRONG

By repeatedly applying J to a sound valuation ν we eventually work our way up to a fixed point ν^*. But once the fixed point is gained, Kripke's inductive construction has reached its conclusion. He was the first to point out that the journey thus far, rewarding though it has been, is far from exhausting our intuitions about truth:

> Liar sentences are not true in the object language, in the sense that the inductive process never makes them true; but we are precluded from saying this in the object language by our interpretation of negation and the truth-predicate (OTT, p. 714).

His reflections on the sources of this expressive limitation are suggestive:

> If we think of the minimal fixed point . . . as giving a model of natural language, then the sense in which we say, in natural language, that a Liar sentence is not true must be thought of as associated with some later stage in the development of natural language, one in which speakers reflect on the generation process leading to the minimal fixed point. It is not itself part of that process (OTT, p. 714).

If this is right, then truth-theorists have their work cut out for them. What is this "later stage in the development of natural language"? How can it be incorporated into semantical theory? These are the questions to be studied.

If we want to make sense of "reflection on the process", two key features of Kripke-style constructions need rethinking. One of them functions to inhibit the evaluation of sentences which ought, on reflection, to be counted false. The other leads instead to the premature falsification of sentences which maturer consideration shows to be uniquely true.

(1) In Kripke-style constructions, truth-value gaps are in an obvious sense ambiguous between *no truth-value so far* and *no truth-value ever.* This is a consequential ambiguity, because our intuitions about how to handle the two kinds of gap are rather different. If ϕ has no truth-value yet but may be getting one later, then $T^{\ulcorner}\phi^{\urcorner}$ must be left unevaluated, pending a decision on ϕ. But when the dust has settled and ϕ is *still* without a truth-value, we are apt to consider that it is not true; and if ϕ is not true, then $T^{\ulcorner}\phi^{\urcorner}$, which says that it *is* true, must be false. If these gaps are as different as they seem, a theory of treating both of them alike may not be getting the entire picture. We need to find a way of elaborating Kripke's procedures so as to allow for the making false of $T^{\ulcorner}\phi^{\urcorner}$ by the untruth of ϕ.[19]

(2) In Kripke-style constructions, $T^{\ulcorner}\phi^{\urcorner}$ is made false not by ϕ's untruth, but by its falsity. Given that his procedures should be elaborated so as to allow for the former, do there remain good grounds for retaining the latter? If the untruth of ϕ is, as it appears to be, sufficient for the falsity of $T^{\ulcorner}\phi^{\urcorner}$, might it not also be necessary? Here is an argument to show that it is. There are, in principle anyway, two kinds of false sentence: those which are false without being true, and those which are false and also true. If ϕ is false without being true, then its untruth has been conceded to be sufficient for the falsity of $T^{\ulcorner}\phi^{\urcorner}$. If ϕ is false and also true, then it is at any rate true. Since ϕ is true, it possesses the property attributed to it by $T^{\ulcorner}\phi^{\urcorner}$, whence $T^{\ulcorner}\phi^{\urcorner}$ deserves to be counted uniquely true, and in particular, *not false.*[20] To sum up, two kinds of false sentence are theoretically possible. The notion that $T^{\ulcorner}\phi^{\urcorner}$ can inherit falsity from ϕ is superfluous in connection with the first, and mistaken in connection with the second. Direct passage from the falsity of ϕ to the falsity of $T^{\ulcorner}\phi^{\urcorner}$ should be pruned from Kripke's procedures.[21]

The basic argument of (1) and (2) can be reformulated as follows. If $T^{\ulcorner}\phi^{\urcorner}$ is to mean "ϕ is true", then the evaluation of $T^{\ulcorner}\phi^{\urcorner}$ should depend on whether ϕ is true *and on nothing else.* Thus if ϕ and ψ are alike *in respect of truth*, i.e., if both are true or both are untrue, then $T^{\ulcorner}\phi^{\urcorner}$ and $T^{\ulcorner}\psi^{\urcorner}$ should be evaluated the same. If ϕ is uniquely true and ψ is both true and false, then ϕ and ψ *are* alike in respect of their truth; since it is agreed that $T^{\ulcorner}\phi^{\urcorner}$ is uniquely true, $T^{\ulcorner}\psi^{\urcorner}$ should be uniquely true too. Similarly, if ϕ is uniquely false and ψ is neither true nor false, then so far as their truth is concerned ϕ and ψ are exactly alike; since $T^{\ulcorner}\phi^{\urcorner}$ is by all accounts uniquely false, so should be $T^{\ulcorner}\psi^{\urcorner}$.

In a fixed-point language of the sort Kripke has constructed, a sentence's semantical status is exactly the same as that of its truth-sentence:

$$T^{\ulcorner}\phi^{\urcorner} \text{ is } \left\{\begin{array}{l}\text{both true and false}\\ \text{uniquely true}\\ \text{uniquely false}\\ \text{neither true nor false}\end{array}\right\} \text{ iff } \phi \text{ is } \left\{\begin{array}{l}\text{both true and false}\\ \text{uniquely true}\\ \text{uniquely false}\\ \text{neither true nor false}\end{array}\right\}$$

Against this, it has been argued that if ϕ is neither true nor false, $T^{\ulcorner}\phi^{\urcorner}$ should be uniquely false, and if ϕ is both true and false, $T^{\ulcorner}\phi^{\urcorner}$ should be

uniquely true:

$$T^{\ulcorner}\phi^{\urcorner} \text{ is } \left\{\begin{array}{l}\text{uniquely true}\\ \text{uniquely true}\\ \text{uniquely false}\\ \text{uniquely false}\end{array}\right\} \text{ iff } \phi \text{ is } \left\{\begin{array}{l}\text{both true and false}\\ \text{uniquely true}\\ \text{uniquely false}\\ \text{neither true nor false}\end{array}\right\}$$

These tables are fairly obvious elaborations of

$$\text{(W)} \qquad T^{\ulcorner}\phi^{\urcorner} \text{ is } \left\{\begin{array}{l}\text{true}\\ \text{false}\end{array}\right\} \text{ iff } \phi \text{ is } \left\{\begin{array}{l}\text{true}\\ \text{false}\end{array}\right\}$$

and

$$\text{(S)} \qquad T^{\ulcorner}\phi^{\urcorner} \text{ is } \left\{\begin{array}{l}\text{true}\\ \text{false}\end{array}\right\} \text{ iff } \phi \text{ is } \left\{\begin{array}{l}\text{true}\\ \text{untrue}\end{array}\right\}.$$

Our theorizing about truth *has* been guided by (W), the weak ideal; but it *ought* to be guided by (S), the strong. Why? Because it is the strong ideal that is *correct*, in the sense of accurately reflecting our intentions about the use of "true".[22]

There is, not surprisingly, an argument on the other side: it purports to show that the strong ideal for truth cannot be correct. On the strong conception of truth, to attempt to assign the Liar any semantical status whatever is to land oneself in contradiction. For either it is true or it isn't: yet L is untrue iff "L is true" is false iff "L is untrue", that is L, is true. On the weak conception, however, the Liar and its associates are (indifferently) under- or overdefined, so the difficulty does not arise. Now which is correct a conception of truth on which a semantical status can be found for every sentence, or one on which every attempt to assign a status to certain sentences collapses into incoherence?

To the extent that one sees the Liar paradox as only apparent, an engaging puzzle defused by the simple device of truth-value gaps (gluts), one will perhaps find this sort of argument congenial. Unfortunately, anyone who has spent time worrying about these matters must know in his or her bones that the paradox is anything *but* apparent. Admittedly there are statements very much *like* the Liar which are not genuinely paradoxical, for example, $\neg T^{\ulcorner}L^{\urcorner}$ interpreted in the manner of the weak conception of truth. But reinterpreting "true" takes us little distance towards resolving the issues the Liar raises. The sad fact is that we use the word "true" in a certain way, and that using it in that way leads to contradictions. It is

simply a question of having the courage of our semantical convictions.

In conclusion, far from thinking that the reemergence of the Liar paradox counts against the strong conception of truth, I think that a conception of truth on which the paradox does *not* arise has purchased consistency at the cost of fidelity to the issue. The choices are two: face the paradox or change the subject. If we choose, as we should, to face the paradox, our challenge is to devise a treatment of truth on which paradox survives, but without undermining the scheme of evaluation which countenances it. These matters are discussed further in Section VIII, and the development of a "paradoxical" semantics begins in Section IX.

VI. STABILITY SEMANTICS[23]

A fundamental insight of fixed point semantics is that one need not *assume* anything about "true" 's interpretation to draw *conclusions* about "true" 's interpretation (i.e., one can start the induction from the empty set). To be able to proceed from no assumptions is an attractive prospect; but someone might still worry that although this manner of proceeding is undoubtedly "sound" – all the conclusions it authorizes are correct – it might not be "complete" – one might not be able to get at *all* the correct conclusions without *assuming* something along the way. On the other hand, assumptions come with their own problems: the assumptions one makes might be wrong. Some wrong assumptions will eventually show themselves up, but others, one fears, will only perpetuate themselves. Fortunately, there seems to be a way to deal with this: randomize over all possible assumptions. Conclusions owing their persistence to wrong assumptions should disappear when those assumptions are varied, whereas correct conclusions ought to maintain their ground. The details are as follows.

Let L_T be a first-order language with truth-predicate, as outlined above. If M is a classical model of the T-less part of L_T, and U is a set of sentences, then $M + U$ is the (classical) extension of M that assigns U to T. Let V, finally, be the classical valuation scheme, i.e., the function mapping classical models of L_T into their associated classical valuations. The sequence $\langle K^\alpha(U) \mid \alpha \in OR\rangle$ can be preliminarily defined as follows:

$$(1) \qquad K^0(U) = U;$$

$$(2) \qquad K^{\alpha+1}(U) = \{\phi \mid \langle \phi, t\rangle \in V(M + K^\alpha(U))\};$$

$$(3) \qquad K^{\lambda}(U) = \lim_{\beta<\lambda} K^{\beta}(U).$$

The reason this definition is only preliminary is that it doesn't yet tell us what $\lim_{\beta<\lambda} K^{\beta}(U)$ is going to be. The $K^{\beta}(U)$'s ($\beta < \lambda$) are capable of varying in almost any way imaginable, so no simple union or intersection rule is likely to yield satisfactory results. Herzberger, Gupta, and Belnap have made separate proposals, each with its own brand of appeal, as to what kind of limit rule makes sense here. Briefly, Herzberger suggests that a sentence deserves to be in the limit interpretation if and only if it is in all of its sufficiently advanced predecessors. Formally, $\phi \in K^{\lambda}(U)$ iff $\exists \beta < \lambda \, \forall \alpha \in [\beta, \lambda) \, \phi \in K^{\alpha}(U)$; or, in the usual notation, $K^{\lambda}(U) = \liminf_{\beta<\lambda} K^{\beta}(U)$. Gupta agrees that at least these sentences should be included, but maintains that our original sentences ought to be retained too, provided that they have not been decisely repudiated in the meantime. In symbols, this comes out to $K^{\lambda}(U) = \liminf_{\beta<\lambda} K^{\beta}(U) \cup (U \cap \limsup_{\beta<\lambda} K^{\beta}(U))$. Belnap contends that both these procedures are in the end arbitrary. His suggestion is that to eliminate the arbitrariness one ought to randomize over ***all*** reasonable limit-taking procedures, where a procedure is reasonable if it locates $K^{\lambda}(U)$ anywhere between $\liminf_{\beta<\lambda} K^{\beta}(U)$ and $\limsup_{\beta<\lambda} K^{\beta}(U)$. Formally, a "bootstrapping policy" Γ is a function from limit ordinals into sets of sentences, and each bootstrapping policy determines an operator K_{Γ} such that $K_{\Gamma}^{\lambda}(U)$ is $\liminf_{\beta<\lambda} K_{\Gamma}^{\beta}(U) \cup (\Gamma(\lambda) \cap \limsup_{\beta<\lambda} K_{\Gamma}^{\beta}(U))$. These differences in limit rule ramify in intriguing ways (see Belnap, 1982; McGee, 1983), but the details do not much concern us here.

Now for the definition of "stable truth". For Herzberger and Gupta, a sentence ϕ is **stably true (false) relative to** U if and only if there is a β such that for all $\alpha > \beta$, ϕ is true (false) in $M + K^{\alpha}(U)$, and **stably true (false)** simpliciter if and only if it is stably true (false) relative to every U. Belnap's definition involves generalization over bootstrapping policies too: ϕ is **stably true (false) relative to** U **and** Γ if and only if there is a β such that for all $\alpha > \beta$, ϕ is true (false) in $M + K_{\Gamma}^{\alpha}(U)$, and **stably true (false)** simpliciter if and only if it is stable true (false) relative to every U and Γ. This is not the place to go into the virtues of stable truth and falsity as just defined (see Herzberger, Gupta, and Belnap); suffice it to say that they are many and considerable. Instead, we look at some intuitions which stable truth and falsity fail to capture.

VII. TRUTH AS GROUNDED

Stability semantics, like fixed point semantics before it, throws a lot of hard light on problems that had lain in shadow for many years. Unfortunately, one theory's strengths are the other's weaknesses. Stability semantics, for example, deals admirably with semantical paradox, where fixed point semantics is comparatively weak (see Section XV). On the other hand, fixed point semantics generally provides logical compounds with the truth-conditions intuition recommends, something that stability semantics does only sometimes. Thus in fixed point semantics a disjunction is true just in case it has a true disjunct; but a disjunction is stably true if and only if it has a true disjunct in all sufficiently advanced stages of any of a certain infinite variety of nonmonotonic progressions. The problem with this is not just that it is cumbersome, but that it seems to be inaccurate, both as a general statement of disjunctions' truth-conditions and in its application to particular cases (see below). To understand what has gone wrong, let's look again at how the semantics operate.

The idea, stripped to essentials, is to use *assumptions* to coax out conclusions not otherwise accessible, randomizing to screen out the effects of such as are mistaken. But notice that the rationale for this procedure depends on a crucial proviso: that one among the assumptions randomized over is *correct*. For if it were to turn out that the *correct* assumption had been left out, then although some of our conclusions might still be correct, others would be quite wrong, reflecting nothing more than the common error of the assumptions chosen. Now the assumptions over which stability semantics randomizes are all possible classical interpretations of the truth-predicate. Unfortunately, we have good reason to believe, and stability semantics itself confirms, that (assuming the language is not utterly benign) no such interpretation can be correct. By the foregoing, this gives us grounds for suspecting that some of the conclusions which stability semantics endorses "reflect nothing more" than the common classicality of the initial interpretations. I will mention just three examples.

(1) All theorems of first-order logic, and in particular sentences like $L \vee \neg L$, are stably true (in all three senses).

(2) Let G be $T^{\ulcorner}G^{\urcorner} \vee T^{\ulcorner}\neg G^{\urcorner}$, so that it says, in effect, that it is either true or false. Then G is stably true (in every sense).

