

DEFLATIONARY TRUTH AND THE LIAR

Received 25 July 1997; revised 8 January 1998

KEY WORDS: correspondence, deflationism, Liar, paradox, truth

The topic of this paper lies at the intersection of two contemporary debates about truth. One debate is concerned with the nature of truth, and its protagonists are the correspondence theorist and the deflationist. The focus of the other debate is the Liar, where the challenge is to provide a theory of truth that adequately addresses semantical paradox. There has been surprisingly little contact between these two debates, despite their common interest in the concept of truth. Yet I think that they do bear on one another, and that is what I shall be arguing here. I shall argue that the deflationary conception of truth is severely compromised by the Liar in ways that the correspondence conception is not.

1. DEFLATIONARY TRUTH

Consider first the debate over the nature of truth. According to the traditional correspondence theory, truth is a substantive property shared by all truths – a property like *correspondence to a fact* or *correspondence to a state of affairs that obtains*.¹ A correspondence theorist might say that the sentence ‘snow is white’ is true in virtue of its corresponding to a fact, and so is the sentence ‘grass is green’. Of course, to make the correspondence theory precise, we will need to say what the truth-bearers are, what facts or states of affairs are, what the correspondence relation is, and what it is to obtain. But we are already in a position to appreciate the deep difference between the correspondence theorist and the deflationist. The deflationist is opposed to the correspondence conception of truth in a special, radical way. The deflationist does not present a competing view of the nature of truth – rather, the deflationist denies that truth has a substantial nature at all.²



In what follows I shall mainly consider one kind of deflationist, the *disquotationalist*. Take the familiar T-sentences:

'snow is white' is true iff snow is white
 'grass is green' is true iff grass is green
 'whales are fish' is true iff whales are fish
 :
 :

According to the disquotationalist, there is no more to truth than is revealed by these T-sentences. One may think of each of them as a partial definition of truth. Tarski writes:

... every equivalence of the form (T) obtained by replacing 'p' by a particular sentence, and 'x' by a name of this sentence, may be considered a partial definition of truth, which explains wherein the truth of this one individual sentence consists.³

The sentence 'snow is white' is true if and only if snow is white – and that's all there is to the truth of 'snow is white'. And the sentence 'grass is green' is true if and only if grass is green – that's all there is to the truth of 'grass is green'. These partial definitions do not ascribe a common feature to the true sentences. For the correspondence theorist, true sentences share a common property: they all correspond to a fact. But for the disquotationalist, there is nothing they share in virtue of being true. Truth is not a genuine property.

On this kind of view – championed by Quine, among others – truth is merely a *device for disquotation*. According to Quine, saying

'snow is white' is true

is just an indirect way of saying something about the world – namely, that snow is white. Truth-talk is *eliminable* – we can semantically descend. Talk about the truth of sentences can be eliminated in favour of direct talk about the world.

So why employ 'true' at all? According to the disquotationalist, the usefulness of the truth predicate emerges when we consider sentences like

- (a) What Claire said is true,
 and
 (b) Every alternation of a sentence with its negation is true.
 In these cases, we refer to sentences not by their quote names, but via a definite description or quantification. But whether we mention sentences via quote-names, definite descriptions or quantification, the truth predicate reminds us that "though sentences are mentioned, reality is still the whole point".⁴ The truth predicate serves "to point through the sentences to the reality".⁵

And according to the disquotationalist, truth-talk is eliminable even in cases like (a) and (b). Following Tarski's lead,⁶ the disquotationalist will point out that every sentence has a quote name, and go on to offer this eliminative analysis of 'true' in (a):

What Claire said is true iff what Claire said = 's₁' & s₁
 or what Claire said = 's₂' & s₂
 or

where 's₁', 's₂', ... abbreviate sentences.⁷ This preserves the spirit of the disquotational account: the truth predicate in (a) points through Claire's utterance to the reality, as expressed by 's₁' or 's₂' or

The disquotationalist will offer a similar account of generalizations over sentences. An eliminative analysis of (b) is given by:

for any sentence x, if x = 's₁' then (s₁ or not-s₁)
 or if x = 's₂' then (s₂ or not-s₂)
 or

where 's₁', 's₂', ... abbreviate the sentences over which x ranges. We can see (b) as a way of expressing an infinite conjunction, say

Tom is mortal or Tom is not mortal
 and Snow is white or snow is not white
 and

where each of these conjuncts is about the world.⁸ Quine writes:

... if we want to affirm some infinite lot of sentences that we can demarcate only by talking about the sentences, then the truth predicate has its use. We need it to restore the effect of objective reference when for the sake of some generalization we have resorted to semantic ascent.⁹

The disquotational treatment of (a) and (b) suggests that 'true' is not just a device for disquotation; it is also a device for expressing infinite disjunctions and conjunctions. These considerations suggest a natural disquotational definition of 'true':

DefT x is true iff (x = 's₁' & s₁) or (x = 's₂' & s₂) or

where 's₁', 's₂', ... abbreviate sentences.¹⁰ And falsity may be defined as follows:

DefF x is false iff (x = 's₁' & ~s₁) or (x = 's₂' & ~s₂) or

Notice that the T-sentences are (easy) logical consequences of the truth definition, given a suitable infinitary logic: Tarski's criterion of adequacy is satisfied.¹¹

DefT is an infinitary definition. Can we find a disquotational definition of truth that is finitely stated? Tarski considers this schematic definition:

$$x \text{ is a true sentence iff } \exists p(x = 'p' \& p).$$

Obvious problems arise if we interpret the quantifier objectively. There is the problem of quantifying into quotes. And the open formula " $x = 'p' \& p$ " is grammatically ill-formed, since what follows the conjunction sign is not a sentence but a name.¹³ Instead, then, we may take the quantifier substitutionally. In the standard terminology, we write:

$$x \text{ is a true sentence iff } \sum p(x = 'p' \& p),$$

where we associate with the variable ' p ' a set of expressions that are acceptable substituends (here, sentences of the language for which 'true' is defined). But how should we read the definiens? We can provide metalinguistic truth conditions: that there is a substitution instance of the open sentence " $x = 'p' \& p$ " which is true. But the specification of these truth conditions is in terms of truth, the very notion we are trying to define. So we need another reading.¹⁴ And it seems that the disquotationalist can do no better than to revert to DefT. The idea is to take " $\sum p(x = 'p' \& p)$ " as abbreviating an infinite disjunction, namely, the right-hand side of DefT.¹⁵

A disquotationalist might abandon a direct definition of truth in favour of a recursive account, according to which 'true' is defined Tarski-style in terms of the more basic notions of reference and satisfaction. Given a language with a finite stock of names and predicates, reference may be disquotationally defined by a finite list of sentences of the form " a refers to a ", and satisfaction by a finite list of sentences of the form " x satisfies ' F ' iff x is F ". In this way, reference and satisfaction are finitely defined – and so truth is finitely defined. But such a recursive disquotationalist is restricted to languages whose sentences have the appropriate kind of logical form. And there is an array of truths that are notoriously hard to fit into the Tarskian mould: belief attribution, counterfactuals, modal assertions, statements of probability, and so on.¹⁶ In the face of these difficulties, the disquotationalist might well prefer a direct definition of truth for sentences, even if it is infinitary.

In what follows I shall work with the disquotational account of truth embodied in DefT. There are two features of this account that are worthy of special attention – they capture the basic disquotational intuitions about truth. First, *there is no more to the truth of a sentence than is given by the*

disquotation of its quote-name. Second, *the truth predicate is eliminable*. DefT is an eliminative definition: truth-talk can in principle always be eliminated in favour of direct talk about the world.

There is another way in which we might present the disquotational account, in terms of a theory of truth rather than a definition. We specify this disquotational theory of truth by specifying its axioms – and its (infinitely many) axioms are just the T-sentences, the instances of the truth-schema " $'p'$ is true iff p ".¹⁷ This is, for example, how Horwich presents his minimal theory of truth for utterances.¹⁸ Clearly, the theory is driven by the same disquotational intuitions: the truth of a sentence is still given by disquotation, and 'true' does not appear on the right-hand side of the T-sentences. Though for convenience I shall work with DefT, it will make no difference whether we think of disquotational truth via the disquotational definition or via the disquotational theory. Either way, disquotation and eliminability are essential features.

2. THE LIAR PARADOX

In the dispute between the correspondence theorist and the deflationist, the Liar has hardly figured at all. And there is a reason for this: deflationists and correspondence theorists agree that since the Liar is a problem for both sides, it can safely be ignored. Marian David puts it this way:

I have completely neglected the paradox of the Liar in my discussion of disquotationalism and the correspondence theory. The reason is simple. Since it afflicts both candidates, considerations concerning the paradox will not be of much help in advancing the debate between substantive and deflationist theories of truth.¹⁹

And it is not just that the Liar is a shared affliction. It is also claimed that the possible cures – the various proposed solutions to the Liar – will be equally available to the deflationist and the correspondence theorist alike. Horwich writes:

There is no reason to suppose that the minimalist answers that are advanced in this essay could be undermined by any particular constructive solution to the paradoxes – so we can temporarily set those problems aside.²⁰

I shall take issue with both these claims. I shall argue that there are features of Liar sentences that cannot be squared with disquotational intuitions about truth, and that the standard ways out of the Liar are incompatible with disquotationalism. In contrast, neither the Liar sentences nor the proposed solutions compromise the correspondence conception.

