Graham Priest

Unstable Solutions to the Liar Paradox

1. Introduction

For about 80 years logicians have participated in a research programme called "solve the paradox". The immediate cause of the existence of the research programme was the proliferation of logical paradoxes around the turn of the century. The central assumption or "hard core" of the programme is the assumption that no contradiction is true, and hence that the reasonings which result in the contradictions must be fallacious. The aim has been to locate the fallacies and to articulate a theory which explains the data: the fallacious, yet highly plausible reasoning. Some time later, which might conventionally be dated at the publication of Ramsey's essay "The Foundations of Mathematics" [1926], the programme bifurcated into one for solving the set theoretic paradoxes and one for solving the semantic paradoxes. The strategy of divide and conquer is a familiar enough one, and often successful. Yet in this case it was a retrograde, or at least defeatist, step: the original aim of the founding fathers, such as Russell, to find a unified solution to the problem had to be given up. If one now considers the semantic branch of the program, it is difficult to avoid the conclusion that it has been somewhat less than successful. (I think that the same is true of the set theoretic branch. However in virtue of the general acceptance–at least by mathematicians–of the cumulative hierarchy, the case is much more difficult to make and I shall not address it here.)

There is certainly no generally accepted solution. The strategies or "heuristics" for solving the problem are but few. There is the "meaningless" strategy, the "neither true nor false strategy" and so on. Yet within this framework, purported solutions have multiplied in a way that makes the breeding habits of rabbits look like family planning. Show that one distinction does not work and a dozen appear in its place; show that a theory runs into trouble with a well-supported philosophical theory and a dozen patched-up versions appear to replace it. This is not the place to chart the historical details of this process, which are, in any case, widely known. However, one can see in this process what Lakatos has called a "degenerating research programme". Characteristic of this is that no essential progress is made towards solving the central problem. Rather, enormous time is spent trying to solve problems of equal or greater acuity created by the programme itself.[1] An extreme form of this is where the proposed problem solution does not really solve the problem at all but merely one of its manifestations. The original problem is then transferred, and appears in a different place. It may sometimes appear in a slightly different guise, whence to a cursory glance the proposed solution may appear more successful than it actually is.

The most recent instalment in this program is an approach to the semantic paradoxes, or rather family of approaches, provided by Anil Gupta [1982] and Hans Herzberger [1982].[2] Their construction is elegant and of clear structural interest, and it might appear that it is a "creative shift in the heuristic of the programme"; but I think that in fact it is merely another phase of the degenerating programme. The major part of this paper is an attempt to show this. Despite this fact, it does seem to me that, although it is incorrect, the idea points the way to a more adequate understanding of the semantic paradoxes. I will return to this in the final section of the paper.

2. The Construction

In order that the paper may be reasonably self contained I will start by outlining the Gupta/Herzberger (hereafter 'GH') construction and pointing out some of its salient features.

We take a first order language, L, with a predicate 'T', which is thought of as the truth predicate. Let M_0 be any first order interpretation for L. The domain of M_0, D, contains a subset, S, which is just the sentences (closed formulas) of L. The extension of T in M_0, U, is arbitrary. However it is simple and natural to let $S \supseteq U$.[3] We now define a transfinite class of structures $\{M_\alpha \mid \alpha \in On\}$ by recursion thus:

i) Given M_0, $M_{\alpha+1}$ is exactly the same as M_α except that the extension of T in $M_{\alpha+1}$ is exactly $\{\varphi \in S \mid M_\alpha \vdash \varphi\}$.

ii) For limit ordinals, the Gupta and Herzberger variants differ slightly.

Let $X_\gamma^+(U) = \{ \varphi \in S \mid \exists\beta<\gamma\ \forall\alpha(\beta<\alpha \text{ and } \alpha<\gamma \Rightarrow M_\alpha \vdash \varphi)\}$

$X_\gamma^-(U) = \{ \varphi \in S \mid \exists\beta<\gamma \ \forall\alpha(\beta<\alpha \text{ and } \alpha<\gamma \Rightarrow M_\alpha \vDash \neg\varphi)\}$

The sentences in $X_\gamma^+(U)$ are *locally stably true* at γ. Those in $X_\gamma^+(U)$ are *locally stably false* at γ. Now let λ be a limit ordinal. Then M_λ is the same as M_α for $\alpha < \lambda$ except that the extension of T in M_λ is:

a) $X_\gamma^+(U)$ (Herzberger)
or b) $X_\gamma^+(U) \cup (U - X_\lambda^-(U))$ (Gupta).[4]

The differences in construction matter not for our purposes.

Let $X_\infty^+(U) = \{\varphi \in S \mid \exists\beta\forall\alpha\geq\beta \ M_\alpha \vDash \varphi\}$
and $X_\infty^-(U) = \{\varphi \in S \mid \exists\beta\forall\alpha\geq\beta \ M_\alpha \vDash \neg\varphi\}$.

We will call the members of $X_\infty^+(U)$ *globally stably true* and those of $X_\infty^-(U)$ *globally stably false.* Let $X_\infty(U) = X_\infty^+(U) \cup X_\infty^-(U)$. The members of $X_\infty(U)$ are *globally stable* (relative to U). If φ is globally stably true with respect to all U we will call it *absolutely* stably true; if globally stably false with respect to all U, it is absolutely stably false. If it is either of these it is absolutely stable.[5] For fixed U, if φ is globally stable, we will call the least ordinal at which φ or $\neg\varphi$ enters the extension of T never to depart, its *stabilisation point.* Let

$$\Sigma(U) = \{\beta \mid \exists\varphi \in X_\infty(U), \beta \text{ is the stabilisation point of } \varphi\}.$$

And let $\sigma(U) = \bigcup\Sigma(U)$. We will call M_α *stabilised* iff $\alpha \geq \sigma(U)$. It is now easy enough to establish the following facts:

0) If M_α is stabilised then if $\varphi \in X_\infty^+(U)$, $M_\alpha \vDash \varphi$ and if $\varphi \in X_\infty^-(U)$ then $M_\alpha \vDash \neg\varphi$.

1) If $\vDash \varphi$ then φ is absolutely stably true.

2) For any U, $X_\infty^+(U)$ is closed under logical consequence, as then is the set of absolutely stably true sentences.

3) If $\varphi \in X_\infty(U)$ then $T\underline{\varphi} \equiv \varphi \in X_\infty^+(U)$ where $\underline{\varphi}$ is a name for φ, which, without loss of generality we may suppose L to contain. (Thus if φ is absolutely stable, its T-sentence is absolutely stably true.)

4) If $\varphi \notin X_\infty(U)$ it does not follow that $T\underline{\varphi} \equiv \varphi \notin X_\infty^+(U)$. However there is no guarantee that it is in $X_\infty^+(U)$.

5) If φ does not contain 'T' then for all α, β, $M_\alpha \vDash \varphi$ iff $M_\beta \vDash \varphi$ Hence $\varphi \in X_\infty^+(U)$ or $\varphi \in X_\infty^-(U)$.

3. Meaning and "Rules of Revision"

So much for the technical construction. Let us now turn to the question of whether it provides a philosophically satisfactory solution to the semantic paradoxes.

