The Liar and the Inconsistency of Language

John Barker

A fairly common assumption about the Liar paradox is that it is a formal problem that can be safely delegated to mathematical logicians, and that its proper resolution can be trusted not to disrupt anyone else’s philosophical theories and projects. Serious students of the Liar know better. While purely formal approaches to the Liar abound, it is not a purely formal problem. Philosophical interpretations of these formal approaches nearly always have non-trivial philosophical consequences. And one of the most striking of these interpretations is Tarski’s¹ assertion that natural languages are inconsistent. In this paper, I want to explore the idea that Tarski was right.

The term ‘Liar paradox’ is the name traditionally given to the fact that the following schema, while very appealing, is logically inconsistent:

(D) 'P' is true if and only if P.

To see that (D) is inconsistent, simply substitute, for P, the following sentence:

(L) Sentence (L) is not true.

Upon substitution of (L) for P, (D) becomes

‘Sentence (L) is not true’ is true if and only if sentence (L) is not true.

Using the empirically obvious fact that sentence (L) = ‘Sentence (L) is not true’, we can deduce

Sentence (L) is true if and only if sentence (L) is not true

which is (classically) inconsistent.

There are (at least) three possible approaches to this problem. The first and most common approach is to view (D) as part of an intuitively appealing but ultimately mistaken folk theory of truth, often called the “naïve theory of truth.” From this perspective, the Liar paradox is simply a reductio of the naïve theory. The naïve theory is intuitively plausible, but it is also contradictory.

¹ See (Tarski 1935).
and must therefore be false, plain and simple. Of course, this raises the question: if the naïve theory is wrong, which theory is right? The philosophical study of the Liar is littered with failed attempts to find this right theory, some of which I will outline below.

The second approach holds that (D), and the naïve theory it represents, is the correct theory of truth after all, notwithstanding that it is inconsistent. Since (D) has contradictions among its logical consequences, this view naturally implies that some contradictions are true. For similar reasons, such a view holds that Liar sentences, such as (L), are both true and untrue. Because of this last feature, such views go by the name of *dialetheism*. Dialetheism certainly has something going for it, in embracing the intuitively appealing naïve theory and rejecting complicated and unintuitive substitutes. But its acceptance of true contradictions and truth value gluts makes it hard for most philosophers to swallow.

These two approaches may seem to exhaust the possibilities. After all, either (D) is true or it is not true. However, there is a third approach, going back to (Tarski 1935). On this approach, (D) is in some sense posited by the language we speak. The Liar paradox is a byproduct of an inconsistency in the rules of language. Whether (D) is true or not is an illegitimate question; the important thing about (D) is not its truth value, but its relation to the rules of language. This approach has been around for a long time, but has not received much systematic development.

This paper defends this third approach. Basically, I will argue that the mainstream approach has a very poor track record, and that the present approach avoids its problems while avoiding dialetheism’s pitfalls. However, there is more at stake than the resolution or diagnosis of the Liar paradox. The larger question is: do languages have semantic rules, and does it make sense to say that those rules are inconsistent? That is by no means an easy question, but I think it is about time we asked it, whether we are interested in the Liar or not.

### 1. Why the Liar Is Hard

The mainstream approach to the Liar – the attempt to correct the naïve theory – has yielded an enormous number of proposed theories of truth, with little or no consensus about which, if any, is right, and with a general sense that none is entirely adequate. The reason for this is not formal,

---

2 The leading proponent of dialetheism is Graham Priest; see e.g. (Priest 1974) and (Priest 1987).  
3 The most notable attempt to develop such a theory is (Chihara 1978). The present paper is largely inspired by Chihara, in principle if not in detail.
but philosophical. Formally, one simply has to formalize the Liar reasoning of the last section and reject one or more premises or inferences of the argument. Intuitively, it is very hard to see one’s way clear to doing so.

The fundamental reason for this is what is known as the revenge problem (also known as the strengthened Liar problem). Consider a Liar sentence, say $S = ‘S$ is not true’. The corresponding instance of (D) is ‘$S$ is true if and only if $S$ is not true’, a (classical) contradiction. Now it might appear to be open to us to reject the $S$-instance of (D), and thereby avoid the contradiction, for any number of reasons. For example, we might think that $S$ is lacking in content, and one might reasonably think that (D) applies only to contentful sentences. In that case, $S$ should certainly be regarded as truth-valueless. But this statement, that $S$ is truth-valueless, gets us into trouble very quickly. If $S$ is truth-valueless, then it is certainly not true. So we have now asserted, as a simple consequence of this account,

(1) $S$ is not true.

But (1) is simply the sentence $S$ itself. We have wound up asserting the very sentence $S$ that we thought was contentless or otherwise truth-valueless.

Now arguably it is propositions, not sentences, that are true or false, and it is natural to suspect (or at least hope) that the Liar paradox will go away once we apply this fact consistently. In particular, Liar sentences might be dismissed as failing to express propositions; and this move might seem to block the Liar altogether, since it is not obvious that one can construct self-referential propositions the way we construct self-referential Liar sentences. But this move simply won’t work; the revenge problem cannot be blocked in this manner. To see this, consider a sentence $S = ‘S$ does not express a true proposition.’ $S$ is certainly a Liar sentence, and so, on the view we are entertaining, it fails to express a proposition. But then a fortiori, $S$ does not express a true proposition. A proponent of the view in question is now in the position of assertively uttering a sentence—namely, ‘$S$ does not express a true proposition’, or in other words, the sentence $S$ itself—that is supposed to express no proposition at all.