Now there is a sense in which both theories are equally afflicted by the Liar. Given either a deflationary or a correspondence theory of truth, we

can produce a Liar couched not in terms of truth, but rather in the terms of the theory itself.

Tarski showed us how to construct a disquotational Liar. Let '*c*' be a typographical abbreviation of the following expression 'the sentence written on the sixth line of this page. We now consider the following statement:

For all *p*, if *c* is identical with the sentence '*p*', then not *p*.

(Intuitively, this is a Liar sentence – taking truth disquotationally, this statement says of itself that it is not true.) As Tarski says, we can establish empirically:

- (α) the sentence 'for all *p*, if *c* is identical with the sentence '*p*', then not *p*' is identical with *c*.

Tarski goes on to assume:

- (β) for all *p* and *q*, if the sentence '*p*' is identical with the sentence '*q*', then *p* if and only if *q*.

Tarski concludes: "By means of elementary logical laws we easily derive a contradiction from the premisses (α) and (β)."²¹ Here we have a version of the Liar couched in disquotational terms.

A correspondence Liar is easily constructed. Consider the sentence:

This sentence does not correspond to a fact,
or the sentence:

This sentence represents a state of affairs that does not obtain.
We can always produce a Liar sentence tailored to the given version of the correspondence theory.

So both theories face their own versions of the Liar: this is a problem that they share. But what about the ordinary Liar, couched not in terms of disquotation or correspondence, but in terms of truth (or falsity)? The ordinary Liar presents a more fundamental challenge to the theories, simply because they are theories of truth. And here, I shall argue, the problem is not equally shared: it is only disquotational intuitions that have to be abandoned in the face of the ordinary Liar.

There are many versions of the Liar. Some turn on the notion of falsity:

- (1) (1) is false.

Others turn on the notions of truth and negation:

- (2) (2) is not true.

Some involve loops. Suppose John says

What Claire is saying is true,
while Claire says

What John is saying is false.

Given certain natural assumptions, all these sentences generate a contradiction.

A cousin of the Liar is the Truth-Teller. Suppose that Jane is mistaken about her own identity, and says:

What Jane is saying now is true.

Then we have a Truth-Teller: a sentence that says of itself that it is true. Unlike the Liar, the Truth-Teller does not generate a contradiction – but it too is intuitively pathological.

A crucial feature of both the Liar and the Truth-Teller is that they are *ungrounded*: in attempting to evaluate them, we are continually led to sentences that involve 'true' – we never reach a sentence free of the truth predicate. Suppose, by way of contrast, that Joanne and Mark produce the following utterances:

Joanne: What Mark is saying is true.

Mark: 'Snow is white' is true.

Here, both utterances are grounded – their evaluation ultimately depends on the evaluation of a sentence (namely, "Snow is white") that is free of the truth predicate.

Still, even though Joanne's utterance is grounded, it presents a *prima facie* difficulty for the disquotational definition of truth. As we have seen, the disquotationalist wishes to accommodate utterances like Joanne's, where the truth predicate is attached to a definite description. The difficulty arises because, in the given circumstances, the definite description refers to an utterance that itself involves a use of 'true'. There is nothing remarkable about this: in ordinary language, we often refer to or generalize over utterances that themselves evaluate other utterances. It is certainly something that the disquotationalist needs to accommodate; it would be unacceptably restrictive and artificial to place utterances like Joanne's outside the scope of the truth definition. Disquotationalists do disagree over the scope of the definition: for some, its scope includes foreign languages as well as the home language;²² for some, it does not go beyond the home language;²³ and for others, it applies only to the sentences or utterances of a single

speaker's idiolect.²⁴ But it seems that no disquotationalist will exclude Joanne's utterance – it is clearly an utterance of English, and clearly an utterance of Joanne's idiolect.

Now consider again the disquotational definition:

DefT x is true iff ($x = 's_1' \& s_1$) or ($x = 's_2' \& s_2$) or

(Here, $s_1, s_2, \dots$ will be sentences or utterances in the disquotationalist's object language – English, perhaps, or Joanne's idiolect.) If we put 'What Mark said' for x , one of the disjuncts will be

- (i) What Mark said = "Snow is white" is true' and 'Snow is white' is true.

And this contains a use of the truth predicate. So here we have a use of 'true' on the right-hand side, in the *definiens*. Indeed, the presence of 'true' on the right-hand side is pervasive: every utterance of the object language will appear quoted and disquoted in the *definiens*, and among these will be all utterances that contain uses of 'true'. And the replacement of any of these uses of 'true' by the disquotational *definiens* reintroduces all these uses of 'true'. This is the *prime facie* difficulty: there is a threat of circularity, or an infinite regrees of infinite disjunctions.

The disquotationalist has a ready response: the circle or regress here is not vicious. Consider again the disjunct (i). What distinguishes this disjunct is that its first conjunct – an identity statement – is true; all the other identity statements, and so all the other disjuncts, are false. Now when we eliminate the use of 'true' from this disjunct via the disquotational definition, one of the infinitely many disjuncts that we obtain is this:

- (ii) 'snow is white' = 'snow is white' & snow is white.

Notice two features of this disjunct: the identity statement is true, and the truth predicate does not appear. The idea is that, given a grounded sentence, we can always 'selectively disquote': we can always trace a path through the disjuncts that contain true identity statements to a sentence free of 'true'. In the case of Joanne's utterance, we trace a path through the disjuncts (i) and (ii) to the sentence 'Snow is white'. Joanne and Mark are both using the truth predicate to point through language to the world – it's just that Joanne's route to the world is a little less direct.

If we present disquotationalism in the axiomatic way mentioned above, as a theory of truth whose axioms are the T-sentences, then the problem presents itself this way: there will be T-sentences where 'true' appears on the right. In response, the disquotationalist may provide a suitably adjusted version of selective disquotation. Since there is a quote name of the

sentence Mark uttered, there will be a T-sentence associated with Claire's utterance, viz.:

- (a) "Snow is white" is true' is true iff 'snow is white' is true.

Here, 'true' appears on the right-hand side. We can avoid any appearance of vicious circularity or regress by the observation that the right-hand side of (a) is the left-hand side of another T-sentence:

- (b) 'Snow is white' is true iff snow is white.

And now we can trace a path through the biconditionals (a) and (b), arriving at the right-hand side of (b), a sentence directly about the world. The general idea is that there is a route from any evaluation to a sentence free of the truth predicate.

But consider again the ungrounded sentences. Consider our Truth-Teller:

- (J) What Jane is saying now is true.

Here, selective disquotation fails. The relevant disjunct on the right of the truth definition is:

What Jane is saying = "What Jane is saying now is true" and what Jane is saying now is true.

And here we cannot trace an appropriate path through the disjuncts to a sentence free of 'true'. (Or, on the alternative presentation of disquotationalism, we cannot trace an appropriate path through the T-sentences to a sentence free of 'true'.) The use of 'true' appears to be ineliminable. And so it seems that we have no disquotational account of truth for the Truth-Teller. Since the argument can be made using any member of the Liar family,²⁵ it seems that we have no disquotational account for ungrounded sentences generally.

The disquotationalist may respond by placing the Liar sentences outside the scope of the truth definition (or, on the axiomatic version, limit the T-sentences that can serve as axioms of the theory). According to Horwich, for example,

we must conclude that permissible instantiations of the equivalence schema are restricted in some way so as to avoid paradoxical results ... certain instances of the equivalence schema are not to be included as axioms of the minimal theory²⁶

Horwich speaks here of avoiding paradoxical results, but the problem is broader than that. An ordinary version of the Liar, say

- (1) (1) is false

generates a contradiction. But other semantically pathological sentences, like Jane's Truth-Teller, do not lead to contradiction – yet still the disquotational theory cannot accommodate them. The problem for the disquotational theory goes beyond the strictly paradoxical sentences and the contradictions they produce. The reason that the theory cannot handle the Liar and the Truth-Teller, and ungrounded sentences generally, is that the theory cannot eliminate 'true' (or 'false') from these sentences, and a vicious regress seems unavoidable.

Suppose then that the disquotationalist restricts the scope of the truth definition. One way to avoid liars, truth-tellers and loops is to define truth for a language in which 'true' does not appear. Horwich writes:

We know that this restriction need not be severe. It need have no bearing on the propositions of science – the vast majority of which do not themselves involve the concept of truth.²⁷

But the restriction to languages without 'true', such as the language of science, is surely too severe – as we saw, deflationists are after an account of truth for natural languages or for speakers' idiolects. Indeed, Horwich himself takes his theory to apply to all possible extensions of a natural language like English.²⁸ And Jane's Truth-Teller, or John and Claire's looped utterances, are ungrounded not in virtue of syntactic or semantic form, but in virtue of the empirical facts; in another's mouth, or at another time, what Jane, John and Claire said would be straightforwardly true or false, and within the intended scope of the disquotational theory.²⁹ If the disquotationalist throws out ungrounded utterances on the grounds that they belong to a language beyond the scope of the theory, then she also must throw out utterances of the same type that the theory is supposed to cover. Moreover, the disquotationalist response must not be *ad hoc*: if the response places restrictions, these restrictions must be principled.³⁰ In short, there must be a positive account of the Liar as it arises in natural languages (or in speakers' idiolects) that is compatible with the disquotational conception. As we have seen, Horwich thinks there is no reason to suppose that deflationism "could be undermined by any particular constructive solution to the paradoxes".³¹ But it remains to be seen whether this is so.