Given that a paradox is an argument with a contradictory conclusion, a necessary condition for a solution is that it locate a step in the argument which is fallacious. Take now a paradox such as the liar, which is Gupta and Herzberger's most favoured case. The argument goes essentially:

Let ψ be ' ψ is false' (1)

The T-scheme gives: ψ is true iff ψ is false. Hence ψ is both true and false.

The point at which the GH construction faults this reasoning is precisely in the application of the T-scheme to ψ. As we have seen, an instance of the T-scheme is not bound to hold in any particular model, let alone all stabilised models. In fact, given that we can find an M_0 in which (1) is true, the negation of the T-sentence for ψ turns out to be absolutely stably false.

Does this solve the paradox? Of course not. It is easy enough to flatly deny a step in the reasoning. It is also straightforward to produce a technically constructed model in which it fails: Tarski's original construction, which both Gupta and Herzberger reject as inadequate does that.[6] The problem is to find an adequate philosophical justification for the construction. This might be approached in a variety of ways; but what, in the end, they must all come down to is that the construction provides an adequate analysis of our conception of truth. (A construction that is openly offered in a revisionist fashion–which is the way the Tarski hierarchy has often been viewed–does not solve the problem. For the aim was to explain what was wrong with the *original* argument, not some revision thereof.) Or, to put it in a form more in line with modern philosophy, the construction must provide an analysis of the meaning of 'true'.

Gupta, at least, sees the matter in these terms too. He says:

> '...I am suggesting that underlying our use of 'true' there is not an application procedure but a revision procedure instead. When we learn the meaning of 'true' what we learn is a rule that enables us to *improve* on a proposed candidate for the extension of truth. It is the existence of such a rule, I wish to argue, that explains the characteristic features of the concept of truth. [1982], p.37

And again later:

> In intuitive terms the conception I have tried to defend is this. When we learn the meaning of 'true' we learn a rule that enables us to determine the extension of truth *provided* that we know the denotations and extensions of all the names, predicates, and function symbols in the language. [1982], p.54.

Thus the meaning of 'true' is the rule which takes us from the extension of 'T' in M_α to its extension in $M_{\alpha+1}$.

Let us then ask whether the construction does give an adequate account of the meaning of 'true'.[7] This question is faced with the difficulty that at present the whole question of how meanings are to be explained is itself a moot one. We have several different accounts of what a theory of meaning for a language should be like: Davidson's account of meaning; Montague semantics; Dummett's verificationism and so on. Unfortunately Gupta's proposal fits into none of them. None of them makes provision for "rules of revision" whatever, exactly, they are. Of course the idea that meanings are rules was commonly aired between the 1930's and 1950's (though the rules in question were normally rules of application). But the theory of meaning has long since passed beyond those tentative and piecemeal days. It is therefore somewhat disconcerting to find the suggestion that meanings are rules (albeit of a new kind) thrown off this casually. If we were to put it uncharitably we could say that since the GH account of the meaning of 'true' is at odds with all the best developed accounts of what a theory of meaning should be like, it is in trouble. More charitably (and I think more accurately) the point is that the GH construction puts a large spanner in the works of theories of meaning. We have here, therefore, an excellent example of a proposed problem solution posing a deep and acute problem purely of its own making.

However, let us not leave the problem there; for the acuity of the problem needs to be emphasized. The mere occupation of a field by a theory, or collection of theories, does not mean that these are right, uncriticizable, or to be taken for granted. Could the GH construction be incorporated in a fully fledged theory of meaning? Obviously it is foolish to deny the possibility of this. (Equally obviously, the onus is on Gupta and Herzberger to show that this is at least plausible, or claims to have solved the paradoxes are somewhat premature.) What would such a theory be like? I do not want to put words into other people's mouths. However, let us try to see what it could be like.

It would seem that a theory of meaning for a language must be an axiomatic theory which for every meaningful sentence of the language, S, has a theorem which spells out what the meaning of S is.[8] Exactly how it does this is a point of some substance. However there appears to be little alternative to Frege's observation that to give the meaning of a sentence is in *some* sense to spell out its truth conditions.[9] The simplest suggestion (*i.e.* Davidson's [1967]) is that, at least for languages which contain no indexical sentences, the meaning of S is spelt out by the instance of the T-scheme for S in a Tarski-type truth theory. This approach is certainly not open to Gupta and Herzberger. Their semantics are of a model theoretic truth-in-a-structure kind, rather than an absolute truth-definition kind. Nor is there any hope (as can be done in possible-world semantics) of nominating a particular structure (some M_α) and identifying truth (*simpliciter*) with truth in that structure. For (on Gupta's account) meaning is essentially *relational,* concerning the generation of one structure from another, rather than being a property of a single structure. Even worse, in unfortunate situations, there will be instances of the T-scheme which fail in all structures, as we have seen. Thus there is no hope of using the T-scheme to state (truly) meanings.[10]

How then to proceed? Somehow we must incorporate the idea of there being a plurality of structures at issue, and important relationships between them. A way to do this is suggested by possible-world semantics. In these, what is thought of as spelling out the meaning of S is the statement of the truth-in-a-possible world condition:

For any world (situation) w, S is true in w iff $\varphi(w)$.

In effect, every sentence is treated as indexical with respect to a world-context. (See, *e.g.* Montague [1970].)

Perhaps a theory of meaning using the GH construction could be based on the idea that the meaning of a sentence S is spelt out by a theorem of the form:

For any stage α, S is true at α iff $\varphi(\alpha)$.

We would have to get a great deal clearer about what the stages were stages of (refining our conception of truth?, approximating the absolute?). However I think we can leave this murky problem aside. For it is quite unlikely that such a theory could be an adequate theory of meaning.

The crucial question to ask is what we would be imputing to speakers of a language with such a theory of meaning. For a theory of meaning spells out what it is that speakers know when they understand a language. They do not, perhaps, have to know the whole theory. Knowledge of the content of sentences which spell out the meanings of sentences might suffice. Neither must we suppose that the speakers can explicitly formulate these claims. Indeed perhaps they may not even have a language in which this can be done. None the less, the meaning-giving sentences of the theory must express what, in some sense, is grasped by speakers of the language, or it is hardly an account of the meaning of *their* language. It follows that concepts used in stating the meaning-giving sentences must be attributed, at least implicitly, to speakers of the language which the theory of meaning is for.

Possible-world semantics are sometimes criticized on just these grounds: that they impute to speakers concepts, such as that of possible worlds (or at least, such set theoretic machinery as is necessary to construct their surrogates), which they do not necessarily have. The objection might be parried by arguing that someone who knows the meaning of a sentence knows not only how to use it in the actual situation but also how they would use it in different situations. In some sense, then, a language speaker must have a conception of possibilities different from the actual. Hence, possible-world semantics are not methodologically vicious. I do not wish to discuss the adequacy of this reply, for the important point here is that there is no similar reply open to a parallel objection addressed to the suggested theory of meaning based on the GH construction. This theory of meaning imputes to speakers not just the notion of possible worlds, but those of an arbitrary ordinal, of ordinal operations and transfinite induction (in the specification of φ). [11] It seems that there could be nothing in the behaviour of most language speakers which would justify this attribution, in which case the semantics cannot be an adequate account of meaning.

It may be that Gupta and Herzberger would wish to formulate their accounts of meaning in some other way. But however it is formulated I do not see how it could sidestep a similar objection. For the transfinite construction is the core of their proposal, and it would seem impossible for meaning-giving sentences in a theory of meaning based on the construction to avoid referring to it.