(3) Let $\langle \phi_m \mid m \in \omega \rangle$ be a sequence of sentences, where for each m ϕ_m is the sentence $(\exists n > m)(T^{\ulcorner}\phi_n \leftrightarrow \phi_{n+1}^{\urcorner})$. Then each ϕ_m says that

there is some subsequent ϕ_n such that it is true that it is equivalent to its successor. Every ϕ_m is stably true (in every sense); in fact, they are all stabilized by $M + K^2(U)$ at the latest.

What is troubling about the attribution cited in (1) is that it has no basis in the truth-values of the evaluated sentence's parts: $L \vee \neg L$ is a true disjunction without a true disjunct to back it up. Intuitively, it seems to me, the *only* way for a disjunction to be true is via the truth of one of its disjuncts (it is just beside the point that it *would come out* true on any development of classical assumptions by methods appropriate thereto, particularly if none of those assumptions is correct anyway). And more generally, Frege taught us that reference was compositional, the references of complex sentences depending on the references of their components. Yet if the conclusions of stability semantics are to be accepted, the principle of compositionality will have to be given up. To wrap this up in an unenlighten ing slogan, truth as we know it is *supported*, yet the truth of stability semantics is not.

The evaluation in (2) feels odd in a different way. It isn't that G is without a true disjunct; it's rather that that disjunct owes its truth to G itself, which makes it hard to see how G could have come by its truth honestly, i.e., non-circularly. That the inheritance of truth-value should be non-circular is the approximate content of the requirement that truth is *forced* (see the next section for the precise definition). The requirement of forcing functions something like a "principle of sufficient reason" in semantics. A sentence's claim to its truth-value is never automatic; it has to be established, typically on the basis of other, similar, claims (this much is the requirement of supportedness). But a claim cannot be established on the basis of others unless those others can be established themselves, and on an independent basis. Sentences cannot make each other (or themselves) true for roughly the same reason that events cannot bring each other about, and citizens cannot appoint each other to positions in government.

The third evaluation is worrisome on yet another account. Circularity is not the problem, because each ϕ_m owes its truth to the equivalence of *later* ϕ_n's. And it isn't that that there aren't consecutive pairs of equivalent ϕ_n's around; for *every* n, ϕ_n and ϕ_{n+1} are true, and hence equivalent. The difficulty is rather that one can't see how ϕ_n and ϕ_{n+1} ever *got* to be true, unless it was through the equivalence of still later ϕ_p's for which the very same problem arises. What is wrong with the stated evaluations of the ϕ_m's

is that they are not *grounded*: support, and indeed independent support, can be found, but the attempt to trace the support back to its foundations leads back forever. Thus we want to require not only that truth is supported, in the sense that every evaluation has a justification, and not only that it is forced, in the sense that every evaluation has a prior justification, but also that it is *grounded*, in the sense that every chain of justifications eventually terminates in the nonsemantical circumstances (see Section VIII for an exact definition).[24]

VIII. SEMANTICAL IDEALS

If truth is strong and grounded, what should an acceptable valuation look like? On the way to an answer we will put together a modest compendium of semantical ideals, of which grounding and strength will be only the most controversial. But there is a problem: unless the language is extremely cooperative, our ideals turn out to be mutually incompatible. Even assuming that they are all *correct*, in the sense of accurately reflecting our semantical intentions, their incompatibility seems to create a strong case for giving some of them up. Or does it? That is something we will talk about shortly. In the meantime, there are the valuations to be characterized.

Ordered pairs of sentences and truth-values are called **facts**, and **(general) valuations** are arbitrary sets of facts. Of course some valuations are better than others; the problem is to say which and why. Commonsensically speaking, the semantical status of a sentence is subject to two main kinds of constraint. First, it has to be legitimately inherited. Second, it has to be properly passed along. With respect to the latter, a correct valuation ought surely to be *logically closed* in the sense already defined, i.e., to satisfy:

(A.1) $\vec{I}(\vec{d}) \in I^t(P) \Rightarrow \langle P\vec{d}, t\rangle \in \nu$
$\vec{I}(\vec{d}) \in I^f(P) \Rightarrow \langle P\vec{d}, f\rangle \in \nu$;

(−.1) $\langle \phi, t\rangle \in \nu \Rightarrow \langle -\phi, f\rangle \in \nu$
$\langle \phi, f\rangle \in \nu \Rightarrow \langle -\phi, t\rangle \in \nu$;

(∨.1) $\langle \phi, t\rangle \in \nu$ or $\langle \psi, t\rangle \in \nu \Rightarrow \langle \phi \vee \psi, t\rangle \in \nu$
$\langle \phi, f\rangle \in \nu$ and $\langle \psi, f\rangle \in \nu \Rightarrow \langle \phi \vee \psi, f\rangle \in \nu$

(∀.1) $\forall c \langle \phi(c), t\rangle \in \nu \Rightarrow \langle (\forall x)\phi(x), t\rangle \in \nu$
$\exists c \langle \phi(c), f\rangle \in \nu \Rightarrow \langle (\forall x)\phi(x), f\rangle \in \nu$;

On the other hand, the requirement

(T.1) $\langle \phi, t\rangle \in \nu \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, t\rangle \in \nu$
$\langle \phi, f\rangle \in \nu \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \nu$

of *semantical* closure is no longer wholly acceptable, reflecting as it does a conception of truth – the weak – that we have lately come to mistrust. Let valuations satisfying (T.1) be **weakly *s*-closed** from this point on; those meeting the improved condition

(T.I) $\langle \phi, t\rangle \in \nu \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, t\rangle \in \nu$
$\langle \phi, t\rangle \notin \nu \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \nu$

will be **strongly *s*-closed**, and the object of our pursuit. Summing up, a correct valuation is at least *l*-closed and strongly *s*-closed, or, as we may say, **strongly closed** simpliciter.

Turning now to the former constraint, that of legitimate inheritance, to be acceptable a valuation will have to be *logically supported*, i.e., it will have to satisfy the converses (A.2)–(∀.2) of (A.1)–(∀.1) above. But the existing requirement

(T.2) $\langle T^{\ulcorner}\phi^{\urcorner}, t\rangle \in \nu \Rightarrow \langle \phi, t\rangle \in \nu$
$\langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \nu \Rightarrow \langle \phi, f\rangle \in \nu$

of *semantical* support is in part weak, and hence objectionable. Let valuations satisfying (T.2) be **weakly *s*-supported**; the more desirable **strongly *s*-supported** valuations will be those meeting condition

(T.II) $\langle T^{\ulcorner}\phi^{\urcorner}, t\rangle \in \nu \Rightarrow \langle \phi, t\rangle \in \nu$
$\langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \nu \Rightarrow \langle \phi, t\rangle \notin \nu.$

Summing up, correct valuations should be *l*-supported and strongly *s*-supported, which we may abbreviate to **strongly supported.**

Demanding that our valuations be strongly supported is on the way to enforcing the forcing requirement, but the latter is strictly stronger than the former: for truth to be forced, each fact requires not just support, but prior support. Formally, a valuation ν is **strongly forced** if and only if there is a strict partial order $<$ on ν such that:

(A.3) $\langle P\vec{a}, t\rangle \in \nu \Rightarrow \vec{I}(\vec{a}) \in I^{t}(P);$
$\langle P\vec{a}, f\rangle \in \nu \Rightarrow \vec{I}(\vec{a}) \in I^{f}(P);$

(−.3) $\langle -\phi, t\rangle \in \nu \Rightarrow \langle \phi, f\rangle < \langle -\phi, t\rangle$;
$\langle -\phi, f\rangle \in \nu \Rightarrow \langle \phi, t\rangle < \langle -\phi, f\rangle$;

(∨.3) $\langle \phi \vee \psi, t\rangle \in \nu \Rightarrow \langle \phi, t\rangle < \langle \phi \vee \psi, t\rangle$ or $\langle \psi, t\rangle < \langle \phi \vee \psi, t\rangle$
$\langle \phi \vee \psi, f\rangle \in \nu \Rightarrow \langle \phi, f\rangle < \langle \phi \vee \psi, f\rangle$ and $\langle \psi, f\rangle < \langle \phi \vee \psi, f\rangle$;

(∀.3) $\langle (\forall x)\phi(x), t\rangle \in \nu \Rightarrow \forall c\,[\langle \phi(c), t\rangle < \langle (\forall x)\phi(x), t\rangle]$;
$\langle (\forall x)\phi(x), f\rangle \in \nu \Rightarrow \exists c\,[\langle \phi(c), f\rangle < \langle (\forall x)\phi(x), f\rangle]$;

(T.3) $\langle T^{\ulcorner}\phi^{\urcorner}, t\rangle \in \nu \Rightarrow \langle \phi, t\rangle < \langle T^{\ulcorner}\phi^{\urcorner}, t\rangle$;
$\langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \nu \Rightarrow \langle \phi, t\rangle \notin \nu$.

Intuitively, $<$ represents an order in which the truth-valuable sentences might have received their truth-values. Conditions (A.3)–(T.3) are designed to ensure that no sentence receives a value unless there are good prior grounds for giving it one.

Grounding is something further yet. For a valuation ν to be grounded, each fact in ν must depend, ultimately, on the nonsemantical circumstances (supplemented by the observation that ν fails to count certain sentences true). Officially, ν is **strongly grounded** iff there is a strict partial order $<$ on ν such that (i) (A.3)–(T.3) are satisfied, and (ii) there are no infinite descending $<$-sequences, i.e., no $\langle \phi_1, v_1\rangle, \langle \phi_2, v_2\rangle, \langle \phi_3, v_3\rangle \ldots$ such that $\ldots \langle \phi_3, v_3\rangle < \langle \phi_2, v_2\rangle < \langle \phi_1, v_1\rangle$. Condition (i) ensures that no descending $<$-path circles back on itself, condition (ii) that every descending $<$-sequence terminates, in the end, in an appropriate atomic fact.

By analogy with our previous definition, logically closed and supported valuations will be **logically balanced**, and semantically closed and supported valuations will be **semantically balanced** (the prefix "strongly" is taken for granted). As before, valuations both logically and semantically closed (supported, balanced) are **closed (supported, balanced)** simpliciter; evidently a valuation is balanced iff it is closed and supported. Finally, valuations satisfying all the desiderata mentioned above, i.e., valuations which are both closed and grounded, will be called **ideal**.

The first thing to notice about ideal valuations is that if there are paradoxes about, there aren't any; in fact, if there are paradoxes about, no valuation is even balanced. To see why not, suppose that the language contains a Liar sentence, i.e., a sentence L identical to $-T^{\ulcorner}L^{\urcorner}$. If ν is balanced, then $\langle L, t\rangle \in \nu \Leftrightarrow \langle -T^{\ulcorner}L^{\urcorner}, t\rangle \in \nu \Leftrightarrow \langle T^{\ulcorner}L^{\urcorner}, f\rangle \in \nu \Leftrightarrow \langle L, t\rangle \notin \nu$, which is a contradiction. And any paradox can be made to lead to the

same result (see Section XV). What is in some ways even worse, closed, *forced* valuations are not always possible even in the absence of paradox; that is, some nonparadoxes (in particular, some "undecidable" sentences, in the sense of Section XIV below) are nonetheless troublesome enough to prevent any valuation from being both.[25]

To these discomfiting revelations two rather different reactions are possible: the sensible and the heroic. The sensible response is to reason that since our desiderata cannot all be satisfied, some of them will have to be given up. To which the heroic retort is that none of them can be given up. If grounding and closure accurately reflect our intentions about the use of our semantical terms, then so long as it is *truth* and *falsity* we mean to talk about, we have no choice but to follow them where they lead.

Since the sensible response is probably second nature to everyone, some time will be spent motivating its opponent. Think of grounding and closure as guiding (or regulative) ideals for the use of "true" and "false". We have discovered that our guiding ideals cannot all be realized. Must we not then sacrifice some of them? But why? It is in the *nature* of guiding ideals that their appropriateness and their capacity to guide do not depend on the possibility of their realization. The unattainability of a perfect valuation can no more repudiate our guiding ideals for truth than the impossibility of a perfectly virtuous person our guiding ideals for human conduct.

The proposal that we compromise our semantical ideals is in essence a proposal that we change what we mean by "true" and "false". The idea seems to be that we can begin with the existing meanings, do away with the parts that trouble us, and retain all the rest as is. Perhaps we would now call the Liar "neither true nor false", or "both true and false", but in connection with ordinary sentences "true" and "false" would mean what they always had. But is this really coherent? Not on any ordinary conception of meaning, for on any ordinary conception of meaning, to change a word's meaning *somewhere* is to change it *everywhere*. (Suppose someone proposed compelling reasons, say a hitherto unsuspected contradiction, why we should change the meaning of "person" very slightly, in order that a certain Mr. Baker should no longer fall into its extension. If the recommended change took place, it would not be only in connection with Mr. Baker that the meaning of "person" had changed: I would mean something different when I called *you* a person.) And now the question is: why should anyone *care* about "truth" in the new sense? We care whether *S* is

"true" in the ordinary sense because we care whether things really are the way *S* says they are. But if "true" were used so that *S*'s untruth was no longer sufficient for the truth of "*S* is untrue", then the constituting link between truth and things being as they were said to be would be gone. Truth thus reconceived would be distinctly less important, and less intelligible, than the notion it replaced. Better to live with the paradoxes and mean what we *feel* like meaning than be done with the paradoxes and be saddled with a notion that no longer speaks to our expressive needs.

The sensible reaction to all this heroic protestation is, well, sensible. Sooner or later we are going to have to violate some of our semantical ideals. Do it now, and we can at least choose where and how the violation occurs. To blunder heroically on is not to escape the necessity of compromise, but to squander one's influence over the form it takes. Of course this strikes our semantical hero as an ignoble collaborationism. If we are to be forced to violate our ideals, so be it, but nothing can make us renounce them. Even if it is granted that some kind of violation is inevitable, that is no argument for capitulating, for one doesn't know in advance how much will have to be conceded; to abandon our ideals now is to run the risk of falling further short of them than was really necessary. Our only conscientious option is to proceed as though the valuations we wanted could be had, and see what results. More specifically, we must employ methods which *would* produce ideal valuations if such existed, and see what they *do* produce. In the next section one such method will be introduced.

IX. STAGE SEMANTICS

The truth of certain sentences is due, and indeed entirely due, to the untruth of certain others. From this it would appear that one cannot identify all of the truths until one has identified all of the untruths. On the other hand, it is not obvious how one can identify even a *single* untruth without first identifying *every* truth (until all the truths have been found, each candidate for untruth is potentially an as yet unidentified truth). If it takes all the untruths to find all the truths, and all the truths to find *any* untruths, the prospects for telling one from the other look dim.