3. DISQUOTATION, UNGROUNDEDNESS AND GAPS

A familiar approach to the Liar adopts a non-classical semantics that allows truth-value gaps. On such an approach, Liar and Truth-Teller sentences or utterances are neither true nor false. The correspondence theory has a natural enough way with gaps. If we characterize the correspondence

theory as follows:

An utterance is true iff it corresponds to a state of affairs that obtains,

we can say that an utterance is false if it corresponds to a state of affairs that does not obtain, and neither true nor false if it fails to correspond to any state of affairs. For reasons that will emerge below, I think we should not regard the appeal to gaps alone as an adequate response to the Liar. But that's a matter independent of the claim that the correspondence intuition can accommodate gaps. Whether gaps are needed (for the Liar, or vagueness, or future contingents, and so on) is one thing; whether a correspondence theorist can admit them is another. Here, and throughout, I am not attempting to explain the notions of *correspondence*, *states of affairs*, or *obtaining*; my concern is only with the pretheoretic correspondence intuition. Truth value gaps do not compromise that intuition.

Are gaps available to the disquotationalist? Recall the disquotational definitions of truth and falsity.

DefT x is true iff $(x = 's_1' \& s_1)$ or $(x = 's_2' \& s_2)$ or ..., and

DefF x is false iff $(x = 's_1' \& \sim s_1)$ or $(x = 's_2' \& \sim s_2)$ or ...,

where ' s_1 ', ' s_2 ', are the members of the given substitution class, call it S . It follows easily from these definitions that a sentence is neither true nor false only if it is not a member of S .³² This consequence is initially troubling in two ways. First, consider again Jane's Truth-Teller (J), or any other empirical version of semantic pathology. Such a meaningful, declarative utterance would seem to belong to S – and indeed other tokens of the same type do. Such empirical cases show that membership in S will sometimes be a matter of the empirical circumstances of the utterance, and not a matter of intrinsic syntax or semantics. Second, if the disquotationalist says no more about gaps, gappy sentences will be lumped along with, say, Julius Caesar, who is also outside S . But Julius Caesar does not suffer a truth value gap (except perhaps in a highly attenuated way). So the disquotationalist must say more: gappy sentences must be distinguished from tokens of the same type that are admitted into S . And they must be distinguished from other entities that also fall outside S , but are not gappy in any natural sense.

At this point, the disquotationalist might embrace the notion of ungroundedness. If a disquotational account can be given of this notion, then perhaps the required distinctions can be drawn. The disquotationalist might

observe that when we instantiate DefT or DefF to an ungrounded sentence, we obtain an infinite regress. And this observation might provide the basis for a disquotational characterization of ungroundedness. We can present informally a procedure for determining whether P is ungrounded. Suppose P is a sentence of the form ' Q is true' (or ' Q is false'). Suppose P is in the substitution class – i.e. is one of the s_i 's – for DefT (or DefF). Now instantiate DefT to P . Of the infinitely many disjuncts on the right-hand side, only one will have a true identity as its first conjunct. If 'true' or 'false' appears in this disjunct, eliminate it in turn via DefT or DefF. Keep going in this way. If we obtain an infinite regress, P is ungrounded. Since the infinite regress is generated by repeated instantiations of the disquotational definitions, it seems plausible to say that ungroundedness may be captured disquotationally. On the present disquotational account, then, we justify gaps by pointing to the ungrounded character of Liarlike sentences, and go on to characterize ungroundedness in disquotational terms.

With gaps on board, the disquotationalist might hope to avoid restrictions on DefT and the T-schema. Perhaps circular instances of the definition and the schema are admissible. Suppose that ' P ' is a Liarlike sentence. Then

' P ' is true iff P

may be counted true if both sides are gappy. In the spirit of the revision theory of truth, we might accept each T-biconditional as a partial definition of truth for all sentences (Liar and Truth-Tellers included) and accept the consequence that truth is a circular notion.³³ Such an approach may well tempt the disquotationalist: even the truth of ungrounded sentences is a matter of disquotation.

However, the disquotationalist cannot take this tack. It is not just that it means giving up the idea that truth is eliminable. Consider again the T-sentence for the Liarlike sentence ' P '. We are taking ' P ' to be gappy. So the right-hand side of the T-sentence is gappy. But the left-hand side is false: it is false that ' P ' is true.³⁴ So the T-sentence is not true.

This is an instance of a more general problem: given a gappy sentence (whether a Liar sentence or not), the corresponding T-sentence is untrue. In order to maintain the truth of such a T-sentence, we might introduce a *weak* notion of truth,³⁵ where " P is true" always has the same semantic status as " P ". (In particular, if ' P ' is gappy, so is " P is true".) The revision theory is a theory of this weak notion.³⁶ But the disquotationalist cannot ignore the *strong* notion of truth, where if we say of a gappy sentence that it's true, we have said something false.³⁷ All truth, weak and strong, is to be deflated.

So it seems that the disquotationalist will do better to exclude the Liarlike sentences from the scope of DefT and the truth-schema. If a disquotational account of ungroundedness can be made out, these exclusions will be principled. Unlike the members of the substitution class, Liarlike sentences are ungrounded; and unlike Julius Caesar, they are tokens of sentence-types of English (or some natural language or idiolect).

Still, the disquotationalist must give up the idea that 'true' can in principle be eliminated via disquotation from every ordinary utterance of English. On the present approach, the disquotationalist must remain silent about uses of 'true' in ungrounded utterances: they are beyond the scope of the disquotational definition. Perhaps the disquotationalist will reply that this is no great loss – these uses of 'true' generate semantic pathology anyway, and it is not embarrassing to admit that a disquotational theory of truth cannot accommodate them. We might be unmoved by this reply: since the concept of truth in ordinary language gives rise to paradox, we might expect a theory of truth to account for paradox rather than to set it aside. We have, however, a more powerful reason for finding the reply unacceptable. There is a pressing problem for the disquotationalist (and for the gap theorist) generated by certain 'strengthened' forms of the Liar.

4. THE STRENGTHENED LIAR

It is well known that there are strengthened versions of the Liar that threaten the adequacy of gap approaches. Consider again the Liar sentence:

(2) (2) is not true.

Suppose we endorse the gap approach, and take (2) to be neither true nor false. Then, in particular, (2) is not true. But that's what (2) says. So (2) is true. And we are landed back in contradiction. Similar paradoxical reasoning can be generated from the sentence:

(3) (3) is false or neither true nor false.

What is the impact of the Strengthened Liar on the present disquotational account?

The disquotationalist owes us an account of (2), and, more generally, of 'not true'. A natural disquotational definition of 'not true' is this:

Def ~T x is not true iff ($x = 's_1' \ \& \ \sim s_1$) or ($x = 's_2' \ \& \ \sim s_2$) or ..., where s_1, s_2 are members of the substitution class S associated with DefT. If we follow the informal procedure outlined in the previous section, we

will find that (2) is ungrounded.³⁸ Notice, though, a questionable consequence of this characterization of 'not true': 'not true' is indistinguishable from 'false'.

An alternative definition of 'not true' incorporates a wide scope negation:

Alt~T x is not true iff $\sim[(x = 's_1' \& s_1) \text{ or } (x = 's_2' \& s_2) \text{ or } \dots]$, where ' s_1 ', ' s_2 ', ... abbreviate the members of the substitution class S . The definiens is equivalent to " x is false or x is not in S ".³⁹ Notice a significant difference between this alternative and Def~T. According to Def~T, something is not true only if it belongs to S . According to Alt~T, something is not true if it does not belong to S (or if it is false). The scope of Def~T is the same as the scope of DefT – but the scope of Alt~T is wider, and now 'not true' is clearly distinguished from 'false'.

But Alt~T comes with an unacceptably high price. Remember that our disquotationalist is treating (2) as ungrounded, along with (J) and (1). But if we adopt Alt~T, the pathological nature of (2) is lost. (2) is analyzed as:

(2) (2) is false or not in S :

Given DefT and DefF, it's easy to check that the assumptions that (2) is true and that (2) is false lead to contradiction. But assume instead that (2) is not in S for DefT, i.e.

$$(2) \neq s_1 \& (2) \neq s_2 \& (2) \neq s_3 \& \dots$$

From this infinite conjunction, and DefT and DefF, we can establish that (2) is not true and (2) is not false, since (2) is not in their associated substitution classes. But no contradiction is forthcoming. (Of course, intuitively we want to say that if (2) is not in S then it follows that (2) is true, given what (2) says. But DefT cannot accommodate this intuitive inference.)

Relatedly, if we prefer Alt~T over Def~T, we will no longer be able to establish the ungroundedness of (2). To suppose that (2) may be plugged into Def~T is equivalent to supposing that (2) is in the substitution class for DefT. And this guarantees that (2) is identical to s_k , for some k ; and from this an infinite regress follows. But when we suppose that (2) may be plugged into Alt~T, it does not follow that (2) is identical to some s_k . But it is only the s_i 's that get disquoted, and the infinite regress is generated by repeated disquotation. And so no regress can be established in the case of (2), and we cannot conclude that (2) is ungrounded.

So the disquotationalist is presented with a dilemma. If 'not true' is characterized by Def~T, 'not true' is indistinguishable from 'false'. And if 'not true' is characterized by Alt~T, the pathological nature of Liar

sentences that turn on the negation of truth is lost. Our present disquotationalist has to seize the first horn: as long as Liar sentences are excluded from DefT in virtue of their ungroundedness, Alt~T is not an option.