The problem here is not that there is no defensible way to restrict the disquotational schema. In fact, there is broad agreement that not all grammatical sentences are truth-apt, and that (D) applies only to truth-apt sentences. The difficulty with exploiting this fact to block Liar reasoning – and the explanation of the revenge problem, in my view – is that asserted sentences must be assumed to be truth-apt. Or more cautiously, the act of asserting a sentence $S$ carries

\[4\] Or rather, that sentence is implied by the $S$-instance of (D), via the substitution of equals for equals.
with it a commitment to S’s truth-aptness, and indeed to S’s truth. In other words, it would seem that we are committed to the following inference rule:

(Assert) From S, infer Tr(‘S’)

The rule Assert applies only to sentences that have been asserted unconditionally (i.e., it does not apply to sentences that depend on hypotheses). It seems that the Assert rule should apply across the board, no matter what sentence S is used. If S is not truth-apt, then it should never be asserted in the first place, and Assert should never be invoked on it. And yet, Assert can generate Liar-type problems in more than one way.

First, suppose we supplement Assert with the following weakened version of (D):

(D-) Tr(‘S’) → S

(D-), like Assert, seems like it should apply without restriction; if S happens not to be truth-apt, then (D-)’s antecedent will simply be false, making (D-) trivially true. Yet (D-) and Assert are mutually inconsistent. To see this, we simply let S be a Liar sentence; for simplicity, assume as an additional axiom

(SR) a = ‘~Tr(a)’

(D-) and (SR) immediately yield Tr(a) → ~Tr(a), and hence ~Tr(a). Since this conclusion has been asserted, we may apply the Assert rule to get Tr(~Tr(a)); using (SR) again, we have Tr(a), contradicting our earlier conclusion ~Tr(a). Thus, (D-) and Assert cannot be jointly maintained.

Now this assumes that the underlying logic is classical, and in discussions of the Liar, this assumption is often questioned. In particular, (D-) might be rejected in a non-classical context, since it may be gappy, not true, when the sentence S is gappy. However, since gappy sentences are not assertable, there is no need to place a similar restriction on Assert. Now Assert, by itself, is consistent, even if the underlying logic is classical, let alone if it is weaker than classical logic. However, there is now a problem about how to express the idea that a Liar sentence, like ~Tr(a), is gappy. To see why there is a problem, consider first the most obvious way to say that a sentence is gappy. It is simply to say that it is neither true nor false: i.e., that neither it nor its negation is true. In particular, then, that a given sentence S is gappy logically

---

5 The seminal work in the truth-value-gap approach to the Liar is of course (Kripke 1975), though the idea existed long before.
implies that it is not true. Now apply this to the sentence \( \neg \text{Tr}(a) \). To claim that that sentence is gappy entails that it is not true – that is, it entails \( \neg \text{Tr}(a) \). So in claiming that \( \neg \text{Tr}(a) \) is gappy, one is committed to, and should be willing to assert, \( \neg \text{Tr}(a) \). But of course any such assertion is immediately rendered inconsistent by the Assert rule. Thus, while we may be convinced that the Liar sentence \( \neg \text{Tr}(a) \) is gappy, the most obvious way of saying that it is gappy is self-defeating.

Of course, saying that the Liar sentence is neither true nor false is merely the *simplest* way of expressing the idea that the Liar is gappy. Other methods exist, such as making a distinction between internal and external negation, or using a special predicate meaning “gappy” or “not definitely true.” However, such devices do not really solve the problem, because they can be used to construct new Liar sentences which are immune to the analysis they are meant to support. For example, “external” negation is an operator that turns a gappy sentence into a true sentence, whereas ordinary “internal” negation turns a gappy sentence into a gappy sentence. Suppose we introduce a new, external negation sign \( ! \). Using that operator, we can easily and consistently express the idea that \( \neg \text{Tr}(a) \) is gappy: we simply say \( !\text{Tr}(\neg \text{Tr}(a)) \) & \( !\text{Tr}(\text{Tr}(a)) \). While this sentence directly implies \( !\text{Tr}(a) \), there is no problem asserting the latter sentence, because it is strictly weaker than \( \neg \text{Tr}(a) \). However, this victory is short-lived, because we can simply form a new Liar sentence

\[
\text{(SR2)} \quad b = '!\text{Tr}(b)'
\]

It is easy to see that \( !\text{Tr}(b) \) is not assertable. However, we now have no acceptable way of saying that \( !\text{Tr}(b) \) is gappy; using external negation to assert that \( !\text{Tr}(b) \) is gappy would immediately imply that it is not (external) true, i.e., it would imply \( !\text{Tr}(b) \). Once more, we are faced with either inconsistency or expressive limitation, in the form of an inability to diagnose the Liar sentence at hand, depending on what formal principles we accept. And finally, this situation appears to be quite generic, as any vocabulary, logical or non-logical, that we may use to theorize about truth and Liar sentences can also be used to generate new Liars.