There is, fortunately, a way out: develop a series of increasingly accurate *approximations* to the set of truths, feeding the sentences not true *on the current approximation* back into the construction of its successor. But

where to begin? Not a single untruth can be identified, however tentatively, until a sizeable collection of truths has been assembled. Fortunately things aren't quite so bad going in the other direction. To find every truth one needs all the untruths, but *some* truths can be discovered before *any* untruths have been. Why not begin with these, that is with the set of all untruth-independent truths? As an approximation to the set of *all* truths, it is admittedly disappointing. But it affords us an estimate of the set of untruths, which is just what is needed for the construction of a second, and hopefully better, approximation to the set of truths. Some of the shortcomings of the second approximation can be eliminated by applying it in a similar way to the construction of a third, and then a fourth, fifth, sixth, and so on. If all goes well, the resulting succession of approximations will somehow converge, with the actual truths and untruths emerging in the process.

The semantics will be (essentially) *dependence-style*, in the sense that a sentence is true (false) if and only if the attempt to trace its truth-(falsity-) conditions back through the maze of its semantical ancestors is ultimately successful. The business of tracing semantical lineages is handled by *dependence relations*, which are defined as follows. Let S be an arbitrary set of facts. An **S-dependence relation** is a binary relation Δ on the set of all facts such that:

(G) $\langle \phi, v\rangle \in S \Rightarrow \langle \phi, v\rangle$ bears Δ to nothing ($v = t$ or f);

Otherwise:

(A.4) $\langle P\vec{a}, v\rangle$ bears Δ to nothing $\Leftrightarrow \vec{I}(\vec{a}) \in I^v(P)$ $(v = t$ or $f)$;
$\langle P\vec{a}, v\rangle$ bears Δ to itself $\Leftrightarrow \vec{I}(\vec{a}) \notin I^v(P)$ $(v = t$ or $f)$;

(¬.4) $\langle \neg\phi, t\rangle$ bears Δ to $\langle \phi, f\rangle$;
$\langle \neg\phi, f\rangle$ bears Δ to $\langle \phi, t\rangle$;

(∨.4) $\langle \phi \vee \psi, t\rangle$ bears Δ to exactly one of $\langle \phi, t\rangle$ and $\langle \psi, t\rangle$;
$\langle \phi \vee \psi, f\rangle$ bears Δ to $\langle \phi, f\rangle$ and $\langle \psi, f\rangle$;

(∀.4) $\langle (\forall x)\phi(x), t\rangle$ bears Δ to each $\langle \phi(c), t\rangle$;
$\langle (\forall x)\phi(x), f\rangle$ bears Δ to exactly one $\langle \phi(c), f\rangle$.

(T.4) $\langle T^{\ulcorner}\phi^{\urcorner}, t\rangle$ bears Δ to $\langle \phi, t\rangle$;
$\langle T^{\ulcorner}\phi^{\urcorner}, f\rangle$ bears Δ to itself.[26]

One sees from clause (A.4) that S-dependence relations are relative to the ground model M.

A finite or infinite sequence of facts is called a **Δ-path** if (a) the first of any two consecutive entries bears Δ to the second, and (b) each entry has only finitely many predecessors. A fact is **Δ-grounded** if it heads no infinite Δ-paths. Finally, a fact is **grounded in S** if it is Δ-grounded for at least one S-dependence relation Δ.

Intuitively, what kind of fact is grounded in S? Let the ***S*-ground** be the set of all facts $\langle \phi, v \rangle$ such that for some S-dependence relation Δ, $\langle \phi, v \rangle$ does not bear Δ to any fact. Evidently the S-ground comprises (i) the members of S and (ii) the obtaining nonsemantic atomic facts (i.e., those mentioned in the first part of condition (A.4)). Facts in the S-ground, because they bear some S-dependence relation to nothing, are "immediately" grounded in S. Any other fact is grounded in S if and only if there is an S-dependence relation Δ such that it bears Δ to facts which bear Δ to facts which . . . belong to the S-ground.

Let S_* be the set of all facts grounded in S.[27] By what was said above, our "take-off" valuation Ω_0 will be Λ_*, the set of all facts grounded in the empty set.[28] To get an idea of what Ω_0 contains, let χ be any atomic truth of the ground model M; then Ω_0 contains $\langle \chi, t \rangle$, $\langle \neg\chi, f \rangle$, $\langle T^{\ulcorner}\chi^{\urcorner}, t \rangle$, $\langle \neg T^{\ulcorner}T^{\ulcorner}\phi^{\urcorner\urcorner}, f \rangle$, $\langle T^{\ulcorner}\chi^{\urcorner} \vee T^{\ulcorner}\neg\chi^{\urcorner}, t \rangle$, $\langle T^{\ulcorner}\chi \vee \neg\chi^{\urcorner}, t \rangle$, and indefinitely many more of the same type. What they all have in common, of course, is that they depend only on other facts *obtaining*, never on other facts *failing* to obtain. Facts with the latter sort of dependence, for example all those of the form $\langle T^{\ulcorner}\phi^{\urcorner}, f \rangle$ or $\langle \neg T^{\ulcorner}\phi^{\urcorner}, t \rangle$, are conspicuously absent from the initial valuation. The reason for their absence has already been mentioned: no untruths can be established, however tentatively, until an approximation to the set of truths has been put forward. Fortunately, the latter approximation is provided by the initial valuation itself, and its complement, i.e., the set of sentences Ω_0 fails to make true, can now serve as our first approximation to the set of untruths.

The next step is dictated by the strong ideal for truth. Sentences untrue on the current approximation deserve – according to the best information now available, but that goes without saying – to have false truth-sentences. Arrange for this as follows: gather every $\langle T^{\ulcorner}\psi^{\urcorner}, f \rangle$ such that $\langle \psi, t \rangle \notin \Omega_0$ into a set, and let the collection of facts grounded therein be Ω_0's successor Ω_1. Although the untruths of Ω_0 aren't describable as such in Ω_0, they are so

describable in Ω_1; and so a certain amount of progress has been made. But notice that the untrue sentences – the sentences that Ω_1 doesn't make true – still don't coincide with the sentences *declared* untrue – the sentences ψ such that Ω_1 makes $T^{\ulcorner}\psi^{\urcorner}$ false and $-T^{\ulcorner}\psi^{\urcorner}$ true (as we might have expected, given that the sentences Ω_1 declares untrue are the untruths of the *previous* valuation Ω_0). To redress the imbalance, we carry out the adjustment again; and again, and again, and again. It only remains to make the foregoing fully explicit.

Let L be an operator on (general) valuations, defined as follows:

$$(L) \qquad L(\nu) = \{\langle T^{\ulcorner}\psi^{\urcorner}, f\rangle \mid \langle \psi, t\rangle \notin \nu\}_*$$

Let the **initial stage** Ω be Λ_*, and for each ordinal α let the **αth stage** Ω_α be $L^\alpha(\Omega)$. Since $[\ ]_*$ is a monotonic operator, $\nu \subseteq \mu \Rightarrow \{\langle T^{\ulcorner}\psi^{\urcorner}, f\rangle \mid \langle \psi, t\rangle \notin \mu\} \subseteq \{\langle T^{\ulcorner}\psi^{\urcorner}, f\rangle \mid \langle \psi, t\rangle \notin \nu\} \Rightarrow \{\langle T^{\ulcorner}\psi^{\urcorner}, f\rangle \mid \langle \psi, t\rangle \notin \mu\}_* \subseteq \{\langle T^{\ulcorner}\psi^{\urcorner}, f\rangle \mid \langle \psi, t\rangle \notin \nu\}_* \Rightarrow L(\mu) \subseteq L(\nu)$, whence L is an antimonotonic operator. $[\ ]_*$'s monotonicity implies also that for all S, $\Omega = \Lambda_* \subseteq S_*$, and this shows at once that Ω is supersound. From the facts that L is antimonotonic, and Ω is supersound, several things follow:

(1) $\{\Omega_\alpha \mid \alpha \in OR\}$ telescopes (Proposition 3);

(2) $\forall \iota \Omega_\iota$ is inferior & $\forall \sigma \Omega_\sigma$ is superior (remarks following Proposition 3);

(3) $\exists \beta\ [\forall \iota \geqslant \beta\ \Omega_\iota = \underline{\Omega}\ \&\ \forall \sigma \geqslant \beta\ \Omega_\sigma = \overline{\Omega}]$ (Proposition 4);

(4) $\overline{\Omega} - \underline{\Omega} = \{\langle \phi, v\rangle \mid \forall \sigma\ \langle \phi, v\rangle \in \Omega_\sigma\ \&\ \forall \iota\ \langle \phi, v\rangle \notin \Omega_\iota\}$ (Proposition 5).

Let's try now to bring the procession of stages into sharper focus. Each stage is a proposed evaluation of our formal language L_T. An ideal proposal would be a stage Ω_α which (among other things) made the truth-sentences of its truths true and those of its untruths false. The initial stage Ω_0 performs rather miserably in this regard, leaving unevaluated many sentences which it "ought" to have made false (e.g., the truth-sentences of its untruths) and many more which it "ought" to have made true (e.g., the negations of those truth-sentences). Ω_0's miserliness is, however, a tremendous boon to $\Omega_1 = \{\langle T^{\ulcorner}\psi^{\urcorner}, f\rangle \mid \langle \psi, t\rangle \notin \Omega_0\}_*$, which feeds on the sentences Ω_0 fails to satisfy. Ω_1's overgenerosity has the effect, in turn, of starving

Ω_2, which responds by growing up only slightly bigger than Ω_0. And evidently the pattern will be repeated indefinitely, with each underfed Ω_{2n} underevaluating the language, and each overfed Ω_{2n+1} overevaluating it. Which takes us to the first limit stage: what happens there? Since the odd-numbered Ω_n's overevaluate L_T, the superior limit of the Ω_n's does too; whence, for the same reasons as before, Ω_ω gives out too few truth-values.[29] In this way the oscillation begins anew, and since a similar scene is enacted at every limit ordinal, the stages perpetually alternate between even-numbered underevaluations, and odd-numbered overevaluations, of L_T.

Aimless though all this meandering might seem, progress is being made, for the underevaluating even stages are gradually growing, and the over-evaluating odd stages are gradually shrinking. Although much remains to be clarified, this much is clear already: as α gets bigger, Ω_α gets better.

X. INTERPRETATION AND EVALUATION

The strong ideal for truth imposes two conditions on the interpretation of T: $T^\ulcorner\phi^\urcorner$ should be true if and only if ϕ is true, and $T^\ulcorner\phi^\urcorner$ should be false if and only if ϕ is not true. The next proposition shows that the first of these conditions is met at every stage.

PROPOSITION 6. $\forall\alpha\,\forall\phi\,[\langle T^\ulcorner\phi^\urcorner, t\rangle \in \Omega_\alpha \Leftrightarrow \langle\phi, t\rangle \in \Omega_\alpha]$.

Proof. If S is a set of facts, let $S|_v$ be $\{\chi \mid \langle\chi, v\rangle \in S\}$. If $S|_t$ is empty, every S-dependence relation relates $\langle T^\ulcorner\phi^\urcorner, t\rangle$ to $\langle\phi, t\rangle$, and to nothing else. Thus for all S-dependence relations $\triangle$, $\langle T^\ulcorner\phi^\urcorner, t\rangle$ is $\triangle$-grounded iff $\langle\phi, t\rangle$ is $\triangle$-grounded, whence $\langle T^\ulcorner\phi^\urcorner, t\rangle \in S_*$ iff $\langle\phi, t\rangle \in S_*$. But every stage Ω_α is equal to S_* for some S with empty $S|_t$; the result follows. □

By Proposition 6, at every stage the truths and the sentences declared true coincide. The relation between the untruths and the sentences declared untrue is more complicated, and more interesting.

PROPOSITION 7. (1) $\forall\iota\,\forall\psi\,[\langle T^\ulcorner\psi^\urcorner, f\rangle \in \Omega_\iota \Rightarrow \langle\psi, t\rangle \notin \Omega_\iota]$

(2) $\forall\sigma\,\forall\psi\,[\langle\psi, t\rangle \notin \Omega_\sigma \Rightarrow \langle T^\ulcorner\psi^\urcorner, f\rangle \in \Omega_\sigma]$

Proof. (1) Let ι be given. If $\iota = 0$, there is no ψ such that $\langle T^\ulcorner\psi^\urcorner, f\rangle \in \Omega_\iota$. If $\iota > 0$, then $\Omega_\iota \subseteq \Omega_{\iota-1}$. But $\Omega_\iota \subseteq \Omega_{\iota-1} \Rightarrow \{\psi \mid \langle\psi, t\rangle \notin \Omega_{\iota-1}\} \subseteq \{\psi \mid \langle\psi, t\rangle \notin \Omega_\iota\} \Rightarrow \{\psi \mid \langle T^\ulcorner\psi^\urcorner, f\rangle \in \Omega_\iota\} \subseteq \{\psi \mid \langle\psi, t\rangle \notin \Omega_\iota\}$.

(2) Let σ be given. By Proposition 4, $\Omega_{\sigma-1} \subseteq \Omega_\sigma$. But $\Omega_{\sigma-1} \subseteq \Omega_\sigma \Rightarrow \{\psi \mid \langle \psi, t\rangle \notin \Omega_\sigma\} \subseteq \{\psi \mid \langle \psi, t\rangle \notin \Omega_{\sigma-1}\} \Rightarrow \{\psi \mid \langle \psi, t\rangle \notin \Omega_\sigma\} \subseteq \{\psi \mid \langle T^\ulcorner \psi^\urcorner, f\rangle \in \Omega_\sigma\}$. □

From Proposition 7 it appears that the even stages are "conservative" in the sense that everything they declare untrue is really, by their lights, untrue; unfortunately, some untruths are left undeclared. By contrast, the more "liberal" odd stages declare each of their untruths untrue, but unhappily they extend the honour to certain true sentences as well.

XI. CONSISTENCY AND COMPLETENESS

A set of S facts is **consistent** if there is no sentence χ such that S contains both $\langle \chi, t\rangle$ and $\langle \chi, f\rangle$, and **complete** if for every sentence χ, either $\langle \chi, t\rangle$ or $\langle \chi, f\rangle$ is in S. Let the **αth interpretation of T**, or τ_α for short, be $\{\langle \chi, t\rangle \mid \langle T^\ulcorner \chi^\urcorner, t\rangle \in \Omega_\alpha\} \cup \{\langle \chi, f\rangle \mid \langle T^\ulcorner \chi^\urcorner, f\rangle \in \Omega_\alpha\}$. (Note that the presence of $\langle \chi, t\rangle$ ($\langle \chi, f\rangle$) in τ_α indicates *not* that χ is true (false) at stage α, but that it belongs to T's extension (antiextension) at stage α.) Then we have the following result about the consistency and completeness of T's interpretations.

PROPOSITION 8. $\forall \iota\, \tau_\iota$ is consistent & $\forall \sigma\, \tau_\sigma$ is complete.

Proof. [σ] Let σ be given. Then for some ι, $\sigma = \iota + 1$. If $\langle \chi, t\rangle$ is in Ω_ι, then it's in $\Omega_{\iota+1}$ (since $\Omega_\iota \subseteq \Omega_{\iota+1}$), and therefore also in $\tau_{\iota+1}$ (by Proposition 6). If $\langle \chi, t\rangle$ is not in Ω_ι, then by definition $\langle T^\ulcorner \chi^\urcorner, f\rangle$ is in $\Omega_{\iota+1}$, whence $\langle \chi, f\rangle$ is in $\tau_{\iota+1}$.