Suppose then that the disquotationalist works with Def~T. Let us take a closer look at a stretch of Strengthened Liar reasoning. Suppose that Aristotle, believing that what Plato has written next door is untrue, writes on the board

(L) The sentence written on the board in room 101 is not true.

But Aristotle is confused about his whereabouts and is himself in room 101. So (L) is a Liar sentence. The familiar reasoning shows that (L) is semantically pathological. So (L) isn't true. That is, we may infer (P). The sentence written on the board in room 101 is not true. But now, given (P), and given what (L) says, we infer

(R) (L) is true.

This reasoning appears intuitive and correct, and any adequate solution to the Liar must account for it. In my view, this reasoning is handled best by contextual theories of truth, according to which 'true' is a context-sensitive term, shifting its extension according to context.⁴⁰ Indeed, the phenomenon of strengthened reasoning provides a primary motivation for contextual views.

According to one development of the contextual line, we may discern differences in the contexts of (L), (P), and (R), to which the truth predicate is sensitive. As well as shifts in speaker, time and place, there are shifts of discourse position, relevant information, and intentions between the contexts of (L), (P), and (R).⁴¹ Moreover, at the first stage of the reasoning there is a pragmatic implicature in place, that (L) is to be evaluated via the schema for the occurrence of 'true' in (L). This implicature is cancelled, once we realize that (L) is pathological, and cannot be evaluated by its associated 'true_L' schema.

The overall structure of the account is as follows. We represent Aristotle's utterance as:

(L) (L) is not true_L,

where 'true_L' abbreviates 'true in the context of utterance of (L)'. In the course of our initial reasoning about (L), we find that it is neither true_L nor false_L. (L) cannot be given truth_L-conditions. If we insert it into the schema

s is true_L iff p

(where 's' is a name of the sentence substituted for 'p'), we obtain a contradiction. In short, (L) is pathological in its context of utterance. We cancel the implicature that (L) can be assessed via the 'true_L'-schema. And we infer that (L) is not true in its context of utterance; that is, we infer (P) (L) is not true_L.

We go on to evaluate (L) in a new 'reflective' context, where we now take into account (L)'s pathological status and what (L) says. Given (P) and given that (L) says that (L) is not true_L, we infer

$$(R) \quad (L) \text{ is true}_L,$$

where 'true_R' abbreviates 'true in the context of (R)'. Think of (R) as a final reflective evaluation of (L), an evaluation based on (L)'s pathologicity and what (L) says.

Is such a contextual analysis available to the disquotationalist? It is not, for a number of related reasons. As we have seen, the present disquotationalist defines 'not true' by Def~T. In particular, the occurrence of 'not true_L' in (L) will be defined this way. This will guarantee the pathology of (L). But this analysis will not allow for the cancelled implicature that accompanies (P). Like (L), (P) will be analyzed via Def~T, and it is built into this definition that (L) is in the substitution class – call it S_L – associated with the disquotational definition of 'true_L'. That is, any use of 'not true_L' carries with it the presumption that the evaluated sentence can be assessed by the schema for 'true_L'. Now, via Def~T, we read (L) and (P) as:

$$[(L) = s_1 \& \sim s_1] \text{ or } [(L) = s_2 \& \sim s_2] \text{ or } \dots$$

On this reading of (P), part of what (P) says is that (L) is a member of S_L; this reading cannot be combined with the cancellation of the implicature that (L) is in S_L.

The immediate problem would be avoided if we analyzed (P) via Alt~T. But we cannot analyze (L) that way if we want to preserve (L)'s pathology – as we surely do if we want to endorse the contextual account. So a consequence of this proposed analysis of (P) is that (L) and (P) will be given distinct readings. Apart from its adhocness, this suggestion is at odds with the final stage of the strengthened reasoning, the reasoning that leads to (R). There it is assumed that (L) and (P) say the same thing: we reason that given (P), and given that (L) says what (P) says, (L) is itself true.

And whether we read (P) via Def~T or Alt~T, (R) itself poses problems for the disquotationalist. There's a specific problem and a general problem. The specific problem in this. We are led to the reflective evaluation (R) through these considerations: (L) says it isn't true (in its context

of utterance), and it isn't (because it's pathological). So what it says is the case: (L) is true. But what does (L) say according to the disquotationalist analysis? Again, (L) says:

$$[(L) = s_1 \& \sim s_1] \text{ or } [(L) = s_2 \& \sim s_2] \text{ or } \dots$$

But if that is what (L) says, then it's *false*: (L) says, among other things, that it's in the substitution class S_L – and it isn't.

The general problem can be put as follows. According to the contextual approach, strengthened liar reasoning shows that we are able to evaluate even pathological or ungrounded sentences as true or false. A Liar sentence like (L) is pathological and so not true; and that's what it says, so it is true. Contextual accounts provide for reflective evaluations of pathological sentences. (L) is true in the context of (R) – (L) is gappy only relative to its context of utterance, and in a suitably reflective context, (L) may be evaluated as true. The Truth-Teller (J) will be reflectively evaluated as false, since it is pathological, and so not true, yet it says it is true. A Liar sentence like (1) ('(1) is false') will also be false, because it is pathological, and so not false, yet it says it is false. The contextual analysis takes into account the pathology of the ungrounded sentences, and accommodates the evaluation of them according to their pathology and what they say. Now recall the disquotational strategy of the previous section: we exclude Liarlike sentences from the disquotational definition of truth in a principled way (via the notion of ungroundedness). To the charge that 'true' was not eliminable from the ungrounded sentences, the disquotationalist's reply was: "But they are pathological anyway; it's acceptable to set them aside." However, according to the present contextual account, these ungrounded sentences are true or false. Ungrounded sentences are truths and falsehoods from which 'true' and 'false' cannot be eliminated. And if there are truths (and falsehoods) expressed by ineliminable uses of 'true' and 'false', then so much the worse for the disquotational theory.⁴²

Moreover, consider a disquotational treatment of a reflective evaluation, say (R). We will obtain:

$$(L) \text{ is true}_R \text{ iff } (L) \text{ is not true}_L.$$

To borrow Tarski's way of talking, just as the truth of 'Snow is white' consists in snow's being white, so the truth_R of (L) consists in (L)'s being untrue_L. So what (L)'s truth_R consists in is inextricably tied to the ineliminable notion of truth_L. To explain what the truth of (R) consists in, then we cannot dispense with the notion of truth.

It is a similar story with the Truth-Teller. We have seen that (J) is reflectively evaluated as false – so here is a falsity from which ‘true’ cannot be eliminated. Further, consider a reflective evaluation of (L):

(R*) (J) is false_{R*}.

From the disquotational definition of falsity_{R*}, we obtain:

(J) is false_{R*} iff it is not the case that (J) is true_J.

The falsity_{R*} of (J) consists in its not being the case that (J) is true_J. As with (R), we cannot explain what the truth of (R*) consists in without appeal to the notion of truth. In general, to account for the truth of certain sentences, the notion of truth is itself indispensable. And truth’s ineliminability compromises the disquotational conception of truth.

It is instructive to contrast the correspondence intuition here. According to the correspondence conception, (L) is true if and only if it corresponds to a state of affairs that obtains. And (L) does. The state of affairs here is a semantic state, that of (L)’s being untrue_L. And this state obtains, because (L) is semantically pathological in its context of utterance. According to the correspondence conception, (L) is indeed true. Or consider the Truth-Teller. According to the correspondence conception, (J) is false if and only if (J) corresponds to a state of affairs that does not obtain. Now (J) corresponds to the state of affairs in which (J) is true_J. And this state of affairs does not obtain, since (J) is pathological in its context of utterance, and so not true_J. So (J) is false. The reflective evaluations (R) and (R*) are in line with the correspondence intuition.⁴³

In correspondence terms, (L) is true in virtue of its correspondence to a state of affairs that obtains; and (J) is false in virtue of its correspondence to a state of affairs that does not obtain. In each case, the state of affairs is a semantic state of affairs. If we put things this way, I think we can see why the disquotationalist finds strengthened reasoning so intractable. (L)’s truth is grounded in a semantic state of affairs, and not a non-semantic state of the world. Recall the disquotational intuition: *the truth predicate is eliminable* – truth-talk can in principle always be eliminated in favour of direct talk about the world. If we accept the contextual analysis, strengthened reasoning shows that this just isn’t so for (L) or any ungrounded sentence. The reasoning establishes certain essentially semantic facts about (L) (that it is pathological, that it is untrue_L), and it’s in virtue of these semantic facts that (L) is true. We cannot represent these facts or states of affairs in non-semantic terms. Truth is not eliminable.

5. HIERARCHICAL VIEWS

We have considered two main approaches to the Liar, one via gaps and the other via the context-sensitivity of ‘true’. Another standard proposal appeals to a hierarchy of languages.⁴⁴ Suppose we take the truth predicate to be stratified, so that any use of ‘true’ is always a use of ‘true in a language’, corresponding to distinct levels of language. Think of English as (cumulatively) stratified in the following way. At the ground level, we have the language L₀, containing sentences free of the truth predicate, like ‘snow is white’. The language L₁ includes L₀ and contains the truth predicate for L₀, ‘true-in-L₀’. So, for example, the sentence “‘Snow is white’ is true-in-L₀” is a sentence of L₁. And so on, up the hierarchy.