As one final example of the revenge problem, let us consider an approach that was specifically designed to avoid the above problem. On some accounts, the natural language predicate ‘true’ actually has a variable extension, i.e., it is indexical. Assertable sentences are indeed true, as the Assert rule suggests. However, a sentence that is assertable in a given context

---

6 For the distinction between truth and definite truth as it relates to the Liar, see especially (McGee 1991) and (Soames 1999).
7 See especially (Burge 1979). The general idea of context dependency as a solution to the Liar has been tried in many different ways; see, e.g., (Russell 1908), (Parsons 1974), (Barwise and Etchemendy 1984), (Gaifman 1992), (Soames 1999), and many others.
may not belong to the extension that ‘true’ has in that context. Rather, asserting that the sentence in question is true might require a new context, in which ‘true’ has a larger extension. Thus, when we apply the Assert rule to a sentence S, we may be shifting the context in such a way that the extension of ‘true’ expands. Post-expansion, when we assert the conclusion ‘S is true’, S belongs to the truth predicate’s extension. Pre-expansion, it may not. This effectively avoids the problem described two paragraphs back. The proper way to describe a Liar sentence like (L) is to say: ‘(L) is not true’. That statement is assertable, but may not be true (i.e., belong to the extension of ‘true’) in the context in which it is asserted. Nonetheless, a new context is always available in which we may correctly assert that ‘(L) is not true’ is true. No contradiction results, because these two assertions take place in different contexts in which ‘true’ has different extensions. The assertions do not contradict each other any more than do ‘It is warm here’ (spoken in Ecuador) and ‘It is not warm here’ (spoken in Greenland).

Does this trick really avoid the Liar problem? Putting aside the matter of why we should think ‘true’ is indexical in the first place, we must consider the revenge strategy of using the account’s own machinery to create new Liars. In this case, the basic machinery of the account is the notion of the context-dependent extension of ‘true’. In other words, the basic primitive of the account is:

Sentence S belongs to the extension of ‘true’ in context c.

And this primitive can easily be exploited to create a new Liar. Specifically, we have

(S+) (S+) does not belong to the extension of ‘true’ in any context.

Assuming we have done nothing illicit so far, it is fairly easy to see that (S+) is not true at all, that it does not belong to the extension of ‘true’ in any context. Unfortunately, that (S+) does not belong to the extension of ‘true’ in any context is exactly what (S+) says. In other words, (S+) is assertable, but not true: not true at all, in any context. Since it is a basic precept of the contextual account that assertable sentences are true in some context or other, the case of (S+) is unacceptable by the account’s own lights.8

---

8 There is much more to be said here. The most common strategy for avoiding problems like (S+) is to claim that the latter is not well formed, because the ‘c’ position in ‘S is true in context c’ cannot be bound by a quantifier as it is in (S+). In other words, ‘S is true in context c’ does not express a simple binary relation between sentences and contexts; it is something more akin to a schema, with ‘c’ a schematic letter. I strongly suspect that this strategy cannot block the revenge problem forever. To the extent that we can generalize over contexts – as we must to some extent
I suspect that the revenge problem is unavoidable, that for any account of the Liar, the concepts that the account employs, and the terminology that expresses them, can be used to construct a new Liar sentence that the account simply cannot handle: it will yield either outright inconsistency or severe expressive limitation. Obviously, demonstrating this assertion would involve an extensive review of existing approaches and would still not be quite conclusive, since it would not include as-yet-unthought-of approaches. Nonetheless, the difficulty of the revenge problem should at least give us pause.

So why not at least consider a simpler hypothesis? What the revenge problem boils down to is that it is very hard to find an alternative to (D) that is intuitively acceptable, in the sense of being a plausible candidate for what ordinary language users learn when they learn the meaning of ‘true’. So why not entertain the idea that it is (D), inconsistent as it is, that ordinary language users learn when they learn ‘true’?

2. The Proposal

The view I want to defend is based on the idea that (D) – inconsistent though it is – is what we mean by ‘true’. Or to put it more bluntly, I want to argue that (D) is analytic. The mainstream approach of trying to improve on the naïve theory of truth is destined to failure (on this view), because the naïve theory is analytic and thus cannot be improved on without changing the meaning of ‘true’. Following Chihara, I call this view the inconsistency theory of truth.

As I noted earlier, this idea has been around for a long time, but it has never been particularly popular. One of the most basic criticisms is that it simply fails to adequately distinguish languages from theories. For example, Burge writes:

According to [Tarski's view], part of the nature of a “language” is a set of postulates that purport to be true by virtue of their meaning or are at least partially constitutive of that “language.” Tarski thought that he had identified just such postulates in natural language as spawning inconsistency. But postulates are contained in theories that are promoted by people. Natural languages per se do not postulate or assert anything. What engenders paradox is a certain naive theory or conception of the natural concept of truth. It is the business of those interested in natural language to improve on it. (Burge 1979, pp. 83–4)
That is, the inconsistency theory views (D) as a posit or axiom of our language; but it is theories, not languages, that have axioms or posits.