Before we leave the area of the theory of meaning, there is one further observation worth making. Both the absolute truth definition of the Tarski approach and the model-theoretic possible-world approach, provide us with a notion of truth, *simpliciter*. Now there is an important connection between truth and assertion. Basically it is that truth is the aim of assertion.[12] The truth is, generically, what we aim to assert when we assert. Thus the fact that there is no notion of truth, *simpliciter,* in the GH construction poses something of a problem. Is the class of sentences we aim at asserting time dependent; so that we can legitimately assert and deny the liar sentence alternately every n minutes? Surely not.[13] There must be a fixed class of sentences at which we aim. What is it? The set of sentences true at some M_α is far too arbitrary to be satisfactory. Similarly, those sentences globally stably true with respect to a particular U have an air of arbitrariness. The only satisfactory class is the class of absolutely stably true sentences. (The class of sentences which are not absolutely stably false will not do since this is inconsistent.) This, I suspect, would be Gupta and Herzberger's line; it seems the only reasonable one.

4. Strengthened Paradoxes

(i) *The Significance of These*

Considerations concerning the theory of meaning are not, I think, the major objections to the proposal that the GH construction solves the semantic paradoxes. The major one is its failure to resolve paradoxes of the "strengthened" variety.[14] I will explain this in subsequent sections. However I want first to say a few words to put the situation in its correct perspective. It might be thought harsh to criticize a novel proposal for failing to solve more contrived paradoxes. After all, the liar paradox is the paradigm problem, and if we sort that out properly, we can hope to get the more peripheral problems sorted out later.

This perspective of the significance of strengthened paradoxes is the exact opposite of the correct one. To see this, look at the liar paradox as follows. We start with a set of sentences; we can call them *bona fide* truths. These are the sentences that are genuinely assertible. On most conceptions these will coincide with the true sentences (though in a many-valued logic they might coincide with the ones of designated value, and so on). Those that are left over, we will call "the Rest". The essence of the liar paradox is a particular twisted construction which forces a certain sentence, if it is in the *bona fide* truths, to be in the Rest (too); and conversely, if it is the Rest, it is in the *bona fide* truths. Since it can't play for both teams at once, the problem is posed.

Now the pristine liar "This sentence is false" is only a manifestation of the problem arrived at by taking the Rest to be the False (and the *bona fide* truths to be the True). In this particular case, then, we can of course get out of the problem by insisting that the False is only a proper part of the Rest. This opens up a gap in which the liar sentence can conveniently lie. But this solves the problem only by showing that it was inadequately posed. For if the False is only properly contained in the Rest, then the pristine liar is not the correct formulation of the problem. What strengthened liar paradoxes, such as "This sentence is false or neither true nor false", do, is to remind us of this fact: If we ever try to get out of the problem by taking a category which is *not* the Rest, we can pose the original problem by describing the Rest in some other way.

To summarise: the basic liar problem is that posed by a construction which destroys the division between assertible sentences and the Rest. As such it is the strengthened liar paradox which reflects the central problem; the ordinary liar paradox is but a special case. Thus a proposed solution to the liar paradox does not solve the essential problem if it leaves the strengthened liar paradox wide open. Such "solutions" are excellent examples of solutions which appear to solve the problem, but actually merely succeed in transferring it to another place or guise. (See sect.1.) Since most proposed solutions to the liar paradox fall foul of some version of the strengthened liar, we could say that such "solutions" are only meta-stable and that the strengthening construction is sufficient to destablise them. This is certainly true of the GH construction as we will now see.

(ii) *The Strengthened Liar: an informal account*

In the particular case at hand, the *bona fide* truths, the assertible ones are, as we have seen, the absolutely stably true ones (or just 'stably true' if no confusion can result). The strengthened liar therefore takes the form:

This sentence is not stably true.

If this sentence is stably true, it is true and hence not stably true. Thus, the sentence is not stably true and we have established this as a logical truth. Hence not only is it true, but it is stably true. (Since all logical truths are stably true sect.2 Fact 1.) Contradiction. I will make this argument more precise in the next part. But first let us examine what Gupta has to say about it. (Herzberger does not mention it.) He says:

> In response to this [*i.e.* the strengthened liar] I observe that the notion of "stable truth" may be viewed in three ways. First, as belonging to the metalanguage. This is the way we have used it above. We have used it in the metalanguage to give an account of the concept of truth in the object language L. This, it seems to me, does not in any way vitiate our account of the concept of truth. Further, when the notion is viewed this way the paradox does not arise. Second, the notion may be viewed as belonging to L itself but under the condition that L has sufficiently weak syntax as far as the predicate that expresses the notion of stable truth is concerned....
>
> Under such conditions we can envisage formulating the entire theory of truth given above in the language L itself. (Of course we will need other notions as well, and the technical details will be messy, if not overwhelming.) The paradox still does not arise. Third, the notion may be viewed as belonging to L when L does not meet the condition of sufficiently weak syntax. Now the paradox is present for the concept "stably true in L". But we must ask how is the concept "stably true in L" added to L? It must be added, it would appear, via a rule of revision. But then can we not give an account of the new paradox parallel to that we gave of the old? [1982], pp.55-6.

Gupta offers three possible responses. Let us take these in turn:

a) 'Stable truth' is part of the metalanguage but not the object language. This merely resurrects the levels of language notion of Tarski. I find it rather sad that Gupta, after so much ingenuity, falls back on this tired old distinction. If this is necessary to get us out of trouble, we might just as well have stuck with the original Tarski construction. Of course this construction will no do. Its inadequacies have been pointed out by many, including Gupta himself,[15] and it is a simple matter to transfer these arguments to the new situation.

b) 'Stable truth' belongs to **L** itself, but the self referential machinery necessary for paradox is absent from **L**. This line is no more promising. Gupta shows that a theory can consistently contain all instances of the T-scheme provided certain syntactic machinery is missing.[16] We can avoid the extended liar paradox in a similar way. However, again, if this is an adequate way out of paradox, we might as well have taken it in the first place. But it is not, as I am sure Gupta realises. No one doubts that virtually anything can be avoided if one weakens the expressive power of the language sufficiently. This is beside the point. For the problem arises with our ordinary concepts/language and our ordinary means of expression. It was the correct semantic analysis of these that is in question and not that of some castrated language.

c) 'Stable truth' is part of **L**, but its semantics is given by a "rule of revision". What *exactly* Gupta has in mind here is not clear. However, it would seem that the suggestion is somewhat disingenuous. What is in question here is precisely a language which can express its own semantic concepts and in which we can formulate Gupta's "entire theory of truth...in the language **L** itself". But then the term "stable truth" does not have to be added to the language. It is a notion *defined* on the basis of the given vocabulary (and defined moreover without the use of 'T') in just the way that Gupta and Herzberger show. Whatever else can be "added" to the language by a "rule of revision", the semantics is already inconsistent.

The fact that these rather inadequate remarks are thrown in, almost as an afterthought to the paper, suggest that Gupta shares the inverted perspective of the significance of strengthened paradoxes that I criticized in the previous part.

(iii) *The Strengthened Liar: a formal account*

In virtue of some of the slippery points involved in the issue of strengthened paradoxes, it is desirable to make the criticism of the previous section more rigorous and precise. To this end a small theorem is useful.