[ι] Let ι be an even successor ordinal. If $\langle \chi, f\rangle$ is in τ_ι, then $\langle \chi, t\rangle$ isn't in $\Omega_{\iota-1}$, or Ω_ι (since $\Omega_\iota \subseteq \Omega_{\iota-1}$), or τ_ι either (by Proposition 6). So for all successor ι, τ_ι is consistent. If ι isn't a successor, $\iota + 2$ is; since $\tau_\iota \subseteq \tau_{\iota+2}$, τ_ι is consistent too. □

Thus the even-numbered interpretations are consistent, and the odd are complete. What about the stages themselves? Let τ be an arbitrary interpretation of T (intrinsically, of course, τ is nothing but a valuation, a set of facts). Then a **(general) model $M + \tau$** of L_T is just like a ground model M, except that where M assigns nothing to the truth-predicate, $M + \tau$ assigns it extension $\{\phi \mid \langle \phi, t\rangle \in \tau\}$ and antiextension $\{\phi \mid \langle \phi, f\rangle \in \tau\}$. Each (general)

model $M + \tau$ induces a (general) valuation $V(M + \tau)$, most conveniently defined as the logical closure (i.e., the closure under (A.1)–(∀.1)) of $\{\langle T^{\ulcorner}\phi^{\urcorner}, v\rangle \mid \langle \phi, v\rangle \in \tau\}$. If for each α we let $M_\alpha = M + \tau_\alpha$, then it is easy to see that Ω_α is just $V(M\alpha)$, the valuation induced thereby. Now let a model (or ground model) of L_T be **total** (**partial**) if it assigns every predicate in its domain jointly exhaustive (mutually exclusive) extension and antiextension. (Note that models both total and partial are essentially classical, and that V maps all such models into their associated classical valuations.) The next proposition relates the character of the ground model M to the consistency and/or completeness of the stages.

PROPOSITION 9. (1) If M is partial, then $\forall \iota\, \Omega_\iota$ is consistent.
(2) If M is total, then $\forall \sigma\, \Omega_\sigma$ is complete.

Proof. If M is partial, then by Proposition 8 M_ι is partial too. Expand M_ι into a classical model M'_ι of L_T. Then $V(M'_\iota)$ is a classical valuation of L_T, and therefore consistent. Since clearly $V(M_\iota) \subseteq V(M'_\iota)$, $V(M_\iota)$ is consistent too. If M is total, then Proposition 8 shows that M_σ is total too. Reduce M_σ to a classical model M'_σ of L_T. Since $V(M'_\sigma)$ is a classical valuation of L_T, it is complete. Clearly $V(M'_\sigma) \subseteq V(M_\sigma)$, so $V(M_\sigma)$ is complete too.

Note that Proposition 9 implies that if M is classical, the even-numbered stages are all consistent, and the odd are all complete.

XII. UNIVALENTS, GAPS, AND GLUTS

Let S be a set of facts. ϕ is a **gap** in S iff neither $\langle \phi, t\rangle$ nor $\langle \phi, f\rangle$ is in S; a **glut** in S iff both $\langle \phi, t\rangle$ and $\langle \phi, f\rangle$ are in S; and **univalent** in S iff it is neither a gap nor a glut in S, i.e., iff $\exists !v\langle \phi, v\rangle \in S$. A univalent sentence ϕ is **uniquely** v, or $v!$, in S, if $v = (\iota u)(\langle \phi, u\rangle \in S)$. To give a sentence's **truth-status** in S is to say whether it is, in S, uniquely true, uniquely false, a gap, or a glut. In this section we ask: how does truth-status develop as α increases?

Let $\langle S_\alpha \mid \alpha \in OR\rangle$ be a sequence of sets of facts, and imagine that we are proceeding through it in the order given. At any given point in our progress, a certain number of sentences are *stabilized*, i.e., possessed of a truth-status they are destined to retain forever. Not it is interesting to note that no

matter how erratically the S_α's vary, the complement of stabilized sentences can never shrink, but only grow, or, after a certain point, remain the same. And in a certain way the fact that every *currently* stabilized sentence will *remain* stabilized seems to reassure us that stabilization is a robust attribute, an attribute with integrity. Until, that is, we begin to reflect that a sentence's current stabilization simply *consists* in its retaining its status, and therefore its stability, forever; at which point the reassurance may begin to seem less solid than before. (Compare: immortality, once acquired, is retained forever; but that is because of logic, not the integrity of the immortal constitution.) Similarly, the persistence of stabilization might seem, at first, to reassure us epistemologically. But not on reflection. Although it is true that the stabilized sentences may be relied on to remain so, short of working through the remainder of the sequence one has no way of telling which sentences they are. (Compare: whoever among us is going to live forever is immortal *now*; but that is again logic, not advance information.)

The persistence of stabilization, because it is basically a logical phenomenon, turns out to be less encouraging than it at first appeared. If we now ask what *would* be encouraging, an answer suggests itself: some sort of persistence of truth-status itself. The next few propositions show that truth-status in τ_α and truth-status in Ω_α do in fact persist.

PROPOSITION 10. $\forall\sigma\,\forall\phi\,\forall v\,[\phi$ is $v!$ in $\tau_\sigma \Rightarrow \phi$ is $v!$ in $\tau_{\sigma+1}]$.

Proof. ϕ is $t!$ in $\tau_\sigma \Rightarrow \langle\phi, f\rangle \notin \tau_\sigma \Rightarrow \langle\phi, t\rangle \in \Omega_{\sigma-1} \Rightarrow \langle\phi, t\rangle \in \tau_{\sigma-1}$ (Proposition 6) $\Rightarrow \langle\phi, t\rangle \in \tau_{\sigma+1}$ ($\tau_{\sigma-1} \subseteq \tau_{\sigma+1}$) $\Rightarrow \phi$ is $t!$ in $\tau_{\sigma+1}$ (since $\tau_{\sigma+1}$ is consistent). ϕ is $f!$ in $\tau_\sigma \Rightarrow \langle\phi, t\rangle \notin \tau_\sigma \Rightarrow \langle\phi, t\rangle \notin \Omega_\sigma$ (Proposition 6) $\Rightarrow \langle\phi, f\rangle \in \tau_{\sigma+1} \Rightarrow \phi$ is $f!$ in $\tau_{\sigma+1}$ (since $\tau_{\sigma+1}$ is consistent). □

PROPOSITION 11. $\forall\iota\,\forall\phi\,\forall v\,[\phi$ is $v!$ in $\tau_\iota \Rightarrow \phi$ is $v!$ in $\tau_{\iota+1}]$,

Proof. ϕ is $t!$ in $\tau_\iota \Rightarrow \langle\phi, t\rangle \in \tau_\iota \Rightarrow \langle\phi, t\rangle \in \Omega_\iota$ (Proposition 6) $\Rightarrow \langle\phi, f\rangle \notin \tau_{\iota+1} \Rightarrow \phi$ is $t!$ in $\tau_{\iota+1}$ (since $\tau_{\iota+1}$ is complete). ϕ is $f!$ in $\tau_\iota \Rightarrow \langle\phi, f\rangle \in \tau_\iota \Rightarrow \langle\phi, t\rangle \notin \Omega_{\iota-1} \Rightarrow \langle\phi, t\rangle \notin \Omega_{\iota+1}$ (since $\Omega_{\iota+1} \subseteq \Omega_{\iota-1}$) $\Rightarrow \langle\phi, t\rangle \notin \tau_{\iota+1}$ (Proposition 6) $\Rightarrow \phi$ is $f!$ in $\tau_{\iota+1}$ (since $\tau_{\iota+1}$ is complete).

PROPOSITION 12. (1) $\forall\alpha\,\forall\phi\,\forall v\,[\phi$ is $v!$ in $\tau_\alpha \Rightarrow \forall\beta > \alpha\ \phi$ is $v!$ in $\tau_\beta]$

(2) $\forall\sigma\,\forall\phi\,[\phi$ is a gap in $\tau_\sigma \Rightarrow \forall\beta > \sigma\ \phi$ is a gap in $\tau_\beta]$

(3) $\forall\iota\,\forall\phi\,[\phi$ is a glut in $\tau_\iota \Rightarrow \forall\beta > \iota\ \phi$ is a glut in $\tau_\beta]$.

Proof. (1) Let ϕ be $v!$ in τ_α. If α is even, then $\forall \iota \geqslant \alpha\ \tau_\alpha \subseteq \tau_\iota$. Since each τ_ι is consistent, $\forall \iota \geqslant \alpha\ \phi$ is $v!$ in τ_ι. $\forall \sigma > \alpha\ \exists \iota \geqslant \alpha\ \tau_\sigma = \tau_{\iota+1}$, so by Proposition 11 $\forall \sigma > \alpha\ \phi$ is $v!$ in τ_σ. Suppose next that α is odd. By Proposition 10, ϕ is $v!$ in $\tau_{\alpha+1}$. Since $\alpha + 1$ is even, the first part of the proof shows that $\forall \beta > \alpha + 1\ \phi$ is $v!$ in τ_β.

(2) and (3) Clearly $\langle \Omega_\sigma \mid \alpha \in OR \rangle$ telescopes $\Rightarrow \langle \tau_\alpha \mid \alpha \in OR \rangle$ telescopes. It follows that τ_σ (τ_ι) is a superset (subset) of every subsequent τ_β. □

PROPOSITION 13. (1) $\forall \alpha\ \forall \phi\ \forall v\ [\phi$ is $v!$ in $\Omega_\alpha \Rightarrow \forall \beta > \alpha\ \phi$ is $v!$ in $\Omega_\beta]$

(2) $\forall \sigma\ \forall \phi\ [\phi$ is a gap in $\Omega_\sigma \Rightarrow \forall \beta > \sigma\ \phi$ is a gap in $\Omega_\beta]$

(3) $\forall \iota\ \forall \phi\ [\phi$ is a glut in $\Omega_\iota \Rightarrow \forall \beta > \iota\ \phi$ is a glut in $\Omega_\beta]$.

Proof. (1) Proof is by induction on ϕ.

[A] Trivial if ϕ is nonsemantic atomic.

[T] $T^\ulcorner\theta^\urcorner$ is $v!$ in $\Omega_\alpha \Rightarrow \theta$ is $v!$ in $\tau_\alpha \Rightarrow \theta$ is $v!$ in $\tau_\beta \Rightarrow T^\ulcorner\theta^\urcorner$ is $v!$ in Ω_β.

[—] $-\theta$ is $v!$ in $\Omega_\alpha \Rightarrow \theta$ is $(-v)!$ in $\Omega_\alpha \Rightarrow \theta$ is $(-v)!$ in $\Omega_\beta \Rightarrow \theta$ is $v!$ in Ω_β (here $-t$ is f and $-f$ is t).

[$\vee$] $\theta_1 \vee \theta_2$ is $t!$ in $\Omega_\alpha \Rightarrow \exists i\ \theta_i$ is $t!$ in $\Omega_\alpha \Rightarrow \exists i\ \theta_i$ is $t!$ in $\Omega_\beta \Rightarrow \theta_1 \vee \theta_2$ is $t!$ in Ω_β. $\theta_1 \vee \theta_2$ is $f!$ in $\Omega_\alpha \Rightarrow \forall i\ \theta_i$ is $f!$ in $\Omega_\alpha \Rightarrow \forall i\ \theta_i$ is $f!$ in $\Omega_\beta \Rightarrow \theta_1 \vee \theta_2$ is $f!$ in Ω_β.

[$\forall$] Similar to [$\vee$].

(2) and (3) By analogy with Proposition 12. □

Once a sentence becomes uniquely true or false, it remains so forever. And once a sentence becomes a gap in some odd stage, or a glut in some even that is how it stays. But where in the procession of stages *do* sentences become uniquely true or false, gaps or gluts? Here there are large differences between the univalents, on the one hand, and the gaps and gluts, on the other. Every stable gap (glut) is a gap (glut) from the initial stage Ω_0 onward, but new unique truths and falsehoods can emerge arbitrarily late in the game, subject only to the language's cardinality and expressive power.

These claims are established as follows. It should be obvious that nonsemantical atomic sentences are gaps (gluts) in Ω_0 if they are gaps (gluts) anywhere. As for semantical atomic sentences, if $T^\ulcorner\chi^\urcorner$ is a gap in Ω_σ, then χ must be true in $\Omega_{\sigma-1}$ (or else $T^\ulcorner\chi^\urcorner$ would have been false in Ω_σ). But if χ is true in $\Omega_{\sigma-1}$, so is $T^\ulcorner\chi^\urcorner$; and since $\Omega_{\sigma-1} \subseteq \Omega_\sigma$, $T^\ulcorner\chi^\urcorner$ is true in Ω_σ too,

contradicting our assumption that $T^{\ulcorner}\chi^{\urcorner}$ was a gap in Ω_σ. The argument against late-coming semantical atomic gluts is similar, and the proof that there are no late-coming gaps or gluts of any kind is a straightforward induction on complexity. To see that unique truths and falsehoods can turn up as late as the language's expressive capabilities allow, consider the sequence $\langle \phi_\alpha \mid \alpha \leqslant \eta \rangle$ defined as follows: ϕ_0 is any nonsemantical atomic gap of Ω_0; for each $\alpha < \eta$, $\phi_{\alpha+1}$ is $-T^{\ulcorner}\phi_\alpha^{\urcorner}$; and for each $\lambda \leqslant \eta$, ϕ_λ is $(\exists \sigma < \lambda) -T^{\ulcorner}\phi_\sigma^{\urcorner}$. Then every ϕ_ι (other than ϕ_0) becomes uniquely false in Ω_ι, and every ϕ_σ becomes uniquely true in Ω_σ.

Such are the basic facts about the stages' development. We turn now to their maturity. Truth, falsity, and related notions are defined in Section XIII, decidability and undecidability are considered in the section following, and the last section deals with paradox.

XIII. TRUTH

At the beginning we set our sights on a "later stage in the development of natural language, one in which speakers reflect on the generation process leading to the minimal fixed point" (OTT, p. 714). No sooner do speakers notice that the forces generating their evaluation are spent than that intelligence demands deployment in the service of a new evaluation. But this applies as much to the second evaluation as the first, as much to the third as the second, and so on indefinitely. "Reflection on the process" is not just the springboard for a single "later stage", but the driving force behind an extended succession of such stages.

Eventually, though, reflection too exhausts itself, in the sense that it will finally have contributed all it can to our estimate of the language's evaluation. This is not to say that a single, final, evaluation ever emerges, because it doesn't. Instead, we find ourselves driven back and forth between a pair of estimates: our sequence's inferior limit $\underline{\Omega}$, and its superior limit $\overline{\Omega}$. But then how is the language to be evaluated?