Consider again Joanne and Mark’s utterances:

Joanne: What Mark is saying is true.

Mark: ‘Snow is white’ is true.

Mark’s utterance is analyzed as a sentence of L₁:

‘Snow is white’ is true-in-L₀,

and Joanne’s as a sentence of L₂:

What Mark is saying is true-in-L₁.

Adopting the hierarchy, the disquotationalist now takes the truth definition to be restricted to truth for a given language. For example, the definition of ‘truth-in-L₁’ will look like this:

$$x \text{ is true-in-} L_1 \text{ iff } (x = 's_1' \& s_1) \text{ or } (x = 's_2' \& s_2) \text{ or } \dots,$$

where $s_1, s_2, \dots$ are the sentences of L₁. The disquotationalist seeks to eliminate Joanne’s use of ‘true’ in her utterance via this definition. Replace x by a name of Mark’s utterance, say ‘M’. One of the disjuncts we will obtain on the right-hand side is this:

M = “‘Snow is white’ is true-in-L₀”

and ‘Snow is white’ is true-in-L₀.

This disjunct (and only this disjunct) is true, and so M is true-in-L₁.

This analysis seems to fit the disquotational conception well enough. Notice that there is no threat of circularity, because we have distinct truth predicates on the left- and right-hand sides. And if we wish to eliminate any occurrence of the truth-in-L₀ predicate on the right hand side, we can

do so via the disquotational truth definition for true-in- L_0 , and arrive at a true disjunct free of 'true':

'snow is white' = 'snow is white' & snow is white.

Joanne and Mark use different truth predicates to point to the world through descending levels of language.

But Joanne's and Mark's utterances are grounded. Consider what happens when we move to more problematic sentences, like Jane's Truth-Teller:

(J) What Jane is saying now is true.

Can the disquotationalist allow (J) to be well-formed if she endorses the hierarchy? According to a hierarchical account, the truth predicate here is assigned some level, say σ , corresponding to the language L_σ .⁴⁵ Jane's utterance is represented as:

What Jane is saying now is true-in- L_σ ,

and Jane's utterance itself is a sentence of $L_{\sigma+1}$. So (J) contains an application of 'true-in- L_σ ' to a sentence of $L_{\sigma+1}$. But 'true-in- L_σ ' is not disquotationally defined for a sentence of $L_{\sigma+1}$. For 'true-in- L_σ ' is defined as follows:

x is true-in- L_σ iff $x = 's_1'$ & s_1 , or $x = 's_2'$ & s_2 , or ...

where the s_i 's are the sentences of L_σ . A disquotational definition of truth provides a stock of quote-names, and the truth of a given sentence is a matter of the disquotation of one of these quote-names. But no s_i is a quote name of (J), and so 'true-in- L_σ ' is not defined for (J). Recall the disquotational intuition: *There is nothing more to the truth of a sentence than is yielded by the disquotation of its quote-name*. In the case of (J), this intuition breaks down.

We can put the point in a slightly more precise way. Consider the relativized schema

x is true-in- L_σ iff $\sum p(x = 'p \& p')$,

where the substitution class associated with ' p ' – call it $S_\sigma -$ is the class of sentences of L_σ . This schema confers disquotational truth or falsity only on members of S_σ : it is only with respect to these sentences that the application of 'true-in- L_σ ' is equivalent to the disquotation of its quote-name. But (J) is not a member of S_σ . And so the schema does not confer disquotational truth or falsity on (J). Jane's use of 'true' cannot be treated disquotationally.⁴⁶

So it seems that the present disquotationalist must reject (J) as ill-formed, or at least as beyond the scope of the disquotational treatment of truth. But this amounts to a refusal to recognize (or at least to treat) members of the Liar family. In contrast, a correspondence theorist can be more accommodating. A correspondence definition of 'true-in- L_σ ' is given by:

x is true-in- L_σ iff x is a sentence of L_σ that corresponds to a state of affairs that obtains.

So (J) can be regarded as attributing to itself the property of being a sentence of L_σ that corresponds to a state of affairs that obtains. Since (J) is a sentence of $L_{\sigma+1}$ and not L_σ , (J) corresponds to a state of affairs that does not obtain. According to the correspondence conception, then, (J) is false (that is, false $_{\sigma+1}$). Compare Russell's treatment of the Liar:

Thus when a man says 'I am lying', we must interpret him as meaning: 'There is a proposition of order n which I affirm and which is false'. This is a proposition of order $n + 1$; hence the man is not affirming any proposition of order n ; hence his statement is false, and yet its falsehood does not imply, as that of 'I am lying' appeared to do, that he is making a true statement. This solves the Liar.⁴⁷

So the correspondence conception is not compromised by a hierarchical treatment of the Liar. In particular, we can give an account of (J)'s truth status in correspondence terms. But we cannot treat the application of 'true-in- L_σ ' to (J) disquotationally.

We do have reason to find the present hierarchical account unappealing in its own right, whether we are disquotationalists or correspondence theorists. On the present account, Liar sentences are either set aside as ill-formed or taken to be unproblematically true or false. This will not satisfy anyone who takes Liar sentences to be well-formed yet gappy, or ungrounded, or in some way semantically pathological. Moreover, on the present account, the English predicate 'true' is divided into infinitely many distinct predicates, and English itself is stratified into a hierarchy of distinct languages. This seems, as Russell once put it, "harsh and highly artificial".⁴⁸ With Tarski, we may doubt "whether the language of everyday life, after being 'rationalized' in this way, would still preserve its naturalness and whether it would not rather take on the characteristic features of the formalized languages".⁴⁹

One might combine the contextual analysis with a hierarchical account, taking the extension of 'true r ' to properly include that of 'true $_L$ '.⁵⁰ This avoids the artificial stratification of the truth predicate: on the contextual view, we have a single language containing a single context-sensitive truth

predicate. And the contextual account incorporates the intuition that Liar-like sentences *are* pathological. But once we admit the pathological nature of the Liarlike sentences, we again need to accommodate strengthened reasoning and reflective evaluations – something that the disquotationalist cannot do. So this more sophisticated version of the hierarchy is no more available to the disquotationalist than the naive version we have been considering.

6. DEFLATIONARY REFERENCE

Horwich finds it 'very plausible' that deflationism about truth 'goes hand in hand' with deflationism about reference.⁵¹ Similarly, Field: "If truth conditions play no central role in meaning, and truth is fully explained by the disquotation schema ... then the same is true of reference ...".⁵² The disquotationalist will propose a definition of reference along these lines:

DefRef x refers to y iff ($x = 'e_1'$ & $e_1 = y$)
or ($x = 'e_2'$ & $e_2 = y$) or ...

where e_1, e_2 , are associated with a substitution class of referring expressions.⁵³ It is easy to see that this definition yields all instances of the references schema: ' e ' refers to y iff $e = y$.

It is well known that the notion of reference, like truth, gives rise to paradoxes (the so-called 'definability' paradoxes, like Berry's, Richard's and König's). And while I cannot provide a full defence here, I would claim that these paradoxes compromise deflationary reference just as the Liar compromises deflationary truth. Consider, for example, a (naive) hierarchical view, according to which 'refers' is split into infinitely many distinct predicates. And consider a version of Berry's paradox, generated by the phrase 'the least integer that cannot be referred to in fewer than twenty-three syllables' (a phrase with twenty-two syllables). Assign the level σ to the occurrence of 'refers' in this phrase, and the level $\sigma + 1$ to the phrase itself. Then, in parallel to the case of the Truth-Teller (T), this phrase will not be in the substitution class associated with the disquotational definition of 'refers'. So an appeal to the simple hierarchy closes off a disquotational treatment of this referring phrase, just as it closed off a disquotational treatment of (T).

Or consider a strengthened paradox of reference. Suppose that at noon 7/1/98 I write on the board the following expressions:

- A. the ratio between the circumference and diameter of a circle.
- B. the positive square root of 36.

- C. the sum of the numbers to which expressions on the board in room 101 at noon 7/1/98 refer.

I believe that room 101 is the room next door, and that written on the board there are expressions that refer to numbers. But I am mistaken about my whereabouts – I am in fact in room 101. It is clear that A and B refer to numbers. But to what number does C refer? We can reason as follows:

Suppose towards a contradiction that C refers to a number, say k . Then the sum of the numbers referred to by A, B and C is $\pi + 6 + k$. But this number is the number to which C refers; so $k = \pi + 6 + k$, which is a contradiction. So C is a pathological referring expression: it appears to refer to a number, but it does not, on pain of contradiction.

From this reasoning, we infer:

- (P) C is pathological, and does not refer to a number.

We can now strengthen our reasoning, building on our conclusion that C is pathological. We can argue as follows:

Suppose that C is indeed pathological, and does not refer to a number. Then A and B are the only expressions on the board that refer to numbers. So *the sum of the numbers to which expressions on the board in room 101 at noon 7/1/98 refer* is $\pi + 6$. Since C and the underlined expression are composed of the very same words with the very same meaning, we infer:

- (R) C (and the underlined expression) refer to $\pi + 6$.

Here we have a strengthened paradox of reference, analogous to the Strengthened Liar. An adequate theory of reference, disquotational or not, must resolve it. Since the reasoning is intuitive we should not block it by artificial, *ad hoc* means. It is better to regard it as data that expresses semantic intuitions we have about the notion of reference. And, as with the Liar, I think that a contextual analysis is called for, one that preserves the validity of the reasoning and respects the data – and is unavailable to the disquotationalist.