Let L be a first order language which contains the language of first order arithmetic and, in addition, the truth predicate 'T'. Let $\{M_\alpha \mid \alpha \in On\}$ be a GH interpretation for L such that M_0 (and thus all the M_α) are extensions of the standard model of arithmetic (or at least are models of Peano Arithmetic). Let M_β be any stabilsed model, and suppose that the set of absolutely stably true formulas of L with respect to M_0 is defined in M_β by a formula of L with one free variable ST(x), *i.e.* φ is absolutely stably true iff $M_\beta \vDash ST(\underline{\varphi}.)$

Now let M be any model which is the same as M_0 except perhaps for the extension of 'T'. Since M extends the standard model of arithmetic, we can code up formulas in the usual way and apply the diagonal lemma[17] to find a formula φ (independent of M) such that

$$M \vDash \psi \equiv \neg ST(\underline{\psi}) \qquad (1)$$

ψ is, of course, just the strengthened liar sentence.

Theorem

ψ is not absolutely stable.

Proof

Suppose ψ is absolutely stable.Then

$M_\beta \vDash T(\underline{\psi}) \equiv \psi$ (Section 2, Fact 3).

i.e. $M_\beta \vDash T(\underline{\psi}) \equiv \neg ST(\underline{\psi})$ by (1).

But $M_\beta \vDash ST(\underline{\psi}) \supset \psi$ since ST defines the set of stably true sentences and M_β is stabilised.

Hence $M_\beta \vDash ST(\underline{\psi}) \supset \neg ST(\underline{\psi})$

i.e. $M_\beta \vDash \neg ST(\underline{\psi})$

i.e. $M_\beta \vDash \psi$ by (1).

So since ψ is absolutely stable and true in some stablised model, it is absolutely stably true,

i.e. $M_\beta \vDash ST(\underline{\psi})$. Contradiction.

Let us, for the sake of interest, record a few corollaries.

Corollary 1

The set of absolutely stably true formulas of **L** with respect to $\mathbf{M_0}$ is not arithmetic.

Proof

If it were, there would be a purely arithmetic formula of one free variable **ST(x)** which would define the set of stably true formulas in every extension of the standard model of arithmetic. Since its extension does not vary, **ST** $(\underline{\psi})$ is absolutely stable, as, therefore is $\neg$**ST**$(\underline{\psi})$ which contradicts the theorem.

Corollary 2

The set of globally stably true formulas of **L** with respect to $\mathbf{M_0}$ is not arithmetic.

Proof

Simply rework the *whole* proof of corollary 1 with 'globally' replacing 'absolutely'.

Corollary 3

The set of absolutely (or globally) stable formulas of **L** with respect to $\mathbf{M_0}$ is not arithmetic.

Proof

The proof is essentially as for absolutely (globally) stable truths. The major difference is that we suppose **Stab(x)** to define the set of stable formulas and then let:

$$\mathbf{M} \vDash \psi \equiv \neg(\mathbf{Stab}\ (\underline{\psi}) \wedge \mathbf{T}(\underline{\psi}))$$

The other modifications are relatively minor.

Let us now return to the philosophical import of the theorem. For it is sufficient to sink the GH construction as a solution to the paradoxes. It forces a dilemma: The theorem holds for *all* (interpreted) languages which contain the language of arithmetic (with its correct interpretation). Now take for **L** the language used by Gupta and Herzberger themselves. This contains the language of set theory and we may certainly therefore take it to contain the language of arithmetic. Gupta and Herzberger do not tell us what the interpretation of their language is supposed to be, yet clearly the arithmetic language must be given the correct interpretation and, on pain of self-refutation, the interpretation must be of the form $\{\mathbf{M}_\alpha \mid \alpha \in \mathbf{On}\}$ where this is a

GH hierarchy. Now consider the predicate of this language "absolutely stably true". Either this does not refer sensibly to the absolutely stably true sentences, or if it does we are faced with an equal absurdity. Take the first horn of the dilemma. In this case the predicate defines the set of stably true sentences in no stablised model. Surely, that it should do this is a minimum condition necessary for us to use the phrase sensibly. We may not care what its extension is at the vicissitudes of lower models. But we want the phrase to mean the right thing at *some* sensible model. In other words, on this horn of the dilemma all the facts amassed by Gupta and Herzberger concerning absolutely stable sentences don't mean what they take them to mean. This is surely absurd.

The other horn of the dilemma is that the predicate does mean what it says, at least in some stablised model M_β. But in that case, the theorem shows that ψ is not stable and *a fortiori* not stably true. But in the interpreted language we are using, this is equivalent to ψ (by 1). Thus we have proved, and are therefore committed to asserting something unstable–which is "as true as false". This contradicts the conclusions about assertion at which we arrived at the end of sect.3.

Either horn of the dilemma is unpleasant, but we can turn the screws even harder. For after all, we *know* how "stably true" was defined in the use-language:

x is absolutely stably true iff for all M' which differ from M_0 in at most the extension of 'T', $\exists\alpha\in On\forall\beta\geq\alpha\ M'_\beta\vDash x$.

Call the whole *definiens* (2). "Stably true" is thus defined in L *without* using 'T'. Hence its extension is the same in any structure which differs from M_0 in at most the extension of 'T'. Thus for any φ, ' φ is (not) stably true' is stable. Hence ψ is stable (by (1)). But the theorem proves that ψ is not stable. Hence the whole construction is inconsistent.

It seems to me that the only possible way out of these problems–a fairly desperate one–is to deny that our use-language can be identified with any language of the form of L. One might argue, for example, as follows. Consider the formula (2) of L. It does not define in M_0 (or any other M_α–it doesn't matter since its extension is constant) the set of stably true sentences. What (2) defines in M_0 is the set of sentences stably true *with respect to the ordinals of* $\mathbf{M}_0$, **On** $^{\mathbf{M}_0}$ (assuming that '⊧' receives its correct interpretation in $\mathbf{M}_0$). Now since $\mathbf{M}_0$ is a set, **On** $^{\mathbf{M}_0}$ cannot contain all the ordinals. In fact, it can contain only those ordinals α such that $\alpha<\lambda$ for some λ. But then the extension of (2) in $\mathbf{M}_0$ is just the set of formulas that are *locally* (absolutely) stably true at λ. This will, in general, differ from the set of globally (absolutely) stably true formulas. It follows that when *we* refer to the stably true formulas of **L** we cannot be using **L** itself. We must be using a different, in fact stronger language (one whose quantifiers range over all, or at least more, ordinals). Thus **L** can at best be part of our language. Our language itself must be conceived of as a hierarchy of languages each of which is of the form of **L** and each of which is adequate to express the semantics of a lower one. Hence we are off up the Tarski hierarchy of metalanguages again, not, this time, with respect to Tarski semantics, but with respect to GH semantics. Nothing has been gained: we can simply turn Gupta's own arguments (and all the others against the hierarchies of language approach) against himself.