But is it obvious that languages do not posit or assert anything? By most accounts, languages do include analytic sentences. Of course, the very existence of analytic sentences has been in dispute ever since Quine’s critique of the analytic/synthetic distinction in (Quine 1951). Nonetheless, a good many philosophers are prepared to acknowledge a limited number of analytic sentences, provided they are trivially analytic and do not do any heavy theoretical lifting. A large part of the reason that the analytic/synthetic distinction fell into disrepute is that it was being misused: contentious philosophical theses, as well as foundational theses of scientific theories, were held to be analytic, mostly because they are not straightforwardly empirical. It is safe to say that few philosophers would now classify such statements as straightforwardly analytic, nor would they accept the positivistic analytic/synthetic dichotomy that grounds such a classification. But none of this shows that there is no such thing as an analytic sentence. It seems obvious to me that ‘An uncle of x is a brother of a parent of x’, for example, is completely different from ‘Grass is green’. The former sentence holds because of the logic of kinship terms; the term ‘analytic’ seems apt here. At any rate, this view of analyticity is, I think, a fairly common one, and I will assume it in what follows.

Now at first blush, the instances of (D) seem every bit as analytic as ‘An uncle of x is a brother of a parent of x’. They seem to hold in virtue of the meaning of ‘true’ just as the latter seems to hold in virtue of the meaning of ‘uncle’. And this impression persists through subsequent bushes. And this fact is, fundamentally, why the Liar is so hard. We could avoid contradiction by simply dropping (D), or dropping enough instances of (D). But doing so is intuitively unacceptable, unless we plan on changing the meaning of ‘true’. All of this argues for regarding (D) as analytic.

However, (D)’s instances are different from ‘An uncle of x is a brother of a parent of x’ in one crucial respect: some of them are logically inconsistent. And this is no small difference, since there is a problem in principle with the idea that an inconsistent sentence could be analytic. Namely, an analytic sentence is usually defined as a sentence that is true by virtue of its meaning. Thus, whatever else it may be, an analytic sentence is true. But if an analytic sentence is also contradictory, then it follows that some sentence is both true and contradictory. If we define a contradictory sentence as one whose negation is analytic, then the existence of an analytic contradictory sentence entails that some sentence S and its negation ¬S are both true. And finally, if a sentence ¬S is true, then it follows that S is not true: this is obvious in any case, but especially so in our case since it is a direct consequence of (D). Thus, if any sentence is both analytic and contradictory, then some sentence is both true and not true. And this is a
contradiction: the claim that a contradictory sentence can be analytic is itself a contradictory claim.\(^9\)

Interestingly, not everyone finds this unacceptable. Dialetheists, such as Graham Priest, simply accept (D) and live with the contradiction. Priest regards natural language as governed by a non-classical “paraconsistent” logic that admits true contradictions but blocks the classical inferences that would otherwise show that everything follows from a contradiction. As the above argument shows, and as Priest admits, when this idea is applied to natural language, we can’t keep the inconsistency safely confined to the object language while we live happily in a classical metalanguage. The inconsistency infects the metatheory, making dialetheism itself an inconsistent theory\(^10\). The only question is how much of a criticism of dialetheism this observation amounts to.

Like most philosophers, I accept the conventional view that inconsistency is a pretty decisive objection against any theory. Thus, I adopt the following as a basic ground rule: the theory of truth to be developed, even though it describes an inconsistency in natural language, must not itself be an inconsistent theory.

Thus, we need to find a better way of capturing the idea that (D) is “what we mean by ‘true’”. This is no easy task, and there are certainly more ways of getting it wrong than right. For example, a natural first stab is the following:

(A1) Competent language users accept (D).

While (A1) is surely true, it is hardly adequate. If ‘accept’ is merely a synonym of ‘believe’, then (A1) reduces to the claim that the disquotational schema is widely believed. And to say this is far from saying anything about the inconsistency of language or of the truth concept. Indeed, it does nothing more than say that there is a widely believed naïve theory of truth.

Much the same objection applies to the following:

(A2) Language users have a disposition to assent to (the instances of) (D).

---

\(^9\) Herzberger used essentially this argument against the Tarskian view of an inconsistent language; see (Herzberger 1966) and (Herzberger 1967).

\(^10\) In paraconsistent logic, one must distinguish two different kinds of inconsistency. A theory is weakly inconsistent if it has both A and \(\sim A\) as theorems, for some sentence A, and strongly inconsistent if every sentence in the language is a theorem of it. Classically, the two notions coincide; but in paraconsistent logic, a theory can be weakly inconsistent without being strongly inconsistent. Unless otherwise noted, I will mean ‘weakly inconsistent’ by ‘consistent’.
At best, (A2) is compatible with a thesis of widespread inconsistent belief, and says nothing about whether the language itself, or the concept of truth, is inconsistent. But it gets worse, because it is far from clear that language users have any such disposition. When presented with an instance of Liar reasoning, a competent language user might react in any number of ways, and in any case is unlikely to embrace the contradictory conclusion. In particular, she may, for any number of reasons, refuse to assent to the relevant instances of (D), if only because she does not know what to make of them.

This should hardly surprise us, since facts about meaning have relatively little to do with dispositions to assent in any case. Someone can have, or lack, a disposition to assent to a particular sentence S for all sorts of reasons, not all of which have much to do with S’s meaning\textsuperscript{11}. And if there are sentences that are, in some sense, both analytic and contradictory (in some suitable sense of those terms), this fact will, if anything, only widen the gap. If a sentence S is analytic, then that fact gives speakers some reason to assent to S. If S is contradictory, then that fact gives speakers some reason to withhold assent from S. And if S is somehow both, then speakers will have both a reason to assent and a reason to dissent, and different speakers will resolve this conflict differently.