Given a sentence ϕ, there are theoretically four possibilities for its truth-status: uniquely true, uniquely false, neither true nor false, or true and false. Since $\underline{\Omega}$ and $\overline{\Omega}$ are the only valuations reflection countenances, the sentences on which they agree are the sentences on whose truth-status reflection yields a definite verdict. These are the *decidable* sentences. If a sentence is not decidable, then it is devoid of truth-status. It would be wrong,

in particular, to think of undecidable sentences as being neither true nor false, because that would be to *allow* them a truth-status.

Officially, a sentence ϕ is **true** iff $\langle \phi, t \rangle$ is in both $\underline{\Omega}$ and $\overline{\Omega}$, and **untrue** iff it is in neither. ϕ is **false** iff $\langle \phi, f \rangle$ is in both $\underline{\Omega}$ and $\overline{\Omega}$, and **unfalse** (sorry) iff it is in neither. A number of important semantical notions can be defined in terms of truth, falsity, untruth, and unfalsity. A sentence ϕ is **uniquely true** iff it is true and unfalse, and **uniquely false** iff it is false and untrue. It is **neither true nor false** iff it is untrue and unfalse, and **true and false** iff it is, well, true and false. A sentence has **truth-features** iff it is true, false, untrue, or unfalse, and a **truth-status** iff it is uniquely true, uniquely false, neither true nor false, or true and false. Finally, ϕ is **decidable** iff it has a truth-status, and **undecidable** otherwise.

The definitions just given force us to acknowledge an ambiguity in our previous use of "untrue". In one sense, an untrue sentence is one which is *not*, in the final analysis, *true*; in another, it is one which *is*, in the final analysis, *not true*. Evidently untruth as just defined is untruth in the second sense. Sentences untrue in the first sense will be **nontrue** from this point on. Since any sentence that is, in the end, not true is not, in the end, true, every untruth is nontrue. But not every nontruth is untrue, as we shall see.

Now that untruth and nontruth have been distinguished, there can be little doubt that we *intend* the truth predicate to apply falsely not just to the untruths, but to nontruths in general. That is, we intend that "ψ is true" should be false whenever ψ fails, in any manner whatever, to be true (we do not insist that ψ go on to *succeed* in being untrue). Unfortunately, this intention of ours, taken in combination with certain others, has turned out to be unfulfillable. Thus it is imperative to distinguish between how we *intend* truth to work and how it *actually* works.

Our semantical intentions are intentions to do something impossible, namely to evaluate sentences in accordance with certain incompatible principles. Notwithstanding the well-known difficulties involved in *doing* impossible things, there seems to be little difficulty in *intending* to do them, even in cases where one knows the thing intended to be impossible.[30] If someone is trapped behind a brick wall in urgent circumstances, I may, and in fact probably should, form the intention of breaking the wall down. Nor is the situation any different when what one intends is logically impossible, and known to be: nothing prevents me from intending to square the circle. Of course, what I *actually* do depends not just on what I mean to do but on

the constraints, be they empirical or logical, under which I operate, and that is why what is done is so often different from what was intended. At the same time, to understand what people actually do one needs to know what they mean to be doing, no less when the intention is foiled than when it succeeds. The upshot of all this for truth is as follows. To illuminate our semantical practice a theory of truth must *accord with*, or even somehow *embody*, our semantical intentions; but it would be pointless to require it to *deliver* truth *as we intend it*, because truth as we intend it is impossible.[31] What the theory can and should deliver is the result of running our intentions up against the bounds of possibility, i.e., truth as it actually works.

How does truth actually work? Rather like this, I think: the truth-predicate applies truly to true sentences, and falsely to untrue sentences. As for the rest, the evidence suggests that in actual fact we simply don't know how to proceed. And this is not because we aren't clever enough; we can actually see that given our semantical intentions, there *is* no rational way to proceed. The next proposition shows that this is how truth works in the theory, too.

PROPOSITION 14.

$$T^{\ulcorner}\phi^{\urcorner} \text{ is } \begin{Bmatrix}\text{true}\\ \text{false}\\ \text{undecidable}\end{Bmatrix} \Leftrightarrow \phi \text{ is } \begin{Bmatrix}\text{true}\\ \text{untrue}\\ \text{undecidable}\end{Bmatrix}.$$

Proof. By Proposition 6, $\langle T^{\ulcorner}\phi^{\urcorner}, t\rangle \in \underline{\Omega}(\overline{\Omega}) \Leftrightarrow \langle \phi, t\rangle \in \underline{\Omega}(\overline{\Omega})$, so $T^{\ulcorner}\phi^{\urcorner}$ is true $\Leftrightarrow$ ϕ is true. Now suppose that $T^{\ulcorner}\phi^{\urcorner}$ is false. Then $\langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \underline{\Omega}$. Since by Proposition 4 $\underline{\Omega} = L(\overline{\Omega}) = \{\langle T^{\ulcorner}\psi^{\urcorner}, f\rangle \mid \langle \psi, t\rangle \notin \overline{\Omega}\}_*$, $\langle \phi, t\rangle \notin \overline{\Omega}$. Since $\underline{\Omega} \subseteq \overline{\Omega}$, $\langle \phi, t\rangle \notin \underline{\Omega}$ either, whence ϕ is untrue. Conversely, if ϕ is untrue, then $\langle \phi, t\rangle \notin \overline{\Omega}$, so $\langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \underline{\Omega}$. Since $\underline{\Omega} \subseteq \overline{\Omega}$, $\langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \overline{\Omega}$ too, so $T^{\ulcorner}\phi^{\urcorner}$ is false. Let ϕ be undecidable. By Proposition 17, $\forall v\ [\langle \phi, v\rangle \notin \underline{\Omega}\ \&\ \langle \phi, v\rangle \in \overline{\Omega}]$. Since $\overline{\Omega} = L(\underline{\Omega})$ and $\underline{\Omega} = L(\overline{\Omega})$, $\langle \phi, t\rangle \notin \underline{\Omega} \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \overline{\Omega}$, and $\langle \phi, t\rangle \in \overline{\Omega} \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \notin \underline{\Omega}$. By Proposition 6, $\langle \phi, t\rangle \in \overline{\Omega} \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, t\rangle \in \overline{\Omega}$, and $\langle \phi, t\rangle \notin \underline{\Omega} \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, t\rangle \notin \underline{\Omega}$. It follows that $T^{\ulcorner}\phi^{\urcorner}$ is undecidable. The argument is reversible, so $T^{\ulcorner}\phi^{\urcorner}$ undecidable $\Rightarrow$ ϕ undecidable. □

Like the truth-predicate of English, L_T's truth-predicate applies truly to its truths, falsely to its untruths, and undecidably to its undecidables. Thus

L_T is capable of representing truth-in-L_T, admittedly not as it was intended, but as it is. Can it represent any other of its semantic notions? Define the predicates F, $T!$, $F!$, GA, and GL, as follows: $F\ulcorner\phi\urcorner =_{df} T\ulcorner\neg\phi\urcorner$, $T!\ulcorner\phi\urcorner =_{df} T\ulcorner\phi\urcorner \& \neg F\ulcorner\phi\urcorner$, $F!\ulcorner\phi\urcorner =_{df} F\ulcorner\phi\urcorner \& \neg T\ulcorner\phi\urcorner$, $GA\ulcorner\phi\urcorner =_{df} \neg T\ulcorner\phi\urcorner \& \neg F\ulcorner\phi\urcorner$, and $GL\ulcorner\phi\urcorner =_{df} T\ulcorner\phi\urcorner \& F\ulcorner\phi\urcorner$.[32] Arguments like the one just given show that these predicates represent, in the sense outlined above, L_T-falsity and all the L_T truth-statuses; for example, $GA\ulcorner\phi\urcorner$ is true/false/undecidable iff ϕ is neither true nor false/true or false/undecidable.

Consider the following four components of the strong ideal for truth: $T\ulcorner\phi\urcorner$ is true $\Rightarrow$ ϕ is true; ϕ is true $\Rightarrow$ $T\ulcorner\phi\urcorner$ is true; $T\ulcorner\phi\urcorner$ is false $\Rightarrow$ ϕ is non-true; and ϕ is nontrue $\Rightarrow$ $T\ulcorner\phi\urcorner$ is false. The first and third fall under the heading of supportedness, the second and fourth under that of closure. As we have defined it, truth satisfies all these conditions but the last (only *un*truths have false truth-sentences); which explains why it is possible to prove truth fully grounded (Proposition 15), but only partly closed (Proposition 16).

PROPOSITION 15. The valuation $\{\langle\phi, t\rangle \mid \phi \text{ is true}\} \cup \{\langle\phi, f\rangle \mid \phi \text{ is false}\}$ is grounded.

Proof. Note first that the valuation in question is $\underline{\Omega} = \{\langle T\ulcorner\psi\urcorner, f\rangle \mid \langle\psi, t\rangle \notin \overline{\Omega}\}_*$. For every $\langle\phi, v\rangle \in \underline{\Omega}$ and $\overline{\Omega}$-dependence relation Δ grounding $\langle\phi, v\rangle$, let $R(\langle\phi, v\rangle, \Delta)$ be $\sup\{R(\langle\chi, w\rangle, \Delta) + 1 \mid \langle\phi, v\rangle \Delta \langle\chi, w\rangle\}$ (recall that $\sup(\Lambda) = 0$), and let $R(\langle\phi, v\rangle)$ be the least of the $R(\langle\phi, v\rangle, \Delta)$'s. Define $\langle\phi, v\rangle < \langle\phi', v'\rangle \Leftrightarrow R(\langle\phi, v\rangle) < R(\langle\phi', v'\rangle)$. Then $<$ is clearly a strict well-founded partial order on $\underline{\Omega}$, and it is easy to check that (A.3)–(T.3) are satisfied. □

PROPOSITION 16. The valuation $\{\langle\phi, t\rangle \mid \phi \text{ is true}\} \cup \{\langle\phi, f\rangle \mid \phi \text{ is false}\}$ satisfies every condition of closure other than the last, i.e., $\langle\phi, t\rangle \notin \nu \Rightarrow \langle T\ulcorner\phi\urcorner, f\rangle \in \nu$.

Proof. The valuation in question consists of all facts grounded in a certain set S $(= \{\langle T\ulcorner\psi\urcorner, f\rangle \mid \langle\psi, t\rangle \notin \overline{\Omega}\})$, i.e., all facts grounded by at least one S-dependence relation Δ. If $I(a) \in I^v(P)$, then by (A.4) $\langle Pa, v\rangle$ is grounded by some (in fact every) such Δ; if $\langle\phi, v\rangle$ is grounded by some such Δ, then (−.4) implies that $\langle\neg\phi, -v\rangle$ is grounded by the same one: and similarly for disjunction, quantification, and application of T. □

XIV. UNDECIDABILITY

Undecidables have been defined as sentences without truth-status. But so far the way has been left open for them to have truth-*features*. For example, if $\overline{\Omega}$ contained both $\langle \psi, t \rangle$ and $\langle \psi, f \rangle$, but $\underline{\Omega}$ contained $\langle \psi, t \rangle$ only, then although ψ would not be decidable, it would be true. The next proposition shows that this theoretical possibility notwithstanding, statusless sentences are featureless as well.

PROPOSITION 17. No undecidable is true, false, untrue, or unfalse.

Proof. It suffices to show that all undecidables are gaps in $\underline{\Omega}$ and gluts in $\overline{\Omega}$. Let ψ be undecidable. If ψ were $v!$ in either $\underline{\Omega}$ or $\overline{\Omega}$, then by Proposition 12 it would be $v!$ in both, contradicting its undecidability. ψ can't be a gap (glut) in both $\underline{\Omega}$ and $\overline{\Omega}$, or it would be decidable, so it must be a gap in one and a glut in the other. Since $\underline{\Omega} \subseteq \overline{\Omega}$, ψ is a gap in $\underline{\Omega}$ and a glut in $\overline{\Omega}$. □

Sentences can be undecidable for a variety of reasons. The most notorious cases are of course the *paradoxes*, for example the Liar sentence L. Although a precise definition must be postponed until Section XV, the outlines of the notion are clear enough: a sentence is paradoxical if the assumption that it has truth-features is inevitably self-refuting. Decidable sentences can not only be *assumed* to have truth-features, they actually have them; so we see why paradoxes are never decidable (this will be proved in the next section).

But not all undecidables are paradoxes. For there are many sentences which can be assumed to have truth-features, but only at the expense of the requirements of support, forcing, or grounding. If I say, "The Liar is true or it isn't", then though there is no objection to *assuming* my statement true, any attempt to *support* this assumption is bound to lead into contradictions. Not so with "This very sentence is true or it isn't", for its truth, once assumed, supports itself; but notice that only its *un*truth is capable of providing an *independent* basis for its truth, whence if the requirement of forcing is observed the ascription of truth will be impossible. Consider, finally, an ω-sequence of sentences, each asserting the truth-or-untruth of its successor (i.e., the nth reads, "the n+1st sentence is either true or it isn't"). The requirements of support and forcing are not violated if we assume all these sentences to be true, but for obvious reasons the requirement of grounding is. So here we have four kinds of undecidable

sentence, arranged in decreasing order of infamy: paradoxes, violators of support, violators of forcing but not support, and violators of grounding but not forcing.

A sentence is decidable just in case it has a truth-status: uniquely true, uniquely false, neither true nor false, or both true and false. But to date we have seen no evidence that the last two possibilities are ever realized. As it turns out, everything depends on the character of the ground model M.

PROPOSITION 18. (1) M is total $\Leftrightarrow (\forall\theta) - (\theta$ is neither true nor false);
(2) M is partial $\Leftrightarrow (\forall\theta) - (\theta$ is both true and false);
(3) M is classical $\Leftrightarrow (\forall\theta)\,(\theta$ decidable $\Rightarrow \theta$ is uniquely true or uniquely false).

Proof. (1) [$\Rightarrow$] Let M be total. By Proposition 9, $\overline{\Omega}$ is complete, so no sentence is a gap in both $\underline{\Omega}$ and $\overline{\Omega}$. [$\Leftarrow$] If M isn't total then there is an n-ary predicate P ($\neq T$) such that $I^t(P) \cup I^f(P) \neq D^n$. Assume without loss of generality that P is unary; let $x \in D - [I^t(P) \cup I^f(P)]$, and choose c so that $I(c) = x$. Since Ω_1 is superior in the sequence of stages, it suffices to show that $Pc \notin \mathrm{dom}(\Omega_1)$. By definition no S-dependence relation grounds $\langle Pc, v\rangle$ unless either $I(c) \in I^v(P)$ or $\langle Pc, v\rangle \in S$. By hypothesis neither $I^t(P)$ nor $I^f(P)$ contains $I(c)$, and obviously neither $\langle Pc, t\rangle$ nor $\langle Pc, f\rangle$ is in $S = \{\langle T^{\ulcorner}\psi^{\urcorner}, f\rangle \mid \langle\psi, t\rangle \notin \Omega_0\}$ either. It follows that $\langle Pc, v\rangle$ is not grounded in S, whence $Pc \notin \mathrm{dom}(\Omega_1)$.