The contextual account runs parallel to the Liar case.⁵⁴ We distinguish the original context in which C is first produced from the subsequent reflective context in which we reason about C. We treat 'refers' as a term that is sensitive to its context of utterance. We analyze (C) this way:

- (C) the sum of the numbers to which expressions on the board in room 101 at noon 7/1/98 refer_C.

where 'refers_C' abbreviates 'refers in the context of utterance of C'. There is an associated reference schema:

'e' refers_C to y iff $e = y$.

If we instantiate to C, and assume that C refers_C to some number k , we get the above contradiction (that $k = k + \pi + 6$). We conclude that C is pathological; it cannot be given reference_C conditions.

Now we reason that A and B are the only expressions on the board that refer_C to numbers, since C does not. So we infer that the sum of the numbers to which expressions on the board in room 101 at noon 7/1/98 refer_C is $\pi + 6$. In producing a token of the same type as C here, we have in effect repeated C. But we have repeated C in a new reflective context, in which we recognize C's pathological character, and no longer provide reference conditions via the 'refers_C' schema. Instead we provide reference conditions for C via the reference schema

'e' refers_R to y iff $e = y$,

where 'refers_R' abbreviates 'refers in the context of utterance of R'. If we instantiate to C, and put ' $\pi + 6$ ' for y , we obtain a true right-hand side, and so a true left-hand side. C refers_R to $\pi + 6$.

As with the Liar, we may capture the pathology of C via the notion of ungroundedness: among the referring expressions that C makes reference to is C itself.⁵⁵ On behalf of the disquotationalist, let us suppose again that ungroundedness can be characterized in disquotational terms. Suppose in particular that the applications of DefRef to the occurrence of 'refers_C' in C leads to the occurrence of 'refers_C' in the definiens, and so on repeatedly. In this way, the disquotationalist can capture the pathology of (C). But now the same general problem presents itself as before. If we accept the contextual analysis, the strengthened reasoning shows us that there are expressions, like C, that do refer, and yet contain ineliminable occurrences of 'refer'.⁵⁶ We arrive at a reference for C only by recognizing certain *semantic* features of C. In parallel with the strengthened Liar, the reasoning here establishes certain essentially semantic facts about C (that it is pathological, that it does not refer_C), and it is in virtue of these semantic facts that (C) refers. We cannot represent these facts or states of affairs in non-semantic terms. Reference is not eliminable.

7. CONCLUDING REMARKS

I have argued that the deflationist will find little comfort in any of the approaches to the Liar we have considered. And prospects look no better if we cast our net a little wider. For example, Barwise and Etchemendy take truth to be a genuine property propositions may have or fail to have, and as a result "... we must abandon anything like the redundancy theory of truth".⁵⁷ According to their 'Austinian' treatment, given any situation, there is a Liar sentence that says its falsity is a fact about that situation.⁵⁸ And according to their treatment of the Strengthened Liar, we may recognize the falsity of such a Liar proposition, step back and express the fact that the Liar proposition is false.⁵⁹ The notions of truth and falsity cannot be eliminated from the expressions of such semantic facts. Again, nothing in this compromises the correspondence conception.

Or consider the dialethic approach, according to which a Liar sentence, say

(1) (1) is false,

is both true and false.⁶⁰ If these are attributions of disquotational truth and falsity, then the disquotational schemas generate a contradiction. Presuming that the disquotationalist is willing to accept this, the problem of ineliminability remains. Any attempt to eliminate the semantical predicate from (1) goes against the dialetheist grain: it is just because (1) attributes falsity to itself, and so generates a semantic paradox, that the dialetheist declares it true and false. We cannot explain what the truth (and falsity) of (1) consists in independently of the notion of falsity.

We might cast our net a little further in another direction. The problems that the Liar presents are not restricted to the disquotational version of deflationism. Consider Horwich's Minimal Theory for propositions. Here propositions, and not sentences or utterances, are the truth-bearers. The axioms of Horwich's minimal theory of propositions are all the propositions whose structure is

(E*) $\langle\langle p \rangle\rangle$ is true iff p),

where ' $\langle p \rangle$ ' is written for 'the proposition that p '.⁶¹ For example, one of the axioms is the proposition expressed by the sentence:

The proposition that *snow is white* is true iff snow is white.

According to Horwich, these axioms together constitute a complete theory of truth; no more needs to be added.

Now, as Horwich notes,⁶² the move to propositions seems to close off any appeal to gaps. As we've seen, the theory of Barwise and Etchemendy, though in terms of propositions, is not available to the deflationist. In general, I cannot see that the standard approaches to the Liar are any more available to Horwich's minimalist than to the disquotationalist.⁶³ And there is anyway a more fundamental problem for the minimal theory.

According to Horwich, certain instances of (E*) must be excluded, on pain of the Liar.⁶⁴ One would expect Aristotle's paradoxical utterance (L) to be one of these, but since it predicates 'true' of a sentence, it is not clear how the minimalist will treat it. Consider instead a version of the Liar more suited to the minimal theory. Suppose Jack unwittingly lands in paradox, saying:

What Jack is saying is not true.

In line with the minimalist theory for propositions, we take 'not true' here to apply to a proposition. The associated instance of (E*) is:

- (e) The proposition that what Jack is saying is not true is true iff what Jack is saying is not true.

What should we make of (e)? What is the proposition referred to on the left-hand side?

If we adopt the contextual line, then, with reflective evaluations in mind, we might say that Jack expressed a true proposition (namely that what Jack said is not true in its context of utterance). And if we take the right-hand side to be true as well, (e) is true, and not contradictory at all. So the contextual theory of truth may lead us to the conclusion that (e) is not a problematic instance of (E*). Now another theory may identify another proposition, or none, as the subject of the left-hand side, and a different truth value for (e). And we might instead be led to the conclusion that (e) is a problematic instance of (E*). Different theories of truth may lead to different conclusions about the status of (e). We can only evaluate (e) if we are already in possession of a theory of truth. Given a Liar sentence, it is a highly non-trivial question as to what proposition, if any, it expresses. It takes a positive account of semantic paradox to determine the reference of "the proposition that *p*" when we put a Liar sentence for '*p*'.⁶⁵ If we couch *axioms* of a theory of truth in terms of the schematic expression "the proposition that *p*", then the game is over before it has begun.

The minimalist, then, is no better off than the disquotationalist. Of course, I have not considered every version of deflationism, nor every resolution of the Liar. But we have seen how inhospitable leading resolutions are to prominent forms of deflationism, and how resistant truth is

to elimination. It seems that there's no getting rid of truth – it cannot be disquoted away.⁶⁶

NOTES

¹ Correspondence theorists need not couch their account in terms of facts or correspondence to a state of affairs. A correspondence theorist may adopt a recursive Tarski-style theory of truth, and characterize the truth of a sentence in terms of connections between substantial elements and the world. See below for more on this kind of account.

² The denial that truth has no substantial nature can take more than one form. It might be the claim that truth has a nature, but a trivial one; versions of disquotationalism might be taken this way. Or it might be the claim that truth is not a property at all; consider for example the prosentential theory, according to which 'true' is not even a predicate.

³ Tarski (1944, p. 50). See also Tarski (1930–1931, p. 155). I should note that although some deflationists have drawn heavily from Tarski's work, it is far from clear that Tarski himself is a deflationist. But I shall not take up that issue here.

⁴ Quine (1970, p. 11).

⁵ *Ibid.*

⁶ Tarski (1930–1931, p. 159).

⁷ The disjunction might be finite, if the language or idiolect to which '*s*₁', '*s*₂', ... belong contains only finitely many sentences. But here, and throughout, the presumption will be that the language in question contains infinitely many sentences, and the corresponding disjunction is infinite.

⁸ The derivation of the infinite conjunction

(*s*₁ or not-*s*₁) and (*s*₂ or not-*s*₂) and ...

from the analytans is straightforward, given an infinitary logic.

⁹ Quine, *op. cit.* p. 12.

¹⁰ Such a characterization is suggested by remarks in Leeds (1978, pp. 121–131 and fn. 10); and versions of it are presented explicitly in Field (1986, p. 58), Resnik (1990, p. 412), and David (1994, Ch. 4 and p. 107).

¹¹ For example, put 'snow is white' for *x*. We will obtain just one true disjunct on the right-hand side. If we eliminate the false disjuncts, the definition yields:

'snow is white' is true iff 'snow is white' = 'snow is white' and snow is white. Dropping the true conjunct of the right-hand side, we obtain the T-sentence:

'snow is white' is true iff snow is white.

¹² Tarski (1930, p. 159).

¹³ Moreover, the component "*x* = '*p*'" is problematic if quantification is objectual. Put names of sentences for '*x*' and '*p*', and the component will identify a sentence with a name of a sentence.

¹⁴ Compare objectual quantification. Given '∃*x*(*x* is a dog)', we can give metalinguistic truth conditions: that there is an object that satisfies the open sentence '*x* is a dog'. But we can also provide an ordinary reading: "There are dogs". We are after an analogue of this ordinary reading in the case of ∑ *p*(*x* = '*p*' & *p*), and there is no obvious candidate. For a detailed discussion of this issue, see David (1994, p. 78–93); and for a broader objection to substitutional quantification in this general vein, see Van Inwagen (1981).