So the claim that language users accept (D), even if true, is too weak for our purposes. We might try to strengthen it as follows:

(A3) Acceptance of the disquotational schema is a necessary condition for being a competent user of the word ‘true’.

But (A3) is probably too strong, at least depending on how we understand ‘accept’. We can easily imagine a competent language user who, on considering some tricky instance of the disquotational schema, decides to suspend judgment. Such a speaker certainly no longer “accepts” that instance, at least for all we have said about acceptance, but still means true by ‘true’.

I think an entirely different approach is called for. We are looking for a feature of analytic sentences that does not automatically entail their truth. One such feature is as follows: by speaking a language, one becomes committed to its analytic sentences. Whether one chooses to assent to these sentences is irrelevant; because they are analytic, one ought to assent to them. Commitment is a normative property; it cannot be (easily) cashed out in terms of a disposition to assent. Commitment, I suggest, is the very property we are looking for. What is special about (D) is precisely this: our language commits us to it.

\textsuperscript{11} (Kripke 1982) contains a well-known discussion of this point.
The idea of language-generated commitment is a fairly natural one, if we take seriously the idea that the facts of language are determined by convention. A convention, being a mutual agreement, generates obligations; in speaking a language, we agree to use that language in a certain way, and thus acquire a commitment to using it in that way. Thus, I am suggesting that (D) represents a language-defined standard for the correct use of ‘true’.

But can a language, in principle, generate a commitment to a contradictory sentence? That it can is suggested, I would argue, by the following thought experiment. Suppose the word ‘tall’ is not currently used by a given language, and that its speakers wish to introduce it by means of an explicit stipulation. First, they might make the following stipulation:

(Tall1) Something is tall if and only if it is at least 5'11" in height.

(Tall1) creates a new predicate ‘tall’ with a precise extension. Alternatively, they might stipulate as follows:

(Tall2) Something is tall if it is at least 6', and not tall if it is less than 5'10".

(Tall2) is a perfectly good way of introducing a predicate, but unlike (Tall1), it under-specifies the extension of that predicate; the predicate is, arguably, vague. Finally, suppose our hypothetical language community made the following stipulation:

(Tall3) Something is tall if it is at least 5'10", and not tall if it is less than 6'.

Whereas (Tall1) specifies a set of individuals precisely, and (Tall2) specifies a set of individuals vaguely, (Tall3) specifies a set of individuals inconsistently. That is, (Tall3) attempts to specify a set of individuals, but fails to do so because it is inconsistent.\textsuperscript{12} Now the natural reaction to (Tall3) is that, unlike (Tall1) or (Tall2), it is not a legitimate way of defining a predicate. In a certain sense, this is clearly right. But suppose our language community adopted (Tall3) as a stipulation anyway, with the intention of introducing a new predicate and with the belief that they had successfully done so. What can we say about the language they now speak, and about the role of ‘tall’ in that language?

Similarly, and more to the point, suppose ‘G’ is not currently being used in our imagined language, and that its speakers wish to introduce it by stipulating the following schema:

\textsuperscript{12} Strictly speaking, of course, (Tall3) is not inconsistent. It simply implies that there are no individuals between 5'10" and 6' in height, and is thus inconsistent with the empirical fact that there are such individuals.
‘S’ is G if and only if S.

(G), like (Tall3), is inconsistent; after all, it is formally identical to (D). But unlike (Tall3), (G)’s inconsistency could easily go unnoticed. In either case, if the inconsistency is unnoticed, the stipulation could form the basis of a more or less well-defined use of the relevant predicate within the language community. Now let us make some observations about the case.

1. In each case, the linguistic community acquires a commitment to the sentence or schema that is being stipulated. It is hard to see how it could be otherwise: by making a stipulation, the community is explicitly holding its use of language, and specifically its use of the introduced predicate, to be bound by that stipulation.

2. In each case, the commitment in question is a convention of language, or at least is generated by the conventions of language. The community’s intention is to add a predicate to their language, not to specify the rules of some non-linguistic practice.

3. While the community acquires a commitment by making a stipulation, the commitment depends on the use of language that the stipulation inaugurates, and not directly on the stipulation itself. Thus, if (Tall3) or (G) were uttered stipulatively but failed to produce the expected language use – if the community failed to follow through on its original intention – then the commitment to (Tall3) or (G) would not be in effect. On the other hand, if the original act of stipulating (Tall3) or (G) were forgotten within the language community, but the corresponding use of language persisted, then so would the commitment to (Tall3) or (G).

If these three observations are correct, then I think the case has been made for the possibility, in principle, of a language-generated commitment to a contradictory sentence. I will defend this claim in greater detail below; for now, I want to focus on further clarifying the view.

First, while the above examples focused on explicit stipulation, I am not claiming that the semantics of ‘true’ is governed by any such stipulation. We do not, and never did, stipulatively define ‘true’ via (D) or anything like it. The claim is simply that (D) is a commitment generated by the language we speak, or a linguistic commitment for short. Explicit stipulation is one way to generate such commitments. Ordinary analyticity is another. In either case, it is the use of language that generates the commitment, whether that use is inaugurated by a stipulation like (Tall3) or (G), or is simply tacitly accepted by language users.