(2) [$\Rightarrow$] If M is partial, $\underline{\Omega}$ is consistent, so no sentence is a glut in both $\underline{\Omega}$ and $\overline{\Omega}$. [$\Leftarrow$] Let $I(c) \in I^t(P) \cap I^f(P)$. It suffices to show that $\langle Pc, t\rangle$ and $\langle Pc, f\rangle$ are in Ω_0. Clearly for any S, if $I(c) \in I^v(P)$ then $\langle Pc, v\rangle$ is grounded in S. Since $I(c)$ is in both $I^t(P)$ and $I^f(P)$, both $\langle Pc, t\rangle$ and $\langle Pc, f\rangle$ are grounded in Λ, i.e., are in Ω_0.

(3) M is classical $\Leftrightarrow M$ is partial and total. The result follows.[33] □

A sentence which was neither true nor false, or both true and false, would be decidable, so Proposition 18 shows that if M is classical, no gaps or gluts exist. This conclusion contrasts sharply with the fairly general impression that the paradoxes *force* us to recognize the existence of gaps and/or gluts. That these devices contribute nothing to the resolution of the paradoxes has already been maintained. Now it appears further that the paradoxes do not

even generate pressure for their introduction. Not that gaps and gluts do not sometimes appear in languages containing paradoxes (on the contrary, L_T contains gaps and/or gluts if M is nonclassical); the point is only that the gaps and gluts are not there *because* the paradoxes are.[34]

XV. PARADOX

Recent analyses of paradox have tended to rely on the following intuition: the paradoxical sentences are those which not only don't, but *can't*, possess semantical attributes. And typically, ϕ *can* have an attribute if and only if there are tenable assumptions, that is assumptions with a certain amount of semantical integrity, relative to which it *does* have it.

Consider first the analysis offered by fixed point semantics.[35] Here the "semantical attributes" are truth and falsity; the "assumptions" are the consistent valuations; and an assumption is "tenable" just in case it is a fixed point of J, or, equivalently, just in case it is weakly balanced (see Section IV). A sentence is "capable of truth (falsity)" if and only if it is true (false) in some tenable assumption, and "paradoxical" if and only if it cannot be true and cannot be false. That at least is how the analysis is usually understood, but a slight emendation may be in order. Since every consistent, weakly supported valuation has a consistent, weakly balanced extension, the insistence on *balance* is ultimately superfluous, and may distract attention from the supportedness that is really doing all the work. It would be better, then, to reconstrue the "tenable" assumptions as the consistent, weakly *supported* valuations, letting sentences unevaluated by any tenable assumption be paradoxical as before.

To a first approximation, the last analysis can be seen as emphasizing the consistency of a sentence's antecedents, the next as fastening instead on a certain kind of consistency in its consequences.[36] In stability semantics, the "semantical attributes" are once again truth and falsity, but the "assumptions" become the language's *classical* valuations. An assumption ν is "tenable" if and only if every sentence stably true (false) *relative to* ν (strictly speaking, to the set of ν's truths) is already true (false) *in* ν; and ϕ is "capable of truth (falsity)" if and only if there is a tenable assumption $V(M + U)$ (and bootstrapping policy Γ) such that ϕ is true (false) in every $V(M + K^{\alpha}_{(\Gamma)}(U))$. As before, a sentence is "paradoxical" if it is incapable both of truth and of falsity.

The present approach agrees that the paradoxical sentences are those which can't have "semantical attributes", and that the latter, in turn, are the sentences to which no "tenable assumption" concedes semantical attributes. But it disagrees about what the "semantical attributes" are, about what an "assumption" ought to look like, and about what "tenability" comes to.

Should the "semantical attributes" be the truth-values, truth and falsity, alone? It may be agreed that the paradoxes are incapable of truth and falsity, but that is a distinction they share with a good many nonparadoxes, e.g., "Viruses are alive" and "The king of France is bald". (Remember that there is no tampering with the ground model, only with the interpretation of T.) Intuitively, what distinguishes the paradoxes from sentences like these is that the paradoxes resist *all* classification in terms of truth and falsity; they not only can't be true, or false, they can't be *un*true, or *un*false. And that seems like a good reason for taking the "semantical attributes" to be truth, falsity, untruth, and unfalsity. Let truth and falsity be the **positive** truth-features, untruth and unfalsity the **negative**. Ordered pairs of sentences and truth-features are **facts, positive** or **negative** according to the character of their second elements. Let an **assessment** be any set of (positive and negative) facts; it will be **incoherent** if it contains a positive fact $\langle \phi, v \rangle$ along with the corresponding negative $\langle \phi, \bar{v} \rangle$, and **coherent** otherwise. **Assumptions**, for us, will be coherent assessments.

The semantical ideals of Section VIII are straightforwardly adaptable to assessments. An assessment Φ is ***l*-closed** if it satisfies (A.5)–(∀.5), and ***l*-supported** if it satisfies their converses (A.6)–(∀.6). (Note that the latter have not been written out; they are obtained from (A.5)–(∀.5) by reversing the directions of all the arrows.)

$$
\begin{aligned}
(\text{A.5}) \qquad & \vec{I}(\vec{a}) \in I^t(P) \Rightarrow \langle P\vec{a}, t \rangle \in \Phi \\
& \vec{I}(\vec{a}) \notin I^t(P) \Rightarrow \langle P\vec{a}, \bar{t} \rangle \in \Phi; \\
& \vec{I}(\vec{a}) \in I^f(P) \Rightarrow \langle P\vec{a}, f \rangle \in \Phi \\
& \vec{I}(\vec{a}) \notin I^f(P) \Rightarrow \langle P\vec{a}, \bar{f} \rangle \in \Phi;
\end{aligned}
$$

$$
\begin{aligned}
(-.5) \qquad & \langle \phi, t \rangle \in \Phi \Rightarrow \langle -\phi, f \rangle \in \Phi \\
& \langle \phi, \bar{t} \rangle \in \Phi \Rightarrow \langle -\phi, \bar{f} \rangle \in \Phi; \\
& \langle \phi, f \rangle \in \Phi \Rightarrow \langle -\phi, t \rangle \in \Phi \\
& \langle \phi, \bar{f} \rangle \in \Phi \Rightarrow \langle -\phi, \bar{t} \rangle \in \Phi;
\end{aligned}
$$

(∨.5) $\langle\phi, t\rangle \in \Phi$ or $\langle\psi, t\rangle \in \Phi \Rightarrow \langle\phi \vee \psi, t\rangle \in \Phi$
$\langle\phi, \bar{t}\rangle \in \Phi$ and $\langle\psi, \bar{t}\rangle \in \Phi \Rightarrow \langle\phi \vee \psi, \bar{t}\rangle \in \Phi$;
$\langle\phi, f\rangle \in \Phi$ and $\langle\psi, f\rangle \in \Phi \Rightarrow \langle\phi \vee \psi, f\rangle \in \Phi$
$\langle\phi, \bar{f}\rangle \in \Phi$ or $\langle\psi, \bar{f}\rangle \in \Phi \Rightarrow \langle\phi \vee \psi, \bar{f}\rangle \in \Phi$;

(∀.5) $\forall c\, \langle\phi(c), t\rangle \in \Phi \Rightarrow \langle(\forall x)\phi(x), t\rangle \in \Phi$
$\exists c\, \langle\phi(c), \bar{t}\rangle \in \Phi \Rightarrow \langle(\forall x)\phi(x), \bar{t}\rangle \in \Phi$;
$\exists c\, \langle\phi(c), f\rangle \in \Phi \Rightarrow \langle(\forall x)\phi(x), f\rangle \in \Phi$
$\forall c\, \langle\phi(c), \bar{f}\rangle \in \Phi \Rightarrow \langle(\forall x)\phi(x), \bar{f}\rangle \in \Phi$.

If Φ is *l*-closed and -supported, it is ***l*-balanced.** Next, Φ is **(strongly) *s*-closed** if it satisfies (T.V), and **(strongly) *s*-supported** if it satisfies (T.V)'s converse (T.VI) (i.e., the result of reversing the direction of the arrows in (T.V)). From this point on the prefix "strongly" will usually be taken for granted.

(T.V) $\langle\phi, t\rangle \in \Phi \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, t\rangle \in \Phi$
$\langle\phi, \bar{t}\rangle \in \Phi \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, \bar{t}\rangle \in \Phi$;
$\langle\phi, t\rangle \in \Phi \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, \bar{f}\rangle \in \Phi$
$\langle\phi, \bar{t}\rangle \in \Phi \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \Phi$.

Assessments both *s*-closed and -supported are ***s*-balanced.** If Φ is *l*- and *s*-closed (balanced), it is **closed (balanced)**; clearly an assessment is balanced iff it is *l*-balanced and *s*-balanced.

Recall that if there are paradoxes in the language, no valuation can be strongly balanced. With assessments matters are different. In fact we are already acquainted, albeit indirectly, with a strongly balanced assessment, namely $\{\langle\phi, t\rangle \mid \phi \text{ true}\} \cup \{\langle\phi, f\rangle \mid \phi \text{ false}\} \cup \{\langle\phi, \bar{t}\rangle \mid \phi \text{ untrue}\} \cup \{\langle\phi, \bar{f}\rangle \mid \phi \text{ unfalse}\}$. And there are more. By analogy with Section IV, let $\mathscr{F}(\Phi)$ be the smallest logically closed extension of $\{\langle T^{\ulcorner}\phi^{\urcorner}, t\rangle, \langle T^{\ulcorner}\phi^{\urcorner}, \bar{f}\rangle \mid \langle\phi, t\rangle \in \Phi\} \cup \{\langle T^{\ulcorner}\phi^{\urcorner}, f\rangle, \langle T^{\ulcorner}\phi^{\urcorner}, \bar{t}\rangle \mid \langle\phi, \bar{t}\rangle \in \Phi\}$. Then every supported assessment is $\mathscr{F}$-sound, and all of $\mathscr{F}$'s fixed points are balanced. By Proposition 2, every supported assessment has a balanced extension.

Now what makes an assumption, i.e., a coherent assessment, tenable? If we follow the lead of fixed point semantics, we will count Φ tenable iff it is balanced, or, what leads to the same results, iff it is supported. But is that a good lead to follow? It is generally acknowledged that the fixed point-semantical analysis of paradox has a number of awkward consequences, most prominently the paradoxicality of logical theorems like $L \vee -L$ and $(\forall\chi)(T^{\ulcorner}\chi^{\urcorner} \vee -T^{\ulcorner}\chi^{\urcorner})$ and semantical laws like $(\forall\chi) -(T^{\ulcorner}\chi^{\urcorner}$ &

$T^{\ulcorner}\neg\chi^{\urcorner}$). Most, if not all, of these anomalies can be traced back to the insistence that tenable valuations be supported. Intuitively, assuming $\psi \vee \neg\psi$ true does not require one to assume the truth either of ψ or of its negation; yet a supported valuation counts the disjunction true only if it can do likewise with one of the disjuncts. In short, supportedness appears to be more than (minimal) semantical virtue requires.

If supportedness is too much to ask, nothing at all is too little. Facts have got to have *something* to say for themselves to earn admission into tenable assumptions, or else anything can be assumed, and nothing is paradoxical. So it will help to look a little more closely at where supportedness goes wrong. Let assessment Φ **support** assessment Ψ if and only if $\Psi \subseteq \mathscr{F}(\Phi)$. Since every supported assumption is sound, every supported assumption supports itself. But why should an assumption have to support *itself* – that is, provide grounds for the inclusion of each of its own members – to be tenable? Certainly tenable assumptions should be *hospitable* to their support, but it is hard to see why they should have to go so far as to *provide* it. Which suggests we require not that Φ support itself, but that it be extendible into a Φ' which supports it. And since tenable assumptions should also be hospitable to their support's support, their support's support's support, and so on, the chain of support should be, in a manner shortly to be made precise, indefinitely continuable. The idea is that an assumption Φ is tenable if it is possible to "fill in" beneath it with a series of further assumptions, each extending Φ, and each provided for by the one (or ones) underneath. Another way of thinking of it is that an *assessment* is tenable if the attempt to "fill in" in this way need not lead into incoherence. The official definition is as follows: an assessment Φ is **tenable** or **supportable** iff there is a totally ordered index set $\langle S, < \rangle$ and a sequence $\langle \Phi_s \mid s \in S \rangle$ of assessments such that

(a) $\langle S, < \rangle$ has a last element 0;
(b) $\forall s \in S\ (s < 0 \Rightarrow s$ has an immediate successor $s')$;
(c) $\langle S, < \rangle$ has no first element;
(d) $\Phi\ (= \Phi_0)$ is the last element of $\langle \Phi_s \mid s \in S \rangle$;
(e) $\forall s \in S\ \ \Phi \subseteq \Phi_s$;
(f) $\forall s \in S\ \ \Phi_s$ is coherent;
(g) if s has an immediate predecessor r in $\langle S, < \rangle$ then Φ_r supports Φ_s;

(h) if s has no immediate predecessor in $\langle S, < \rangle$, then for some coherent Ψ between liminf $\{\Phi_r \mid r < s\}$ and limsup $\{\Phi_r \mid r < s\}$, Ψ supports Φ_s.

Note that supportable assessments are always coherent, hence always assumptions. And note also that for assumptions, supportedness can be construed as the simplest case of supportability, the one in which it's possible to let $\langle S, < \rangle$ be $\{\ldots -3, -2, -1, 0\}$ in the order given, and $\Phi_s = \Phi$ for each s in S.

Now to the definition of paradox. A sentence ϕ **can be true (false, untrue, unfalse)** if and only if $\langle \phi, t \rangle$ ($\langle \phi, f \rangle$, $\langle \phi, \bar{t} \rangle$, $\langle \phi, \bar{f} \rangle$) belongs to some supportable assumption. ϕ is **paradoxical** if and only if it cannot be true, or false, or untrue, or unfalse. Equivalently, paradoxes are sentences not in the domain of any supportable assumption. There follow examples of the notion's application to various kinds of sentence.

(1) *Liar*: Let L be $-T^{\ulcorner}L^{\urcorner}$. Then L is, unsurprisingly, paradoxical.

(2) *MetaLiar*: Let L' be what Brian Skyrms calls a MetaLiar, that is a sentence *other* than L which also says that L is untrue. Then L' is, like L, paradoxical, reflecting the fact that on the present approach the problem is not that (i) L is untrue, but can't itself *say* so, but rather than (ii) L is genuinely paradoxical, hence unamenable to construal as true *or* untrue.

(3) *κ-Liar*: Let $T^{\alpha+1}\phi$ be $T^{\ulcorner}T^{\alpha}\phi^{\urcorner}$, and $T^{\lambda}\phi$ be $(\forall \beta < \lambda)T^{\beta}\phi$. Then a κ-Liar is a sentence ϕ identical to $-T^{\kappa}\phi$. κ-Liars are always paradoxical.

(4) *Weakened Liar*: Let W be $F^{\ulcorner}W^{\urcorner}$, i.e., $T^{\ulcorner}-W^{\urcorner}$. Then W says that W is false. It is sometimes observed that intuitively, W is no less paradoxical than L: if it were false, then given what it says, it would be unfalse; and if it were unfalse, then given what it says it would be false. As we might have hoped, W is paradoxical in the sense defined.