(iv) *The Inevitability of Semantic Ascent*

Moreover, semantic ascent of this kind is no accident. The aim of the problem is to solve the semantic paradoxes. These appear to occur when a language can express its own semantic notions. Thus the aim requires us to give a consistent semantical theory that can handle the semantics of the theory itself.[18] But such is a classical chimera. For to give an adequate account of the semantics of a theory we require at least the following. First we need to spell out an interpretation of the language in question. This may be an absolute truth definition, a GH construction, or whatever. We then demand that the theory be proved sound with respect to this interpretation. (We may also demand a proof of completeness with this notion suitably formulated. However, a proof of soundness is a minimum necessary condition for claiming to have given a suitable semantics for the theory.) But classically, soundness implies consistency. Hence giving a semantics for the theory entails proving consistency. Now if all this can be done in the theory itself, it follows that the theory can establish its own consistency. But provided that the theory is "sufficiently strong" and based on classical logic, we know that no consistent theory can prove the canonical assertion of its own consistency. This is Goedel's second incompleteness theorem, and the failure of the GH construction but a corollary of this.

What the second Goedel incompleteness theorem shows is that, classically, consistency can be maintained only by giving the semantics of a theory in a different theory. Thus any (consistent) theory must fail to be capable of giving its own semantics either by the requisite notions failing to be expressible in the language of the theory, or by requisite principles about them failing to be provable in the theory. The theory must therefore be either expressively incomplete or proof-theoretically incomplete.[19]

To summarize: incompleteness is the price paid for consistency. All the "solutions" to the semantic paradoxes ring the changes on this theme, one way or another. In particular, when our proof procedure is naive, so that there are no prior axiomatic constraints on provability, we must have expressive failure, *i.e.* we must, to be consistent, consider ourselves to be using a meta-language. This is why this idea comes up again (Tarski) and again (Kripke) and again (Gupta and Herzberger). Each of these solutions sought to improve on the former, but each falls to what is merely a new way of expressing the same point. The solutions are therefore inherently unstable and a fundamental cause of the degeneration of the paradox-solution research programme is exposed.

5. In Praise of Inconsistency

We have seen that the programme of solving the paradoxes is doomed to failure. Yet we still need a coherent approach to the paradoxes, and we still need to understand how a language such as English can handle its own semantic notions. How is this to be done? We will find a way suggested if we look at a third criticism of the GH construction. This one deals with the heuristic of the construction.

In his paper [1982a] Herzberger spells out some of the heuristic ideas behind his technical construction. In particular, Herzberger invites us to consider how, by simple steps of reasoning we obtain the flip-flop pattern associated with the liar paradox. Let us spell this out in slightly more detail. Suppose that a is a name of the sentence 'a is not true'. Then a is either true or not. Without loss of generality suppose it is the former. So

a is true

i.e. 'a is not true' is true

i.e. a is not true.

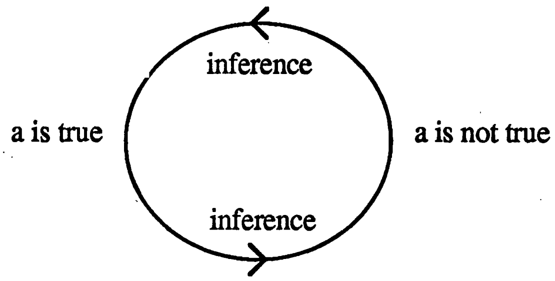
Call this progression **θ**. This chain of reasoning is supposed to be modeled in some sense by the construction which takes us from $\mathbf{M}_\alpha$ to $\mathbf{M}_{\alpha+1}$, and specifically the change in the extension of **T**. Thus Herzberger says:

> I believe that this kind of construction does so far incorporate the "ordinary rules" [*i.e.* of inference] that Wittgenstein remarked upon;...There is indeed an inconsistency between the valuations at one stage and those at another...I offer it as a reconstruction of what some people have felt to be the inconsistency of natural language.... [1982a], p.487-8.

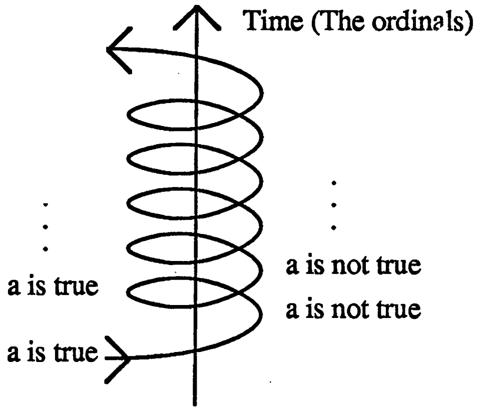
But is it? The process of someone making inferences in accordance with **θ** is a process in time, and though Herzberger does not say that the construction through the ordinals is supposed to model a temporal progression, the language he uses makes it difficult not to think of it in those terms (*e.g.* "On this picture, our language has an inner *dynamics* of a highly regular sort, based on a *process* of *progressive* semantic evaluation" [1982a], p.492. My italics.) And if we think of the progression as temporal, the idea of the extension of the truth predicate changing over time becomes quite a tempting one. After all, novel arguments do force us to revise what we take to be true.

However, the illusion of temporality is a spurious one. Though someone who reasons through **θ** may make a temporal progression, the progression **θ** itself is not a temporal one, but a progression of logical support, which is timeless. Once one grasps this then the thought that when we arrive at the end of **θ** we must change our interpretation to bring it into line with our conclusion is not at all enticing. Indeed, such a change, changing as it does the truth of the premise on which the conclusion was based, undercuts the very rationale for making that change. What the progression **θ** shows us is that, far from anything changing, the truth of 'a is true' commits us to the view that it is not true too *at the same time.*

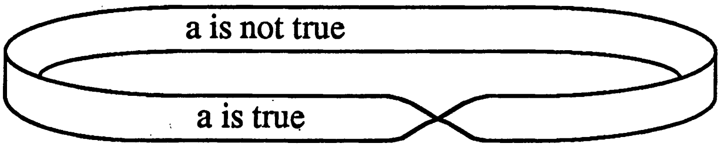
In fact, although Herzberger claims that in "Naive semantics [*i.e. his* semantics], the paradoxes will not be made to disappear" ([1982a], p.480), this is precisely what they do. And they are made to disappear by enforcing a misplaced temporal metaphor,[20] to try to make the talk of change natural. Thus the natural circular image of the paradox:



is changed to a spiral one:



(See Herzberger [1982a], pp.483-4.) But the temporality is out of place. The picture of the circle is much more appropriate than the spiral. Indeed, if a picture is desired, the best one is that of a Moebius strip:



To summarise: natural reasoning itself forces us to the conclusion that both the liar sentence and its negation are true; that it is both true and not true. It is not, therefore, the extension of the truth predicate that needs revising (whatever, in the end, this is supposed to mean) but our incorrect rejection of this fact. This outcome is perhaps a strange one. But there is little that is truly novel in philosophy, and we can find travellers that have been here before us. One such is Hegel. Hegel claimed that any two contradictory categories, A, A', have an "interpenetrating" boundary. That is, that the concepts (or perhaps better the analytic principles characterising them) force us to recognise the existence of something *per impossibile* in both categories; indeed, such that the thing's being in A is exactly its being in A'. Hence we are forced to recognise a novel category, the supersession (or synthesis) of A and A', which transcends the distinction. (It is the "being of each in the other".) Whatever we are to make of this claim for all categories, it is certainly true for the pair true/false. For the liar paradox forces us to the conclusion that something is both. Indeed, the liar sentence's being false is exactly its being true. Thus we must accept the category of *dialetheia* –something both true and false. [21]

Moreover, Hegel very astutely realised what would happen if one tried to consistentise the situation: the result would be an instability that would ultimately be futile (the only resolution of the situation being to recognise the transcendent category). Bearing in mind that Gupa and Herzberger have projected the oscillation into a realm that Hegel could never have guessed at (the transfinite) let us allow Hegel to speak for himself:

> If we let somewhat and another, the elements of determinate Being, fall asunder, the result is that some becomes other, and this other is itself a somewhat, which then as such changes likewise, and so on *ad infinitum*. This result seems to superficial reflection something very grand, the grandest possible...[However the] progression to infinity never gets further than a statement of the contradiction involved in the finite, viz. that it is somewhat as well as somewhat else. It sets up with endless iteration the alternation between these two terms, each of which calls upon the other...[So] such a progression to infinity is not the real infinite. That consists in being at home with itself in its other, or, if enunciated as a process, in coming to itself in its other. Much depends on rightly apprehending the notion of infinity, and not stopping short at the wrong infinity of endless progression.
>
> Hegel [1830] sect.94. (Quotation rearranged.)