In this connection, it is important to get at least somewhat clear on what we mean by the “use” of language. That is obviously a very difficult question, but we can at least say this much. When we speak of the use of language, we are not simply referring to the sum total of linguistic behavior. We are speaking of the standards of correct behavior that are accepted by a given language community. The concept of language use is a normative, not descriptive, concept, as is
the concept of linguistic commitment. Of course, normative facts about language use may ultimately be grounded in descriptive facts about language users, but that is not my concern in this paper. I am simply making a claim about language use in the normative sense of ‘use’.

I think it is fairly natural to suppose that a social convention, linguistic or otherwise, can give rise to conventions. However, it may be less clear what it means for a linguistic convention to commit speakers to a sentence. One way to become committed to a sentence S is to believe or assert a proposition, or set of propositions, that implies the proposition expressed by S. This might be an appropriate notion of commitment for our purposes, provided a number of conditions are satisfied. First, linguistic conventions would have to be understood as, in part, mutual agreements about what to believe; then a linguistic convention could generate commitment to one or more beliefs that imply (D), say. Second, it would have to be maintained that the sentences in question express well-defined propositions. For example, in the language into which ‘tall’ has been introduced via (Tall3), sentences involving ‘tall’ would have to express well-defined propositions. Third, these propositions would have to bear well-defined implication relations to each other.

Now all of these conditions could be defended, but they do seem somewhat questionable in the present case. In particular, since ‘tall’ in our example seems to lack a well-defined extension, one may question whether sentences involving ‘tall’ really express propositions that bear well-defined implication relations to each other. Fortunately, there is a more straightforward conception of linguistic commitment.

First, an analogy. Suppose you and others have adopted a certain formal deductive system for some purpose, and that this system includes an inference rule R. Now suppose a formal sentence S is derivable, via R, from sentences you have already asserted. You are now licensed to assert S, i.e., you have a permission to assert S. However, it seems that in addition to this permission, you also have an obligation. The conclusion S is not optional for you at this point. Were someone else to assert S, you would be forced to accept that assertion, given what you have already asserted. You are not merely licensed to assert S; you are committed to S, in a way that is hard to cash out in terms of anything more basic but which seems to have something to do with assertions and obligations. You are, in some sense, obligated to assert S; or rather, you are obligated to assert S if the alternative is denying S or explicitly withholding judgment. And you are obligated to accept other people’s assertions of S.

This idea is not without precedent. See, e.g., (Lewis 1983), where a convention of language is held to be a convention of truthfulness and trust in an abstract language L. To be trustful-in-L is to believe that other people’s utterances are true-in-L; thus, trust-in-L is a pattern of belief, and on Lewis’s account, this pattern of belief is a convention.
Now language is not a formal deductive system, but the analogy seems appropriate. To be linguistically committed to a sentence $S$ is, roughly, to be committed to assertively uttering $S$, at least under suitable conditions. To be linguistically committed to an inference rule is to be linguistically committed to their conclusions once one asserts, or becomes committed to, their premises. A schema, such as (D), can be viewed either as a set of sentences (the instances of the schema) or as an inference rule with no premises; either way, linguistic commitment to it amounts to linguistic commitment to its instances.

I have already suggested that linguistic commitment is closely related to analyticity. Since the “true by virtue of meaning” interpretation of analyticity is so well-entrenched, let us coin a new term and call a sentence quasi-analytic\footnote{My use of “quasi-analytic” should not be confused with that of (Tappenden 1993). Tappenden defines a sentence to be quasi-analytic in a language if it is not false on any admissible precisification of the language’s predicates.} in L if the conventions of L generate a commitment to it. Likewise, we may call a sentence quasi-contradictory in L if L’s conventions generate a prohibition from asserting L. Now the inconsistency theory can be stated more sharply: it is possible for a sentence to be both quasi-analytic and quasi-contradictory in a language, and some instances of (D) are in fact both quasi-analytic and quasi-contradictory. In other words, speakers of English have a linguistic commitment to (D), and if ‘true’ has the same meaning in your idiolect that it does standardly, then this fact linguistically commits you to (D) as well, inconsistency and all.

Now that the inconsistency theory is on the table, let us review its rationale. Its primary rationale, of course, is the persistent difficulty of the mainstream approach to the Liar, more specifically the revenge problem. If one approach appears problematic, then it behooves us to consider alternatives. In fact, I suspect that the primary motivation for the mainstream approach is the perception that there are no alternatives, that the correct theory of truth, in the form of a consistent refinement of (D), must be out there and that it is simply our job to find it. According to the inconsistency theory, that is not the case at all. It is impossible to improve upon (D) while remaining true to the ordinary concept of truth, because (D) is what ‘true’ means; a replacement would, at best, represent a revision of the ordinary concept of truth.