(5) *Paradox Without Self-Reference*: Here is an example designed to show that self-reference is not essential to paradox. For each $m \in \omega$, let ϕ_m be $(\forall n > m) - T^{\ulcorner}\phi_n^{\urcorner}$, so that each ϕ_m says that every succeeding ϕ_n is untrue. An intuitive argument shows that every one of these sentences is paradoxical. If ϕ_m were true, then given what it says, every succeeding ϕ_n would be untrue; but if so then every ϕ_n after ϕ_{m+1} is untrue, whence ϕ_{m+1} is true after all. If ϕ_m were untrue, then there would be an $n > m$ such that ϕ_n was true; but then by the argument just given ϕ_{n+1} would be both true and untrue. Once again, each ϕ_m is paradoxical in the sense defined above.

(6) *Logical Theorems and Semantical Laws*: $L \vee -L$ is stably true in stability semantics, but paradoxical in fixed point semantics! Intuitively, though, it is neither true nor paradoxical, but somewhere in between. Likewise on the present theory. It is not true because truth is logically supported, and neither L nor $-L$ is true. At the same time, it *can* be true, as is shown by the support sequence $\{\Phi_{-n} \mid n \in \omega\}$, where $\Phi_0 = \{\langle L \vee -L, t\rangle\}$, $\Phi_{-(2n+1)} = \{\langle L, t\rangle, \langle L \vee -L, t\rangle\}$, and $\Phi_{-(2n+2)} = \{\langle L, f\rangle, \langle L, \bar{t}\rangle, \langle L \vee -L, t\rangle\}$. That *no* logical theorems, and few if any semantical laws, are paradoxical follows from item (8) below.

(7) *Decidables*: Let $\Phi = \{\langle \phi, t\rangle \mid \phi \text{ is true}\} \cup \{\langle \phi, f\rangle \mid \phi \text{ is false}\} \cup \{\langle \phi, \bar{t}\rangle \mid \phi$ is untrue$\} \cup \{\langle \phi, \bar{f}\rangle \mid \phi$ is unfalse$\}$. It follows from Proposition 17 that a sentence is decidable iff it is in Φ's domain. Since no sentence is both true and untrue, or both false and unfalse, Φ is coherent, and it is easy to check that it supports itself. Thus the support sequence $\langle \ldots \Phi, \Phi, \Phi, \Phi\rangle$ shows that Φ is supportable, and from this it follows that every paradox is undecidable.

(8) *Stability Paradoxes*: Let ϕ be nonparadoxical in the sense of Belnap. Then there are U, v, and Γ such that $\langle \phi, v\rangle \in \bigcap_\alpha V(M + K_\Gamma^\alpha(U))$. Cardinality considerations show that there must be β and γ $(\beta \neq \gamma)$ such that $V(M + K_\Gamma^\beta(U)) = V(M + K_\Gamma^\gamma(U))$. Fix β and γ, and let $\langle S, <\rangle$ be $\{\langle -n, \alpha\rangle \mid n \in \omega, \beta < \alpha \leqslant \gamma\}$, lexicographically ordered. If for each element $\langle -n, \alpha\rangle$ of S $\Phi_{\langle -n, \alpha\rangle}$ is defined as $\{\langle \phi, t\rangle, \langle \phi, \bar{f}\rangle \mid \langle \phi, t\rangle \in V(M + K_\Gamma^\alpha(U))\} \cup \{\langle \phi, f\rangle, \langle \phi, \bar{t}\rangle \mid \langle \phi, f\rangle \in V(M + K_\Gamma^\alpha(U))\}$, then the resulting sequence $\langle \Phi_s \mid s \in S\rangle$ meets conditions (a)–(h) above, whence $\Phi_{\langle 0, \kappa\rangle}$ is supportable and ϕ is nonparadoxical. Thus every sentence nonparadoxical in the sense of Belnap is nonparadoxical in the sense defined above. It is not hard to see that Belnap's nonparadoxes include Herzberger's and Gupta's, so any sentence nonparadoxical in any known scheme of stability semantics is nonparadoxical for us. Whether our notion coincides with Belnap's I do not know.[37, 38]

XVI. CONCLUSION

Many topics have not been covered, in most cases because I don't know quite what to say about them. Would it be possible to add a decidability predicate to the language? What about stronger connectives, like exclusion negation or Lukasiewicz implication? Would an expanded language do better at expressing its own semantics? Would it contain new and more terrible paradoxes? Can the account be supplemented with a workable notion of inherent truth (see note 36)? In what sense does stage semantics

lie "between" fixed point and stability semantics? In what sense, exactly, are our semantical rules inconsistent? In what sense, if any, does their inconsistency resolve the problem of the paradoxes?

The ideals of strength, grounding, and closure together define an intuitively appealing conception of truth. Nothing would be gained by insisting that it was *the* intuitive conception of truth, and in fact recent developments make me wonder whether such a thing exists. However that may be, until the alternatives are better understood it would be foolish to attempt to decide between them. Truth gives up her secrets slowly and grudgingly, and loves to confound our presumptions.

NOTES

[1] I would like to thank George Bealer, Charles Chihara, Donald Davidson, Anil Gupta, Hans Herzberger, Paul Kube, Shaughan Lavine, Vann McGee, George Myro, Albert Visser, and Peter Woodruff for many enlightening and encouraging conversations. I am especially grateful to Hans Herzberger, who sparked my interest in truth, and who gave me the idea it was possible to think for myself. When this paper was in preparation Peter Woodruff sent me a copy of his penetrating "Paradox, Truth, and Logic I". Our approaches overlap in several areas. First, we are alike in using *arbitrary* subsets of {sentences} × $\{t, f\}$ – what he calls "approximate predicates" and I call "general valuations" – both to interpret T and to evaluate the language. Second, there is the application of Kleene's strong valuation scheme to the associated models (this Professor Woodruff traces back to Dunn). Third, there is the emphasis, much more explicit in his work than mine, on the complementarity of completeness and consistency. These formal similarities notwithstanding, a preliminary look at "Paradox, Truth and Logic II" (in preparation as I write) suggests to me that our responses to the paradoxes are somewhat different. Here I am closer to Terence Parsons, whose "Assertion, Denial, and the Liar Paradox" draws (what I think is) the crucial distinction between semantical fiat and semantical fact. Thanks to the editor and the referee for a number of useful criticisms, and thanks to the Social Sciences and Humanities Research Council of Canada for financial assistance.

[2] Thanks to Albert Visser for emphasizing this point to me.

[3] The illustration is borrowed from John Searle, who used it in a different connection.

[4] Of course, this reasoning provides no guarantee that there will be an explicit, formal way of doing it. Our ability to tell truths from untruths (given the relevant non-semantical facts) might turn out to be irreducibly informal, more like our ability to identify faces than our ability to identify prime numbers. Of course we hope it does not.

[5] Gupta speaks of the "descriptive problem of explaining our use of the word 'true', and, in particular, of giving the meaning of sentences containing 'true'" (TP, p. 1). His problem sounds something like mine, so to avoid unnecessary multiplication of terminology I have borrowed his term.

[6] I certainly do not claim that Kripke, Herzberger, Gupta, or anyone else would *endorse* this as a standard of success for his/her own theory.

[7] Kripke draws the strong/weak distinction nicely, identifying his own theory as weak rather than strong:

> The approach adopted here has presupposed the following version of Tarski's "Convention T", adapted to the three-valued approach: if "k" abbreviates a name of the sentence A, $T(k)$ is to be true, or false, respectively iff A is true, or false. This captures the intuition that $T(k)$ is to have the same truth-conditions as A itself; it follows that $T(k)$ suffers a truth-value gap if A does. An alternate intuition would assert that, if A is either false of undefined, then A is *not true* and $T(k)$ should be *false*, and its negation *true* (OTT, p. 715).

Graham Priest's theory, although less thoroughly elaborated than Kripke's, seems also to rely on the weak conception. From the Tarski biconditional $\ulcorner A$ is true iff $A\urcorner$ he deduces that

> A is true only (true and not false) iff $\ulcorner A$ is true$\urcorner$ is true only. A is paradoxical [i.e., both true and false] iff $\ulcorner A$ is true$\urcorner$ is paradoxical. A is false only iff $\ulcorner A$ is true$\urcorner$ is false only (LP, p. 238).

In "Presupposition, Implication, and Self-Reference", van Fraassen observes that one *could* say

> that when P is neither true nor false, then $T(P)$ does not have a truth-value either – as opposed to: then $T(P)$ is false. But we shall then have no way of formulating the assertion that a sentence is *not true* (p. 144).

Charles Parsons, commenting on the foregoing, remarks that

> it would be natural to say that if A is not true, then Ta is false (TLP, p. 383).

And Peter Woodruff, in "Logic and Truth-Value Gaps", proposes the following as a truth-table for the unary truth-operator T (the blank space indicates a gap),

p	Tp
t	t
	f
f	f

remarking only that the above "clearly reflects the intended interpretation" of Tp, the intended interpretation being "it is true that p" (LTG, p. 124).

Although the strong and weak conceptions are not usually explicitly distinguished, it is often possible to make out which conception underlies a given discussion. For example, one used often to hear that truth-value gap approaches, though more than a match for the "ordinary" Liar "This sentence is false", were powerless to deal with the "strengthened" Liar "This sentence is not true". On the weak conception of truth (and falsity), this is so, but on the strong, neither Liar submits to the truth-value gap resolution. Thus an insistence on the particular viciousness of the strengthened Liar

vis-à-vis the ordinary is often a tipoff that the weak conception is at work. (Note that in this paper the Liar will be "This sentence is not true"; when we talk about "This sentence is false" later on it will be called the "weakened" Liar, in deference to tradition.)

[8] This is assuming that negation is read weakly, so that the negation of a gap (glut) is another gap (glut). Interestingly, Kripke calls the Liar neither true nor false, and Priest calls it both true and false. It should be mentioned, though, that Priest sees himself as accepting, rather than attempting to resolve, the paradoxes.

[9] This is adapted from a formulation of George Myro's.

[10] A sequence is *increasing* if earlier entries are subsets, possibly improper, of later ones. If earlier entries are proper subsets of later ones, the sequence is *strictly* increasing.

[11] For more on inductive spaces, see "Grounding, Dependence, and Paradox".

[12] $\text{limsup}_{\beta<\alpha} L^{\beta}(K)$ is defined as $\cap_{\beta<\alpha} \cup_{\beta<\gamma<\alpha} L^{\gamma}(K)$, and referred to as the *superior limit* of $\langle L^{\beta}(K) | \beta < \alpha \rangle$. For future reference, $\text{liminf}_{\beta<\alpha} L^{\beta}(K)$, $\langle L^{\beta}(K) | \beta < \alpha \rangle$'s *inferior limit*, is defined as $\cup_{\beta<\alpha} \cap_{\beta<\gamma<\alpha} L^{\gamma}(K)$.

[13] It is important to realize that the results on antiinductive spaces do not essentially depend on any "special" features of the limit rule selected. For the only reasonable limit rules in this context are those relying on the inferior and superior limit operations described in note 12, and it can be shown that all such rules lead to essentially the same results. More exactly: Let $\langle K'_{\alpha} | \alpha \in OR \rangle$ be just like $\langle K_{\alpha} | \alpha \in OR \rangle$ except that K'_{λ} can be *either* $L(\text{limsup}_{\beta<\lambda} K'_{\beta})$ *or* $L(\text{liminf}_{\beta<\lambda} K'_{\beta})$. Then $\langle K'_{\alpha} | \alpha \in OR \rangle$ is a telescoping sequence, and for each α either $K'_{\alpha} = K_{\alpha}$ or K'_{α} lies between $K_{\alpha-1}$ and $K_{\alpha+1}$. This shows, in particular, that K's upper and lower closures $\overline{K}$ and $\underline{K}$ are invariant under reasonable perturbations of the limit rule.

[14] It must be emphasized that what follows is not reliable as a descriptive account of Kripke's theory of truth. For that, see Kripke's "Outline of A Theory of Truth", or my "Grounding, Dependence, and Paradox".

[15] Two remarks. First, observe that the models just defined are, as their name suggests, significantly more general than the sort familiar from the literature. In classical models a predicate's antiextension is the complement of its extension; in partial models (the kind Kripke considers) a predicate's antiextension is disjoint from its extension; in general models the relation between extension and antiextension is entirely unconstrained. The corresponding valuations will be arbitrary relations between sentences and truth-values (see below). These liberalizations allow us to consider without prejudice the connections, if any, between paradoxes, gaps, and gluts. Second, it will be convenient to treat expressions of the form Ta, where $I(a)$ is not a sentence, as ill-formed. If a semantical criterion for well-formedness seems objectionable, imagine that there is a syntactically distinguished class of sentence-names. The reader is requested to make necessary allowances.

[16] Note that a valuation satisfies (—.1)–(∀.1) and (—.2)–(∀.2) iff it is strong Kleene.

[17] Compare note 24 of "Outline of A Theory of Truth".

[18] See "Outline of A Theory of Truth", where the minimal and maximal intrinsic fixed points are singled out for special attention. Note also that the requirements of forcing and grounding (see Sections II and VIII), which could be pressed into service here, have their origins in Kripke's work.

[19] Actually, Kripke does briefly consider such an elaboration, namely "closing off" T's fixed-point interpretation by throwing the complement of its extension into its antiextension (the extension is left unchanged). See OTT, p. 715.

[20] Note carefully that there is a big difference between saying that ϕ is true, period, and saying that it is *uniquely* true, or true without being false. $T^{\ulcorner}\phi^{\urcorner}$ just says that ϕ is true, as it surely is if ϕ is both true and false.

[21] To guard against possible misunderstanding, I would like to make a couple of explanatory remarks. (1) It has been argued that even if ϕ is a gap or glut, $T^{\ulcorner}\phi^{\urcorner}$ is neither. But if I have "conceded that sentences can "in principle" be neither true nor false, why can't this also apply to some sentences of the form $T^{\ulcorner}\phi^{\urcorner}$?" And of course a similar query can be raised about true-and-false sentences. (Thanks to the editor for pointing this out to me.) My answer is that the argument is supposed to *show* that the conceded possibilities cannot extend to sentences of the form $T^{\ulcorner}\phi^{\urcorner}$. For if they did, what would the semantical status of ϕ be? Whatever answer you give, $T^{\ulcorner}\phi^{\urcorner}$ comes out either uniquely true or uniquely false. Perhaps the confusion arises because ϕ might itself be of the form $T^{\ulcorner}\theta^{\urcorner}$. If ϕ can be a gap (glut), then since ϕ is in this case $T^{\ulcorner}\theta^{\urcorner}$, $T^{\ulcorner}\theta^{\urcorner}$ can be a gap (glut). It is true that the argument temporarily assumes that ϕ is neither true nor false (true and false), but in this case the assumption is strictly *per absurdum*. If ϕ did have the envisaged status, then it would not; so it does not. (2) Someone might propose that since no sentence is, as a matter of fact, both true and false, there can be no harm in passing directly from the falsity of ϕ to that of $T^{\ulcorner}\phi^{\urcorner}$. For incidental reasons I am inclined to agree that no sentence is both true and false, but even so the objection seems unconvincing. First, the fact that no two-valued sentence exists, if it is a fact, is not something a theory can take for granted, but rather something it should try to account for. Procedures whose appropriateness *depends* on no true sentence being false have little to contribute to an explanation of how it comes *about* that no true sentence is false. Second, even if there are not, in the end, any sentences of this kind, it may be that on the way to establishing this one necessarily passes through stages where they are temporarily countenanced. Procedures whose appropriateness depends on no truth being false are liable to mishandle such stages.