- 15 See Field (1986, pp. 57–59), and David (1994, p. 98).
 16 Problems for the recursive disquotationalist are discussed in David (1994, pp. 117–119), Field (1994, p. 269), and Horwich (1999).
 17 Of course, we will expect the T-sentences to be theorems of a correspondence theory of truth. But the correspondence theorist will do more than just affirm the T-sentences – she will also make substantial claims about the nature of truth.
 18 Horwich (1990, pp. 103–109).
 19 David (1994, p. 191), see also pp. 7 and 70.
 20 Horwich (1990, p. 42).
 21 Tarski (1930–1, p. 162). I omit the details of the derivation.
 22 See Horwich (1990, fn. 4 on pp. 19–20, and pp. 105–107).
 23 See Resnik (1990, pp. 413–414); in Resnik 1997, this restriction is eased.
 24 See Field (1994, p. 250), where the disquotational conception is characterized in terms of the sentences that a given speaker understands: "... for a person to call an utterance true in this pure disquotational sense is to say that it is true-as-he-understands-it. ... a person can meaningfully apply 'true' in the pure disquotational sense only to utterances that he has some understanding of ...".
 25 For example, consider a version of the Liar generated by

- (1) (1) is false.

Instantiating DefT to (1), and taking some obvious steps, we obtain

- (1) is false iff it's not the case that (1) is false.

This is contradictory – we have a case of strict paradox. And an infinite regress is also easily generated via repeated instantiations of DefT:

- (1) is false iff it's not the case that (1) is false iff it's not the case that it's not the case that (1) is false iff ...

26 Horwich (1990, pp. 41–42).

27 *op. cit.*, p. 41.

28 *op. cit.*, pp. 20–21, fn. 4; pp. 105–107).

29 As Kripke (1975) has pointed out, there is a riskiness to our ordinary uses of 'true': under suitably unfavorable circumstances, virtually any use of 'true', however innocent it may look, leads to paradox.

30 And it should not be thought that a theory of truth can be guided solely by the desire to maximize the scope of the truth definition, or to maximize the number of T-sentences we accept as axioms. Take a theory in which the diagonal lemma can be proved. Vann McGee has shown that, in the context of such a theory, there are many consistent sets of T-sentences that are maximal and mutually incompatible (see McGee, 1992).

31 Horwich (1990, p. 42).

32 Here is a proof. Take a gappy sentence, and suppose, towards a contradiction, that it is a member ' s_k ' of the substitution class. Given that s_k is not true, we obtain the negation of the infinite disjunction that forms the right-hand side of DefT. And from this we obtain $\sim s_k$. Similarly, given that s_k is not false, we obtain $\sim \sim s_k$. So we have $\sim s_k$ & $\sim \sim s_k$, a contradiction.

- 33 The revision theory of truth is developed in Gupta and Belnap, 1993. The revision theory of truth

is a consequence of combining a general theory of definitions and Tarski's suggestion that the biconditionals be viewed as partial definitions. Tarski's suggestion brings out clearly the circularity in the notion of truth ... (p. 142).

I should emphasize that the revision theory itself does not appeal to gaps.

- 34 Compare an argument of Dummett's (in Dummett, 1959), which arises in the course of examining Frege's claim that 'It is true that P' has the same sense as P:

Suppose that P contains a singular term which has a sense but no reference: then, according to Frege, P expresses a proposition which has no truth value. This proposition is therefore not true, and hence the statement 'It is true that P' will be *false*. P will therefore not have the same sense as 'It is true that P', since the latter is false while the former is not (p. 4).

- 35 Gupta and Belnap present this distinction in 1993 (p. 22), citing Yablo (1985).
 36 See Gupta and Belnap (1993, pp. 22 and 29, and fn. 52 on p. 29).

- 37 Strong truth is related to strengthened versions of the Liar, discussed below.
 38 Following our procedure, we assume that (2) is in S, and instantiate Def $\sim$ T to (2). Of

the infinitely many disjuncts on the right-hand side, only one will have a true identity as its first conjunct. It's easy to see that we obtain:

- (2) is not true iff it's not the case that (2) is not true.

This is contradictory – we have a case of strict paradox. And an infinite regress is also easily generated via repeated instantiations of Def $\sim$ T:

- (2) is not true iff it's not the case that (2) is not true iff it's not the case that it's not the case that (2) is not true iff ...

- (2) is ungrounded, as we would expect. (Compare the case of (1) in fn. 25.)

- 39 The equivalence may be demonstrated as follows. Consider:

- (1) $\sim[(x = 's_1' \& s_1) \vee (x = 's_2' \& s_2) \vee \dots]$

and

- (2) $[(x = 's_1' \& \sim s_1) \vee (x = 's_2' \& \sim s_2) \vee \dots] \vee \sim[x = 's_1' \vee x = 's_2' \vee \dots]$.

(1) is the definiens, and (2) symbolizes ' x is false or x is not in S '. Observe that (2) is equivalent to (3):

- (3) $[x = 's_1' \vee x = 's_2' \vee \dots] \rightarrow [(x = 's_1' \& \sim s_1) \vee (x = 's_2' \& \sim s_2) \vee \dots]$.

To show (1) $\rightarrow$ (2).

We'll show that (1) $\rightarrow$ (3). Suppose (1). Assume the antecedent of (3), and suppose $x = 's_k'$ for some k . Then, eliminating the false disjuncts from (1), we obtain

$$\sim['s_k' = 's_k' \& s_k]$$

and then

$$\sim s_k.$$

So

$$'s_k' = 's_k' \& \sim s_k.$$

And from this we obtain

$$[(x = 's_1' \& \sim s_1) \vee (x = 's_2' \& \sim s_2) \vee \dots],$$

and we're done.

To show (2) $\rightarrow$ (1). Suppose (2). Suppose first

$$(x = 's_1' \& \sim s_1) \vee (x = 's_2' \& \sim s_2) \vee \dots.$$

Then

$$x = 's_k' \& \sim s_k \text{ for some } k.$$

So

$$\sim (x = 's_k' \& s_k).$$

So

$$\sim [(x = 's_1' \& \sim s_1) \vee (x = 's_2' \& \sim s_2) \vee \dots],$$

and we're done.

Suppose second

$$\sim [x = 's_1' \vee x = 's_2' \dots].$$

Clearly

$$[(x = 's_1' \& \sim s_1) \vee (x = 's_2' \& \sim s_2) \vee \dots] \rightarrow [x = 's_1' \vee x = 's_2' \dots].$$

So we obtain

$$\sim [(x = 's_1' \& \sim s_1) \vee (x = 's_2' \& \sim s_2) \vee \dots],$$

and we're done.

40 Different versions of the contextual view may be found in Parsons (1974), Burge (1979), Barwise and Etchemendy (1987), Gaifman (1992), Simmons (1993).

41 For a detailed account of these contextual shifts, see Simmons (1993, pp. 102–106). The pragmatic implicature here is an instance of a more general implicature; as Burge puts it, "Sentences being referred to or quantified over are to be evaluated with the truth schema for the occurrence of 'true' in the evaluating sentence." (Burge, 1979, p. 95.)

42 Strengthened reasoning also presents problems for the prosentential theory of truth. According to the prosentential theory, 'true' is used in forming prosentences. In the discourse:

Mary: Chicago is large

John: If that is true, it probably has a large airport

the expression 'that is true' is a prosentence, which shares its content with its antecedent, namely 'Chicago is large'. In a parallel fashion, the Liar sentence 'This is false' relies on its antecedent for its content,

but it is, unfortunately, its own antecedent and, as such, fails as an antecedent supplier of content (Grover, 1992, p. 124).

According to the prosentential account, Liar sentences fail to have content.

Consider now the strengthened reasoning, and (P) in particular. In a discussion of strengthened reasoning, Grover remarks that "if 'true' is prosentential", then (P) "fails to express a proposition" (*op. cit.* p. 203). It is a consequence of the prosentential theory that (P) fails to have content, because according to the theory, (P) relies for its content on its antecedent and its antecedent (L) fails to have content. But it is surely highly counterintuitive that (P) is without content. Further, if we accept reflective evaluations like (R), then the prosentential account is closed off to us. If Liar sentences are true or false (on reflection), then they cannot be without content.

43 The Strengthened Liar will pack some surprise for the correspondence theorist (as it will for any truth theorist). On the correspondence account, (L) does correspond to a fact – the fact that it is not true_L. But we cannot relate (L) to this fact via 'true_L' – and here is the surprise. But the basic correspondence intuition remains intact – (L) is true (on reflection) in virtue of its correspondence to a fact.

44 The division is somewhat artificial. As we shall see, there are proposals that combine a limited appeal to gaps, context-shifts and the hierarchy.

45 The hierarchical approach owes us an account of how the level is established – but let us leave that to one side.

46 Of course, it will not help to expand the substitution class S_σ to include (J). For though we may then put (J) for x , the quote name for (J) will appear on the right-hand side, and the definition of 'true-in- L_σ ' will be circular.