Another consideration in favor of the inconsistency theory is simply that (D) seems analytic, and that this is, all else being equal, an excellent reason to think that it is analytic. The Liar is a problem because Liar reasoning seems valid, even though it results in a contradictory conclusion. This, in turn, is due to the fact that the instances of (D) seem mandatory, and that denying them would be contradictory. As I mentioned earlier, there is no formal problem about avoiding the contradictory conclusions of Liar reasoning: one simply has to reject enough instances of (D). The trouble is that intuitively, doing so seems no more acceptable than denying
a truth of logic. And most importantly, this perception remains even when the relevant instances of (D), and the Liar reasoning in which they are invoked, are brought into focus. I think our linguistic intuitions are telling us something here. We make judgments of analyticity based on linguistic intuition; and if we apply the same standard to (D), we will naturally classify it as analytic also. Of course, such considerations are defeasible and further consideration might yet lead us to reject (D)’s analyticity. However, the only apparent reason to do so is, again, the perception that we must do so, for the reasons spelled out in Herzberger’s argument. The inconsistency theory offers an alternative. The correct interpretation of our linguistic intuitions about (D), I suggest, is that it is quasi-analytic.

Finally, the inconsistency theory has the great virtue of simplicity. If one thing is clear about mainstream theories of the Liar, it is that they tend to be rather complicated. It also seems that they are becoming more complicated as simpler theories are tried and found wanting. By contrast, the inconsistency theory could hardly be simpler: the semantic rules governing ‘true’ generate a commitment to (D), not to some sophisticated replacement of (D). In fact, there is more to this simplicity than meets the eye. It is not just that the inconsistency theory is a simple theory, though it certainly is that. Perhaps more importantly, it is very sparing in what it attributes to ordinary speakers. Many, though by no means all, mainstream theories of truth attribute to speakers a mastery of a complex set of semantic rules that do the basic work of a truth predicate while avoiding paradox. By contrast, the inconsistency theory attributes to speakers an utterly simple semantic rule – namely that described by (D) – and does not ascribe any mechanism for avoiding paradox. In this respect, the inconsistency theory is far more economical.

None of this, of course, amounts to a proof of the inconsistency theory. And that should come as no surprise, as philosophical theories cannot, in general, be proved or disproved. The point is simply to show that the inconsistency theory is an attractive alternative to other approaches – provided, of course, that it can withstand objections, to which we now turn.

3. Objections

The most pressing question about the inconsistency theory is whether it, at the end of the day, succumbs to the Herzberger objection that we considered earlier. That is: is the inconsistency theory also an inconsistent theory? Does an inconsistency theory of a language L confine the inconsistency to the language L, or does the inconsistency theory spill over into the theory itself? If so, then the inconsistency theory is, at best, no different from dialetheism, and avoiding that result is a basic ground rule of the present inquiry.
However, the inconsistency theory is not inconsistent in any obvious way. It *reports* an inconsistency in a set of commitments, but there is nothing inconsistent about that, any more than it is inconsistent to point out an inconsistency in a theory or belief system. The fundamental difference between the present theory, a dialetheist theory and a mainstream theory is the attitude they take toward (D). Dialetheism simply asserts (D). The mainstream approach denies (D) and looks for an alternative. The present theory simply observes that our language generates a commitment to (D), but does not endorse this commitment by actually asserting (D).

This may seem like an awfully fine distinction, but it may become clearer if we consider inconsistency theories of foreign languages. Again, imagine a non-English language L into which ‘tall’ has been introduced via something like (Tall3). Assuming for the sake of discussion that the pattern of linguistic commitments in that language is as I am claiming it is, we would speak accurately by saying that (Tall3) is a linguistic commitment of the language L. But it would clearly not be appropriate for us to actually assert (Tall3) (or rather its English translation). That would go beyond simply reporting a pattern of commitments. It would be tantamount to taking those commitments on board. Asking whether a given individual is tall, or whether ‘tall’ applies-in-L to that individual, would be like asking whether the moon is still made of cheese. (Tall3), or its English translation, should be neither asserted nor denied by a theory of L. And (D) should be neither asserted nor denied by a theory of the English predicate ‘true’.

Of course, as a purely practical matter, we use English to state the inconsistency theory of truth. By speaking English, we commit ourselves to (D). But this is not the fault of the inconsistency theory; it is the fault of the language in which, by practical necessity, we state the inconsistency theory. It does not make the inconsistency theory inconsistent.

I am assuming, however, that it is possible in principle for a system of obligations to be inconsistent, and some would doubt this. Might inconsistent obligations be excluded by the very logic of obligation? If they are, then the inconsistency theory is in trouble. When a language generates a linguistic commitment to a contradictory sentence, its speakers have inconsistent obligations in the following sense: they have a commitment to S, and also one to ~S, for some sentences S. Indeed, speakers of an inconsistent language have a commitment to *every* declarative sentence of their language, for the simple reason that everything follows from a contradiction – at least, it does if we accept classical logic, and I have offered no reason not to. They also have inconsistent obligations in the following, stronger sense: they have both an obligation and a prohibition concerning the same speech act, namely that of asserting a certain inconsistent sentence S.

When I explain the inconsistency theory, people often assume that the obligations in question have built-in provisos for handling such conflicts. They don’t. The obligations in question are unconditional. Of course, as a *pragmatic* matter, we might find ways of resolving,
or at least avoiding, these conflicts in practice. I don’t believe that this is due to any special
mechanism for using the truth predicate consistently; it is simply an instance of our general
ability to deal with inconsistencies. But in any case, the fact that people can work around a
conflict of obligations in a particular case does not mean that those obligations did not genuinely
conflict to begin with.