As Hegel insists, we must recognise dialetheias. I do not deny that this raises important philosophical issues; but it is a necessary first step in an adequate approach to the semantic paradoxes and to semantically closed languages. I shall not try to explain this in detail here. Much of it is already in the literature. I will, however, briefly explain its salient features and its relation to some of the points I have discussed.

First, the idea that a sentence may be both true and false (or that both a sentence and its negation may be true) must be built into a coherent semantics (*e.g.* as in Priest [1979], [1980]). We can then take the paradoxical arguments to be what they appear, *prima facie,* to be, viz. sound arguments with contradictory conclusions. In particular, it is possible to produce inconsistent but non-trivial theories, such as set theory, based on these semantics (*e.g.* Routley [1977], Brady [1986]). Moreover it is possible to produce a semantically closed theory[22] *e.g.* an axiomatic theory which can prove all instances of its T-scheme. (See Priest and Crosthwaite [198+].)

A few further points are worth noting in the present context. First the semantics can be fitted into any orthodox theory of meaning (for example a Davidsonian one; see Priest and Crosthwaite [198+]), which can therefore be taken as providing an analysis of the meaning of 'is true'. Hence we have a satisfactory account of our naive conception of truth. Moreover, the T-scheme is, what it appears to be, an analytic principle governing the concept of truth, and not a "hasty generalization".[23] Second, the standard connection between (absolute) truth and assertibility is preserved. Third, any version of strengthened paradoxes one cares to formulate (if indeed there is any way one can sensibly distinguish them from the ordinary variety) can be handled in the same way as the ordinary variety, viz. left to stand. Fourth, the flip-flop behaviour of the sentence 'This sentence is not true'. (Suppose it is true; then it is not true; then it is true...), which is cited by Gupta and Herzberger in favour of their account, is explained in the best possible way: it is a valid *sorites*. Finally, this approach provides once again the possibility of a uniform treatment of the logical paradoxes.[24]

This approach to semantic closure clearly exhibits highly desirable features. Of course, it goes without saying that the logic generated by the dialetheic semantics is not classical. For otherwise the contradictions involved would trivialise the issue. In this context it is worth saying a final word about Herzberger's attitude to classical logic. Herzberger defends his use of classical logic partly on the ground that it is classical logic which in fact produces the paradoxes:

> [My construction] uses only ordinary models and classical two-valued valuations. This seems appropriate, inasmuch as it is reasoning in accordance with classical logic which in the first instance gives rise to the semantic paradoxes. [1982], p.61.

This is incorrect. It is naive reasoning (concerning the concept of truth) which generates the paradoxes. The liar paradox was known 2,000 years before "classical" model theory; and intuitionistic (and relevant) logic equally produce the contradiction. Of course it could be that classical logic is an adequate formalization of our naive reasoning procedures. But this certainly cannot be assumed without further ado. It is slightly disconcerting to find that many logicians, including apparently Herzberger, have forgotten that classical logic is just a *theory* of what naive reasoning is, with both strengths and weaknesses. (It is clear that the logical paradoxes are an Achilles' heel of classical logic, one that will, in the end, I think, take it in the same direction as Achilles.) For example Herzberger says:

> All standard [semantic] schemes are weaker than the classical valuation scheme, and consequently no one of them seems to be altogether free from intuitive wrinkles. [1982], p.92.

The remark is a casual one made in the context of discussing valuation schemes within Kripke's theory of truth. Nonetheless, its implication is clearly that deviation from classical logic is *ipso facto* a defect ("wrinkle"), and, correlatively, that classical logic has no "wrinkles". This is, of course, not the case. For one thing, it is the very strength of classical logic which has been under attack from various directions throughout this century (Brouwer, C.I. Lewis, Anderson and Belnap).

In a time of normal science (to use the language of T.S. Kuhn) the dominant theory is so much taken for granted that its problematic nature is suppressed into the "collective subconscious" of the scientific community. It is time that the problematic nature of classical logic is firmly brought back into the "conscious". The logical paradoxes are just the thing to do this.[25]

6. Appendix

The preceding sections of this essay were written in 1982. Since they were written some other papers using constructions related to the GH construction have appeared. In this appendix I want to comment briefly on one of these, Yablo [1985].[26] (Subsequent page references are to this.) There are many of the philosophical comments in Yablo's paper that I quite agree with. There are also a number of things that I disagree with.[27] However, I will, in this appendix, discuss only those aspects of the paper which relate directly to points made earlier in this essay.

An important difference between Yablo's construction and the GH construction, is that whilst the latter works with classical two-valued evaluations the former works with four-valued evaluations, according to which a sentence may be assigned true (t), false (f), both, or neither. In virtue of the fact that dialetheias are explicitly countenanced, it might be thought that I should have no quarrel with the construction. However, Yablo argues that supposing there to be dialetheias does not solve the problem of strengthened paradoxes (p 302), and that a construction like his is called for. I will take issue with both these claims.

Let us take the first point first. The argument is premised on the claim that truth is 'strong', in the sense that if φ is true (and possibly false as well) then $\mathbf{T}\underline{\varphi}$ is true and true only, and if φ is not true (and possibly not false either) $\mathbf{T}\underline{\varphi}$ is false and false only. In other words, the truth predicate is always classically valued. Now consider the extended liar:

(ψ) ψ is not true

ψ can have neither of the classical truth values for the usual reasons. Moreover it can have neither of the other values since the truth predicate is always classically valued. Thus ψ can have no consistent value.

Now, first, I would take issue with the claim that truth is strong. If φ is true, then I certainly agree that $\mathbf{T}\underline{\varphi}$ is true, by the T-scheme; and if φ is not true, I agree that $\mathbf{T}\underline{\varphi}$ is false. But I see no reason to suppose that $\mathbf{T}\underline{\varphi}$ *cannot* be false (as well as true) if φ is. I (now) think that it may or may not be false, depending on φ. (The reasons are explained in ch. 4 of Priest [198+].) Yablo argues for his position; but his argument seems to me not to be cogent. It goes thus (p 301):

'Pa' is true iff **a** has the property **P**.
'Pa' is false iff **a** does not have the property **P**.
Hence, taking **T** for **P** and φ for **a**:
' $\mathbf{T}\underline{\varphi}$' is true iff φ is true.
' $\mathbf{T}\underline{\varphi}$' is false iff φ is not true.
It follows that truth is strong.