Instead, the disquotationalist might keep S_σ as it is, and nevertheless allow the substitution of '(J)' for ' x ' in the definition of 'true-in- L_σ '. Then we may obtain:

$$(J) \text{ is true-in-} L_\sigma \text{ iff } (J) = 's_1' \& s_1, \text{ or } (J) = 's_2' \& s_2, \text{ or } \dots,$$

where each s_i is a sentence of L_σ . An apparent advantage of this move is that, since the right-hand side is false (in virtue of the false identities), the left-hand side is also false, which is the result we expect for Jane's Truth-Teller. Nevertheless the disquotationalist ought not to be tempted by this move. Apart from its *ad hoc* nature, it is quite against the spirit of disquotationalism. For it permits the application of 'true-in- L_σ ' to a sentence outside S_σ . But it is only the sentences in S_σ whose quote-names are disquoted in the definitions of 'true-in- L_σ '. So the application of 'true-in- L_σ ' to a sentence outside S_σ can never be accounted for disquotationally. In general, if 'true' applies to a sentence outside the associated substitution class, we cannot confer disquotational truth on that sentence.

47 Russell (1908), in van Heijenoort (1967, p. 166).

48 Russell (1903, p. 528).

49 Tarski (1930–1931, p. 267).

50 See especially Parsons (1974) and Burge (1979). The contextual idea can also be developed in the direction of a non-hierarchical 'singularity' account, according to which neither extension is more comprehensive than the other. According to the singularity theory, a use of 'true' applies almost everywhere, with the exception of certain singularities,

sentences to which it applies neither truly nor falsely. (For example, (L) is a singularity of the use of 'true' in (L).) See Simmons (1993).

⁵¹ See Horwich (1999). See also Horwich (1990, pp. 122–124). Of truth, reference and satisfaction, Horwich writes: "We should expect no deeper analyses of any of these semantic phenomena than are provided by their minimal theories ..." (Horwich, 1990, p. 53, fn. 3).

⁵² Field (1994, p. 261).

⁵³ For versions of this definition, see for example Resnik (1990, pp. 414–415), Horwich (1990, p. 124), Field (1994, p. 261), David (1994, p. 114), and Horwich (1999).

⁵⁴ The shifts in contextual parameters and pragmatic implicature are broadly similar to those in the strengthened Liar discourse.

⁵⁵ In Simmons (1994), ungroundedness is captured formally by reference trees that have infinite branches.

⁵⁶ As in the case of the Liar (see fn. 42), there is a problem here for the prosententialist as well as for the disquotationalist. In her account of the Berry paradox, Grover writes:

The phrase 'integer described in less than 19 syllables' should inherit from a set of antecedents, but because one of the antecedents happens to be the Berry description itself, there's an ungrounded inheritor. The Berry expression fails to refer."

But the strengthened reasoning about C suggests that a phrase can have itself as an antecedent, yet still refer. (For strengthened reasoning about the Berry phrase that runs parallel to the strengthened reasoning about C, see Simmons, 1994.)

⁵⁷ Barwise and Etchemendy (1987, p. 12).

⁵⁸ See *op. cit.*, pp. 124–125.

⁵⁹ See *op. cit.* p. 152, and also p. 138.

⁶⁰ See Priest (1979, 1984, 1987).

⁶¹ Horwich (1990, p. 18). There is a difficulty with Horwich's presentation of the minimal theory here. On p. 19, Horwich says that "[E*] is a function from propositions to propositions." I cannot see how it is. What do we put for each occurrence of 'p' in (E*)? Suppose we put a name of a proposition for each occurrence. Consider the first replacement: the angle brackets will be around the name of a proposition, and the inappropriate result is the name of a name of a proposition. Consider the second replacement – that will yield a name as the right-hand side of a biconditional. A natural suggestion is to abandon the claim that (E*) is a function from propositions to propositions, and treat (E*) substitutionally. Horwich, though, will reject this suggestion because substitutional quantification is characterized in terms of truth (see pp. 26–27).

⁶² *Op. cit.*, p. 80.

⁶³ This general point is encouraged by Horwich's claim that the minimal theory for propositions is equivalent to the minimal (or disquotational) theory of truth for utterances (*modulo* two further principles that Horwich finds uncontroversial – we can "easily derive" one theory from the other). See *op. cit.*, pp. 107–108.

⁶⁴ *Op. cit.*, p. 42.

⁶⁵ Davidson (1996) raises a related question for Horwich's theory:

How are we to understand phrases like 'the proposition that Socrates is wise'? In giving a standard account of the semantics of the sentence 'Socrates is wise', we make use of what the name 'Socrates' names, and of the entities of which the predicate 'is wise' is true. But how can we use these semantic features of the sentence 'Socrates is wise' to yield the

reference of 'the proposition that Socrates is wise'. Horwich does not give us any guidance here.

Perhaps, as Davidson suggests, there is a general problem for Horwich concerning the reference of 'the proposition that p'. But there is at least an acute problem when we put a Liar sentence for p.

⁶⁶ My thanks to Peter Alward, Dorit Bar-On, Simon Blackburn, Adam Constarbaris, Bill Lycan, Andrew Mills, Michael Resnik, and an anonymous referee for their valuable comments.

REFERENCES

- Barwise, Jon and Etchemendy, John: 1987, *The Liar*, Oxford University Press.
- Burge, Tyler: 1979, 'Semantical Paradox', *Journal of Philosophy* 76, 169–198; reprinted in Martin (1984, pp. 83–117). Page references are to Martin (1984).
- David, Marian: 1994, *Correspondence and Disquotational*, Oxford University Press.
- Davidson, Donald: 1996, 'On the Folly of Trying to Define Truth', *Journal of Philosophy* 93, 263–278.
- Dummett, Michael: 1959, 'Truth', *Proceedings of the Aristotelian Society* 59. Reprinted in Dummett, *Truth and Other Enigmas* (Duckworth, 1978). Page references are to *Truth and Other Enigmas*.
- Field, Hartry: 1986, 'The Deflationary Conception of Truth', in *Fact, Science, and Morality: Essays on A. J. Ayer's 'Language, Truth and Logic'*, edited by G. MacDonald and C. Wright. Oxford: Basil Blackwell, pp. 55–117.
- Field, Hartry: 1994, 'Deflationist Views of Meaning and Content', *Mind* 103(411), 249–285.
- Gaifman, Haim: 1992, 'Pointers to Truth', *Journal of Philosophy* 89, 223–261.
- Grover, Dorothy: 1992, *A Prosentential Theory of Truth*, Princeton University Press.
- Gupta, Anil, and Belnap, Nuel: 1993, *The Revision Theory of Truth*, MIT Press.
- Horwich, Paul: 1990, *Truth*, Oxford: Basil Blackwell.
- Horwich, Paul: 1999, 'Truth', in *Truth*, Oxford Readings in Philosophy, edited by Simon Blackburn and Keith Simmons, Oxford University Press, forthcoming.
- Kripke, Saul: 1975, 'Outline of a Theory of Truth', *Journal of Philosophy* 72, 690–716; reprinted in Martin (1984, pp. 53–81). Page references are to Martin (1984).
- Leeds, Stephen: 1978, 'Theories of Reference and Truth', *Erkenntnis* 13, 111–129.
- Martin, Robert L. (ed.): 1984, *Recent Essays on Truth and the Liar Paradox*, Oxford University Press.
- McGee, Vann: 1992, 'Maximal Consistent Sets of Instances of Tarski's Schema (T)', *Journal of Philosophical Logic* 21, 235–241.
- Parsons, Charles: 1974, 'The Liar Paradox', *Journal of Philosophical Logic* 3, 381–412; reprinted in Martin (1984). Page references are to Martin (1984).
- Priest, Graham: 1979, 'The Logic of Paradox', *Journal of Philosophical Logic* 8, 219–241.
- Priest, Graham: 1984, 'Logic of Paradox Revisited', *Journal of Philosophical Logic* 13, 153–179.
- Priest, Graham: 1987, *In Contradiction: A Study of the Transconsistent*, Nijhoff.
- Quine, W. V.: 1970, *Philosophy of Logic*, Prentice-Hall.
- Resnik, Michael: 1990, 'Immanent Truth', *Mind* 99(395), 405–424.
- Resnik, Michael: 1997, *Mathematics as a Science of Patterns*, Oxford: Clarendon Press.

- Russell, Bertrand: 1903, *The Principles of Mathematics*, Cambridge University Press.
- Russell, Bertrand: 1908, 'Mathematical Logic as Based on the Theory of Types', *American Journal of Mathematics* 30, 222-262; reprinted in van Heijenoort (1967), pp. 150-182).
- Simmons, Keith: 1993, *Universality and the Liar*, Cambridge University Press.
- Simmons, Keith: 1994, 'Paradoxes of Denotation', *Philosophical Studies* 76, 71-106.
- Tarski, Alfred: 1930-1931, 'The Concept of Truth in Formalized Languages', in Tarski, *Logic, Semantics, Metamathematics*, 2nd edition (Hackett, 1983).
- Tarski, Alfred: 1944, 'The Semantic Conception of Truth', *Philosophy and Phenomenological Research* 4, 341-375; reprinted in A. P. Martinich (ed.), *The Philosophy of Language*, Oxford University Press (1985). Page references are to Martinich (1985).
- van Heijenoort, Jean: 1967, *From Frege to Gödel*, Harvard University Press.
- van Inwagen, Peter: 1981, 'Why I don't understand substitutional quantification', *Philosophical Studies* 39, 281-285.
- Yablo, Stephen: 1985, 'Truth and Reflection', *Journal of Philosophical Logic* 14, 297-349.

*University of North Carolina at Chapel Hill,
Department of Philosophy,
Chapel Hill, NC 27599-3125, U.S.A.*