[22] Kripke himself seems to have thought that the weak conception of truth was in some sense prior to the strong. He remarks that "the primacy of the first intuition [i.e., the weak conception] can be defended philosophically," presumably on the ground that "the alternate intuition [i.e., the strong conception] arises only after we have reflected on the process embodying the first intuition" (OTT, p. 715). I cannot see that this is so (although it is possible I have misunderstood). Admittedly the strong conception necessarily involves reflection, and admittedly *some* conception must guide the process on which we reflect, but why can't the latter conception be strong also? Of course, the reflective *elements* of the strong conception cannot come into play until the process has been concluded, but that does not seem to be problematic in itself. In the meantime, i.e., within the process, the strong conception dictates that we refrain from calling the truth-sentences of falsehoods false, in anticipation of the reflection to come. (See Sections VIII and IX below.)

[23] What follows is not meant to be an accurate guide to stability semantics as its authors – Hanz Herzberger and Anil Gupta – conceive it. For that, see Herzberger's "Notes on Naive Semantics" and "Naive Semantics and the Liar Paradox", and Gupta's "Truth and Paradox".

[24] In fairness, I very much doubt that Herzberger or Gupta would be impressed by the worries I have raised. They would probably accuse me of a confusion between truth, on the one hand, and stable truth, on the other (I believe Gupta has said that the English word "true" never means "stably true"). Truth, the quantity perpetually up

for reassessment, *is* supported; stable truth neither is nor ought to be. I think there is considerable justice in this accusation, but let me say a few things by way of response. First, even if truth is, in stability semantics, supported, it is not, in general, forced or grounded. But *something* is forced and grounded (or else my intuitions are very far off) and if not truth, or stable truth, then what? Second, what are the connections between truth, stable truth, and ideal assertability? If ideal assertability lines up with truth, then stable truth seems to be left out in the cold, without any discernible import for the human practices of evaluation and assertion. But if, as I suspect is the case with Gupta, ideal assertability lines up with stable truth, then ideal assertability is not supported. Yet intuitively, it seems to me, ideal assertability is supported, witness the uneasiness one feels about asserting the following: "Either the Liar is true or it isn't true". Third, even if Gupta is right that the meaning of "true" is given by a rule of revision rather than one of application, there remains the problem of saying which sentences are true and which aren't. But if "ordinary" truth and stable truth are different, then this is a problem stability semantics does not address. It tells us whether ϕ is stably true, and (if one accepts that all and only stable truths are assertable) it tells us whether "ϕ is true" may be asserted, but it does not seem to tell us whether ϕ is true. (Notice that it won't do to say that ϕ is true iff "ϕ is true" is assertable, because then ϕ is true iff "ϕ is true" is assertable iff "ϕ is true" is stably true iff ϕ is stably true, i.e., ϕ is true iff stably true after all.)

[25] An example is the sentence $G = T^{\ulcorner}G^{\urcorner} \vee T^{\ulcorner}{-}G^{\urcorner}$ (introduced in Section VII). Suppose towards a contradiction that G is in the language, and that ν is closed and forced. By supportedness (a consequence of forcing) $\langle G, f\rangle \in \nu \Rightarrow \langle T^{\ulcorner}G^{\urcorner}, f\rangle$, $\langle T^{\ulcorner}{-}G^{\urcorner}, f\rangle \in \nu \Rightarrow \langle G, t\rangle, \langle {-}G, t\rangle \notin \nu$. By closure, $\langle {-}G, t\rangle \notin \nu \Rightarrow \langle G, f\rangle \notin \nu$; so $\langle G, f\rangle$ cannot be in ν. On the other hand, by forcing $\langle G, t\rangle \in \nu \Rightarrow \langle G, t\rangle < \langle G, t\rangle$ or $\langle {-}G, t\rangle < \langle G, t\rangle$. Since the former is impossible, $\langle G, t\rangle \in \nu \Rightarrow \langle {-}G, t\rangle < \langle G, t\rangle \Rightarrow \langle {-}G, t\rangle \in \nu \Rightarrow \langle G, f\rangle \in \nu$ by l-supportedness. Since $\langle G, f\rangle \notin \nu$, $\langle G, t\rangle \notin \nu$ either. But then $\langle T^{\ulcorner}G^{\urcorner}, f\rangle \in \nu$ by s-closure, and since by the above $\langle {-}G, t\rangle \notin \nu$, s-closure puts $\langle T^{\ulcorner}{-}G^{\urcorner}, f\rangle$ into ν too. Since $\langle T^{\ulcorner}G^{\urcorner}, f\rangle$ and $\langle T^{\ulcorner}{-}G^{\urcorner}, f\rangle$ are in ν, $\langle G, f\rangle$ is in ν by l-closure, contradicting our recent conclusion that $\langle G, f\rangle$ was not in ν. Conclusion: if G is in the language, no valuation can be both closed and forced.

[26] Note that these are not quite the dependence relations of "Grounding, Dependence, and Paradox", i.e., the dependence relations associated with the operator J taking μ to the least logically closed extension of $\{\langle T^{\ulcorner}\phi^{\urcorner}, v\rangle \mid \langle \phi, v\rangle \in \mu\}$. But they are the dependence relations associated with the operator j described in note 27. Lawrence Davis's semantic trees (see his "An Alternate Formulation of Kripke's Theory of Truth"), we note for completeness, are essentially the dependence relations associated with the operator taking μ to $J_l(\mu) \cup J_s(\mu)$.

[27] Note that S_* is exactly the closure of S under rules (A.1)–($\forall$.1) and the first part of (T.1). If j is defined as the function taking μ to $J_l(\mu) \cup \{\langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \mid \langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \mu\} \cup \{T^{\ulcorner}\phi^{\urcorner}, t\rangle \mid \langle \phi, t\rangle \in \mu\}$, then for all j-sound S, S_* is the fixed point S generates under the operation of j.

[28] Lest anyone think there is something gratuitous about our use of Λ_* as starting point, let me suggest a different perspective on the sequence of stages. Think of (A.1)–($\forall$.1) and the first part of (T.1) as "positive" rules, the second part of (T.1) as "negative" (from the point of view of the theory of inductive definitions, that is exactly what they are). Then our procedure is essentially this: starting with the empty set, close under the positive rules (stage 0), apply the negative, close under the positive

again (stage 1), apply the negative again, close under the positive again (stage 2), and so on. (See Richter and Aczel, "Inductive Definitions and Reflecting Properties of Admissible Ordinals", p. 337, for a similar picture of nonmonotonic induction.)

[29] Over- and underevaluation should be understood here as either proper *or* improper. The intersection of a decreasing sequence of proper overevaluations need not be a strict overevaluation, but the intersection of a decreasing sequence of overevaluations does have to be an overevaluation.

[30] I owe this observation to Donald Davidson.

[31] George Myro tells a story which illustrates the point nicely. Philosopher A explains his theory, whereupon philosopher B promptly produces a counterexample. Philosopher A puzzles a moment, then responds: "You must not have understood the theory as I intended it, for I intended it not to *have* any counterexamples!" That we intend our intuitive semantical theory to be free of contradictions does not show that a theory which does involve contradictions is not our intuitive semantical theory.

[32] Thanks to Shaughan Lavine for pointing out the importance of a falsity predicate, and for ideas on how to define one.

[33] It follows from Proposition 18 that if M is classical, sentences are divided between the true, the false, and the truth-featureless. That sounds like a surprising result. Consider, for example, the truth-teller sentence K $(= T^{\ulcorner}K^{\urcorner})$. Call it true, and since it says it is true, it is; call it nontrue, and since it says it is *true*, it is false. Surely a sentence which is indifferently true or false is a genuine, uncontroversial, gap. Unfortunately for the objection, the preceding remarks do not establish that K is indifferently true or false. What they do suggest is that *if* it were a matter of indifference whether K was true or nontrue, *then* it would be a matter of indifference whether it was true or false. But of course it is not a matter of indifference whether K is true or nontrue. By the requirement of forcing, K's (i.e., $T^{\ulcorner}K^{\urcorner}$'s) truth requires the prior truth of the sentence whose truth it asserts. Since that sentence is K itself, its prior truth is out of the question. Thus there is nothing to make K true, and it is consequently *not* true. Since K is not true, any sentence saying it *is* true deserves to be false. And since K is itself among the sentences which say this, K deserves to be false. (It is important to see that this argument does *not* assume that whatever isn't true deserves to be false, but only that any truth-sentence of a nontruth deserves to be false.) The example helps bring out the "reflective" character of truth as presently conceived. It is not until we reflect on K's nontruth that the reason for its falsity emerges. (Note that K's dual K' $(= -T^{\ulcorner}-K'^{\urcorner})$, which says in effect "I am not false", is true.)

[34] Call a sentence *Kripke-true* (*-false*) if it is true in J's minimal fixed point Λ^*. If the ground model M is partial (in particular if it is classical), every Kripke truth (falsity) is true (false) in our sense. It suffices to show that $\nu = \{\langle \phi, t\rangle \mid \phi \text{ is true}\} \cup \{\langle \phi, f\rangle \mid \phi \text{ is false}\}$ is closed under (A.1)–(T.1). That ν is closed under (A.1)–($\forall$.1) and the first part of (T.1) follows from Proposition 16. It remains to show that $\langle \phi, f\rangle \in \nu \Rightarrow \langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \nu$. $\langle \phi, f\rangle \in \nu \Rightarrow \langle \phi, f\rangle \in \underline{\Omega}$ and $\langle \phi, f\rangle \in \overline{\Omega}$. By Proposition 17 ϕ is decidable, so if $\langle \phi, t\rangle$ were in $\overline{\Omega}$, ϕ would be both true and false, contradicting Proposition 18. So $\langle \phi, t\rangle \notin \overline{\Omega}$, whence $\langle T^{\ulcorner}\phi^{\urcorner}, f\rangle \in \underline{\Omega}$ and $T^{\ulcorner}\phi^{\urcorner}$ is false. This completes the proof. (Note that we wouldn't expect the Kripke truths (falsehoods) to be true (false) in our sense if M weren't partial, because for us the truth-sentences of gluts aren't false.)

[35] The following is not to be read as a neutral exposition of Kripke's account of paradox. The idea is Kripke's, but the formulation and emphases are not.

[36] Some caveat as note 35, reading "Herzberger and Gupta" for "Kripke".

[37] It follows from some work of Vann McGee's that not every non-paradox can be supplied with a support sequence of order-type $*\omega$. An example is $L_\omega \vee -L_\omega$, where L_ω is the recently defined ω-Liar.

[38] One further possible application of supportability deserves brief mention. Call a sentence **inherently true (false, untrue, unfalse)** iff it is capable of truth (falsity, untruth, unfalsity) but not of untruth (unfalsity, truth, falsity). Notice that inherent truth and Kripke's intrinsic truth are not the same (all tautologies are inherently true, but not all are intrinsically true); inherent truth has more in common with the property, well-known but I think unnamed, of being true in some fixed points and false in none. Furthermore, inherence and decidability are entirely independent: L is neither decidable nor inherently anything; "7 + 5 = 12" is decidably and inherently true; $L \vee -L$ is inherently true but not decidably true; and the truth-teller K ($= T^\ulcorner K^\urcorner$) is decidably false but not inherently false. Sentences with the same decidable, but different inherent, truth-features have different semantical flavours: for example, "Snow is black" and "I am true" are both decidably false, but only "Snow is black" is *inherently* false. Inherence can also be applied on the side of undecidability. Gupta suggests that fixed point semantics has difficulty explaining why when A says "Everything B says is true" and "Something B says isn't true", and B says "At most one thing A says is true", B seems to win the argument. But we can explain this by pointing out that the conjunction of A's statements is inherently false, whereas what B says is inherently true. And Gupta's refinements of the example submit to similar treatment. (Incidentally, it is not clear that a parallel move is not open to Kripke.) Problem: we know that there is a largest intrinsic, weakly balanced, valuation, but is there a largest inherent, strongly balanced, assessment?

BIBLIOGRAPHY

N. Belnap, 'Gupta's rule of revision theory of truth', *Journal of Philosophical Logic* **11** (1982), 103–116.

C. Chihara, 'The semantic paradoxes: a diagnostic investigation', *The Philosophical Review* **88** (1979), 590–618.

L. Davis, 'An alternate formulation of Kripke's theory of truth', *Journal of Philosophical Logic* **8** (1979), 289–296.

A. Gupta, 'Truth and paradox', *Journal of Philosophical Logic* **11** (1982), 1–60.

H. Herzberger, 'Notes on naive semantics', *Journal of Philosophical Logic* **11** (1982), 61–102.

H. Herzberger, 'Naive semantics and the liar paradox', *Journal of Philosophy* **79** (1982), 479–497.

S. Kripke, 'Outline of a theory of truth', *Journal of Philosophy* **72** (1975), 690–716.

V. McGee, 'Technical notes on three systems of naive semantics', typescript.

Y. N. Moschovakis, *Induction on Abstract Structures*, North-Holland, Amsterdam 1974.

C. Parsons, 'The Liar paradox', *Journal of Philosophical Logic* **3** (1974), 381–412.

T. Parsons, 'Assertion, denial and the Liar paradox', *Journal of Philosophical Logic* **13** (1984), pp 137–152.

G. Priest, 'The logic of paradox', *Journal of Philosophical Logic* **8** (1979), 219–241.

W. Richter and P. Aczel, 'Inductive definitions and reflecting properties of admissible

ordinals', in J. E. Fenstad and P. G. Hinman (eds.), *Generalized Recursion Theory*, North-Holland, Amsterdam, 1974.
B. van Fraassen, 'Presupposition, implication, and self-reference', *Journal of Philosophy* **65** (1968), 136–152.
P. Woodruff, 'Paradox, Truth, and Logic, Part I: Paradox and Truth', *Journal of Philosophical Logic* **13** (1984), pp. 213–232.
P. Woodruff, 'Logic and truth value gaps', in Lambert (ed.), *Philosophical Problems in Logic*, D. Reidel, Dordrecht, 1970.
S. Yablo, 'Grounding, dependence, and paradox', *Journal of Philosophical Logic* **11** (1982), 117–137.

Department of Philosophy, Moses Hall,
University of California at Berkeley,
Berkeley CA 94720,
U.S.A.