But how is it supposed to follow that $\mathbf{T}\underline{\varphi}$ cannot be both true and false? This would follow only if it were impossible for φ to be both true and not true, in other words, if the truth predicate behaves consistently. But this is exactly what we are supposed to be showing. To bring out the question-begging nature of the argument, note that if it were right it would prove not only that **T** is a classical predicate, but that every predicate is classical. It would therefore under-cut the whole rationale of four-valued semantics.

But even supposing that truth is strong, does the argument against the dialetheic account of extended paradoxes work? No. It shows only that ψ can be given no *consistent* evaluation. But the *point* of dialetheism is to allow precisely for this. I claim that ψ is both true and false. Assuming truth to be strong, it follows that ψ is both true and not false. This is, indeed, a contradiction. But equally, this is exactly what we should expect. As I argued in section 4, the liar paradox is a construction which violates all semantic boundaries–those of four-valued semantics included. Dialetheism is designed precisely to cope with this situation.

Even though the objection against a dialetheic account of the paradoxes is incorrect, it brings home an important point. One cannot adopt a dialetheic position on paradoxes without being inconsistent oneself. Yablo (and some other writers)[28] are prepared to countenance a half-hearted dialetheic view according to which a sentence may be both true and false, but they balk at describing situations inconsistently themselves. But this cannot even seem reasonable unless we enforce a rigid distinction between the (inconsistent) object theory and the (consistent) meta-theory. And if this is a reasonable move then we might as well invoke the object/metalanguage distinction to get rid of an inconsistent object theory in the first place. But this is not reasonable, as I discussed in section 4 above. A red-blooded dialetheism is the only viable option.

Let us turn now to Yablo's own construction. How, exactly, this is supposed to solve the paradox is not, surprisingly enough, spelled out.[29] Still, the construction is supposed to provide an analysis of the notion of truth, and to take account of the paradoxes, in some sense. The construction is a good deal more complex than the GH construction, but it is similar in the following ways. A semantics is defined for a language with its own truth predicate. This is done by defining a hierarchy of interpretations by transfinite induction on the ordinals. The interpretation of the truth predicate (which is the only thing that varies in the hierarchy) at level $\alpha + 1$ is produced by a uniform operation on its extension at level α. During the process of ascent through the ordinals, a certain stability emerges, and this can be used to define a set of absolute categories for formulas of the language.

The exact details of the construction need not concern us here. All we need note is that the construction provides a *pair* of "equally good" evaluations, $\underline{\Omega}$ and $\overline{\Omega}$ (p 331) representing the limit situation, in terms of which we may define φ to be:

true if *both* $\underline{\Omega}$ *and* $\overline{\Omega}$ assign t to φ;
untrue if *neither* $\underline{\Omega}$ *nor* $\overline{\Omega}$ assigns t to φ;
false if *both* $\underline{\Omega}$ *and* $\overline{\Omega}$ assign f to φ;
unfalse if *neither* $\underline{\Omega}$ *nor* $\overline{\Omega}$ assigns f to φ.

These categories are neither mutually exclusive (except for the first pair and the second) nor exhaustive. We note also that (p 332):

T$\underline{\varphi}$ is true iff φ is true;
T$\underline{\varphi}$ is false iff φ is untrue.

With these details under our belts we can see that exactly the same objections apply to this approach as apply to the GH approach. In particular, the same features of the degenerating research programme are present. First, the account of truth, depending as it does on transfinite ordinals and induction, is susceptible to the argument based on the theory of meaning that I used in section 3. Yablo distinguishes between the psychological problem of determining how people actually operate with the notion of truth and the descriptive problem of characterising truth (p 229-300). He might therefore object to this argument on the ground that to criticise the characterisation in this way is untoward psychologism. However, it is agreed that the descriptive problem is essentially one of giving an account of *meaning* (fn. 5); and though meaning cannot be defined in psychological terms, there are certainly psychological constraints on what can count as an adequate theory of meaning.[30]

More crucially, exactly the same situation concerning the strengthened liar paradox again arises. The Rest, in this construction, is just the set of sentences that are not true. Yablo calls these *non-true* to distinguish them from the untruths (p 331). Now, consider the sentence:

(ψ) ψ is non-true.

By the usual argument ψ is true iff it is non-true. Hence it is both true and not true. Note that the conclusion is not that ψ is true and false, which would be alright. Note also that Yablo's metalanguage is quite classical. Thus, ψ is either true or it is not, and not both.

As in section 4, we can work this argument into a proof that the class of non-truths cannot be represented in the theory, in the following sense: assuming the the base interpretation extends the standard model of arithmetic, there is no formula of one free variable, N(x), which is true of just the non-true sentences.[31] For if there were we could, by diagonalisation, produce a formula, ψ, which has the same value in any evaluation in the hierarchy as $N(\underline{\psi})$. It follows that ψ is true iff $N(\underline{\psi})$ is true. But $N(\underline{\psi})$ is true iff ψ is non-true. Contradiction.

Since Yablo talks about the non-true sentences, it follows that if he is to be consistent and mean what he says, he *must* be talking in a language other than the one he is discussing, a metalanguge; the strategy is thus forced into semantic ascent, as we noted in section 4 that it must be. We noted also the self-defeating nature of this move if the aim is to produce a semantically closed theory. Of course, since the underlying semantics of the language does allow for things to be both true and false, it would be possible for Yablo to refuse the semantic ascent by accepting the contradiction, but only by himself becoming a dialetheist, and, as we noted above, he is not prepared to be this red-blooded. Moreover, if one is, then the motivation for the construction seems to be under-cut.[32] A semantically closed theory, with truth behaving as we think it does, is much more simply obtainable, as I indicated in section 5.

Thus we see that Yablo's construction merely adds another epicycle to the "solve the paradoxes" research programme, as any attempt to face the semantic paradoxes and remain consistent, must.

Graham Priest
Department of Philosophy
University of Western Australia
Nedlands, Australia

Notes

"A successful research programme bustles with activity. There are always dozens of puzzles to be solved and technical questions to be answered; even if *some* of these--inevitably--are the programme's own creation. But this self-propelling force of the programme may carry away the research workers and cause them to forget about the problem background. They tend not to ask any more to what degree they have solved the original problem, to what degree they gave up basic positions in order to cope with the internal technical difficulties. Although they may travel away from the original problem with enormous speed, they do not notice it. Problemshifts of this kind may invest research programmes with a remarkable tenacity in digesting and surviving almost any criticism.

Now problem shifts are regular bedfellows of problem solving and especially of research programmes. One frequently solves very different problems from those which one has set out to solve. One may solve a more interesting problem than the original one. In such cases we may talk about a 'progressive problemshift'. But one may solve some problems less interesting than the original one; indeed, in extreme cases, one may end up with solving (or trying to solve) no other problems but those which one has oneself created while trying to solve the original problem. In such cases we may talk about a *'degenerating problemshift* '." Lakatos [1968], pp.128-9.

A third variant is given by Belnap [1982].

This is an inessential modification of the GH construction.

Belnap gives yet a third possibility, viz. $X_\lambda^+(U) \cup (Z_\lambda - X_\lambda^-(U))$ where Z_λ is an arbitrary (sub) set (of S).

Some care needs to be taken over the terminology. What I call 'globally stable', Herzberger calls 'stable' and Gupta calls 'relatively stable'. What I call 'absolutely stable', Herzberger calls 'naively stable' and Gupta calls 'stable'. Gupta's definitions of the various notions is also slightly different but equivalent.

Specifically, if φ is a sentence in language of order **n+1** in the Tarski hierarchy and $\underline{\varphi}$ is its name in the language of order **n+1** (we can allow names for the sentences of all the languages to occur in each language), then $\mathbf{T}_n\underline{\varphi}\equiv\varphi$ may fail, where $\mathbf{T}_n$ is the truth predicate in the language of order **n+1**.

The question of the sense in which an application of the rule in general "improves" the extension of **T** is, fortunately, an issue we can avoid.

See, *e.g.* Davidson [1967].

There are totally psychologistic accounts of meaning, such as that of Grice. But such theories are incapable of dealing with the compositionality of meaning.

It might be suggested that only the absolutely stable sentences are meaningful, and hence that we require of a theory of meaning only that it deliver the T-sentences for absolutely stable sentences. However, this will not work since, as we shall see, for sufficiently rich languages, the set of absolutely stable sentences is not arithmetic. Thus, assuming that we can effectively tell a meaning-giving sentence when we see one, the truth-theory could not be axiomatic.

The construction must be iterated beyond the finite since, in general, stabilisation will not occur at finite levels.

The point is made in Dummett [1973]. See esp.p.320.

I will return to the question of temporality in the final part of the paper.

For the terminology, see Haack [1978], ch.8.

See his [1982], *e.g.* pp.27-30. My own shot is in Priest [1984].

[1982] sect.II. A simpler proof of this fact can be found in Priest [1984].

See, *e.g.* Boolos and Jeffrey [1974], p.176.

This was, of course, emphasized by Tarski. But the current literature trying to show how a theory can handle its own truth predicate is a further narrowing of the programme of solving the paradoxes, predicated on the assumption that all semantic relations can be defined in terms of truth (or at least satisfaction). But all the proposed theories work with other semantical notions–such as stability–which cannot be defined in terms of truth (or satisfaction). The very theories therefore show the further narrowing of the programme to be untenable.

Thus, ZF cannot prove its own consistency because it cannot quantify over proper classes, whereas NBG cannot prove its own consistency since one cannot prove certain "impredicative" classes to exist.

The strategy itself of avoiding contradictions by postulating a temporal dimension is hardly a new one; it can be found *e.g.* in Kant. See von Wright [1968] sect.11.

The term was coined in Priest, Routley and Norman [1986].

Pace Herzberger [1982a], p.481.

Pace Gupta [1982], p.51.

There is perhaps one final observation worth making. Gupta isolates a phenomenon he calls failure of local determination [1982], p.21ff. What this amounts to in effect is a failure of compositionality. The semantic value of a sentence is determined by things other than the relevant semantic values of its components. When we look at his proof that local determination may fail, we meet an old friend: Curry's paradox. In a recursively based account of (absolute) truth, compositionality must hold. However, in a semantically closed theory, the principle $A \leftrightarrow (A \rightarrow B)/B$ must fail if Curry paradoxes are to be avoided.

For a further discussion of these matters, see Priest [1986].

Another is Woodruff [1984], which I have discussed in Priest [1984a]. A number of the comments made there also apply to

Yablo's construction.

The comments on the genuineness, inevitability etc. of paradox I quite endorse. But because of this, I do not think that it is necessary to produce yet another construction which tries to avoid them. I also disagree with Yablo's "cosmological argument" for groundedness, pp 316-7.

E.g., Rescher and Brandom [1979]. See esp. ch 26.

A rather swift comment is made on the matter in fn.1. This is all.

For a discussion of these see Davies [1981] ch.1. Note also that Yablo is not against using psychologistic considerations to try to make his account plausible. (See, *e.g.*, p 330.)

Note that, in particular, $\neg T(x)$ will not do. $\neg T(x)$ is true iff $T(x)$ is false iff x is *un*true.

See Priest [1984a] section 4.

Bibliography

Belnap, N., "Gupta's Theory of Truth," *Journal of Philosophical Logic*, 1 (1982) 103-16.

Boolos, G. and Jeffrey, R. *Computability and Logic*. Cambridge University Press, 1974.

Brady, R., "The Non-Triviality of Dialectical Set Theory," in Priest, Routley and Norman [1986].

Davidson, D., "Truth and Meaning,"*Synthèse*, XVII (1967) 304-23.

Davies, M. *Meaning, Quantification, Necessity*. Routledge and Kegan Paul, 1981.

Dummett, M. *Frege*. Duckworth, 1973.

Gupta, A., "Truth and Paradox,"*Journal of Philosophical Logic*, 11 (1982) 1-60.

Haack, S. *Philosophy of Logics*. Cambridge University Press, 1978.

Herzberger, H., "Notes on Naive Semantics," *Journal of Philosophical Logic*, 11 (1982) 61-102.

Herzberger, H., "Naive Semantics and the Liar Paradox," *Journal of Philosophy*, LXXIX (1982) 479-497. [1982a]

Hegel, G.W.F. *Logic: Part One of the Encyclopedia of the Philosophical Sciences*: English translation by W. Wallace. Oxford University Press, 1975 (originally published, 1830).

Lakatos, I., "Changes in the Problem of Inductive Logic," Ch.8 of Vol.2 of Lakatos' *Philosophical Papers*. Eds. J. Worrall and G. Currie. Cambridge University Press, 1978 (originally published 1968).

Montague, R., "English as a Formal Language," Ch.6 of Montague's *Formal Philosophy*. Ed. R.H. Thomason. Yale University Press, 1974.

Priest, G., "Logic of Paradox," *Journal of Philosophical Logic*, 8 (1979) 219-41.

Priest, G., "Sense, Entailment and *Modus Ponens*," *Journal of Philosophical Logic*, 9 (1980) 415-35.

Priest, G., "Semantic Closure," *Studia Logica*, 43 (1984) 117-129.

Priest, G. "Logic of Paradox Revisited," *Journal of Philosophical Logic*, 13 (1984) 153-179. [1984a]

Priest, G., "Classical Logic *Aufgehoben*," in Priest, Routley and Norman [1986].

Priest, G. *In Contradiction*, to appear. [198+]

Priest, G. and Crosthwaite, J., "Relevance, Truth and Meaning," in *Directions of Relevant Logic*. Eds. R. Routley and J. Norman. Martinus Nijhoff (forthcoming). [198+]

Priest, G., Routley, R. and Norman, J. *Paraconsistent Logics* Philosophia Verlag, 1986.

Ramsey, F., "The Foundations of Mathematics," in *The Foundations of Mathematics and other Logical Essays*. Ed. R.B. Braithwaite. Routledge and Kegan Paul, 1931 (originally published 1926).

Rescher, N. and Brandom, R. *The Logic of Inconsistency.* Blackwell, 1979.

Routley, R., "Ultralogic as Universal," Appendix to *Exploring Meinong's Jungle and Beyond.* Australian National University, Research School of Social Sciences, 1980 (originally published 1977).

Woodruff, P., "Paradox, Truth and Logic. Part I: Paradox and Truth," *Journal of Philosophical Logic*, 13 (1984) 213-232.

von Wright, G. *Time, Change and Contradiction.* Cambridge University Press, 1968.

Yablo, S., "Truth and Reflection," *Journal of Philosophical Logic,* 14 (1985) 297-349.