Again, however, the inconsistency theory assumes that it is possible for someone to be
obligated to perform an act A and simultaneously prohibited from performing A. Fortunately,
there seem to be plenty of examples of such conflicting obligations. For example:

• A poorly crafted body of laws can generate conflicting legal obligations, by both
mandating and forbidding a given action in a given situation. When this happens, judges
may of course find a way of resolving the conflict. But when they do, they are implicitly
amending the conflicting statutes, not simply interpreting them.
• Contracts and other promises can conflict in this manner, either internally or, more
likely, with other promises.
• A game can have inconsistent rules. Imagine, for example, a variant of chess that
actually requires a player to capture the opposing player’s queen if it is threatened by the
player’s pawn. Such a capture is sometimes illegal; the capturing pawn might be pinned
against the king, for example. In this variant of chess, the capture is both required and
prohibited.

In general, any system of rules generates a pattern of permissions and obligations: it determines
what one may or must do, if one is to continue following the rules. And an inconsistent system of
rules – a system that makes conflicting mandates on those who follow the rules – is certainly
possible, albeit undesirable.

Now one may certainly question whether such rule systems generate conflicting all-
things-considered obligations. If one finds oneself subject to conflicting laws, then surely one is
not really obligated to do the impossible. Promises, including contracts, are not absolute but can
be voided by extenuating circumstances, and one promise can be cancelled when it conflicts with
a higher-priority promise. And in the case of an inconsistent game, one can simply opt out of the
game, or improvise. So in all of these cases, the obligations in question are not absolute. This is
true, but it completely misses the point. The obligations I am talking about are not all-things-
considered obligations. They are simply the obligations one takes on by working within a system
of rules. They are what the rules tell one to do. All things considered, it may be the best course of
action to opt out of a system of rules, or to modify it. Crucially, it may be advisable to disregard
a set of rules precisely because it generates conflicting obligations. If we do not admit that rules can generate conflicting obligations in the first place, then I find it hard to even make sense of the scenarios described above, which I think everyone must admit are in some sense possible. It is because two laws conflict that courts must resolve the conflict. It is because two contracts conflict that one must find a way to resolve the conflict. Without conflicting obligations, it is hard to even describe the basic facts in such cases.

So again, if there is any internal contradiction in the inconsistency theory, it is not an obvious one. There may nonetheless be good reasons to be squeamish about it. It is, after all, a skeptical account, and skeptical accounts in general have the burden of proving themselves. However, in the case of the Liar, I think it is more or less inevitable that a skeptical account will be called for, whether in the form of the inconsistency theory or otherwise. There is just no getting around the fact that when it comes to the Liar, our ordinary reasoning about truth is deeply flawed. The only question is where to locate that flaw: in our language, or somewhere else.

Overall, I think that whether the inconsistency theory is found to be appealing or unappealing will depend mostly on one’s basic views on language, meaning and representation. To believe the inconsistency theory, one must believe that there are meaningful predicates that lack application conditions, which goes against the grain if one tends to identify meanings with application conditions or truth conditions. One must also believe that languages can, in effect, make posits or assertions, and is not simply a neutral vehicle for stating theories. Obviously an entire paper (at least) could be written about either topic. Nonetheless, let me end by indicating why I think we should at least be open to the view of language that the inconsistency theory implies.

Broadly speaking, there are two approaches to language and meaning: a (late) Wittgensteinian tradition that stresses the social aspects of language, and an opposing tradition that tends to view language as a vehicle for stating facts. Now the objections mentioned in the last paragraph should not be very troublesome from a Wittgensteinian perspective. After all, if a meaningful expression is one that has an established use in a language community, who cares if it lacks a well-defined extension? And as the example of (Tall3) shows, it is quite conceivable for a predicate to have a well-defined use that is based on inconsistent rules. That would presumably qualify it as meaningful from a Wittgensteinian perspective. Thus, I will focus instead on the opposing tradition, which focuses on the representational side of language.

From this perspective, the main job of a declarative sentence is, arguably, that of denoting a state of affairs, and the main job of a predicate is, arguably, that of picking out an extension. This thought has led some to identify the meaning of an expression (i.e., a sentence, predicate, etc.) with its extension; or better, with its intension, or its extension in a possible world.
w as a function of w. If we restrict this to sentences, we get the classic truth-conditional account of meaning. However, the identification of meanings with intensions is, at best, an idealization that leaves certain things out. In particular, it is well known that two expressions can have the same intension without being synonymous. For example, ‘1+1=2’ and ‘Arithmetic is incomplete’ are hardly synonymous, even though they have the same intension, namely the necessarily true proposition. Why is this?

The sentences ‘1+1=2’ and ‘Arithmetic is incomplete’ may pick out the same intension, but they do so in different ways. They represent different conceptualizations of that intension. And, I would argue, the connection between a linguistic expression and an intension is always mediated by a conceptualization in this manner.¹⁵ Now in the case of ‘tall’ (in the (Tall3) example), or in the case of ‘true’, I think we have a predicate that aims at picking out an intension, and which is connected to a conceptualization in the ordinary manner. It just happens that in these cases, the usual mechanism breaks down. The conceptualization associated with the predicate, while it purports to pick out an intension, simply fails to do so.

So in short, I am not denying that predicates aim at picking out extensions, or that sentences aim at stating facts. I am allowing that they usually aim to do so, but do not always succeed. Language is, to a considerable extent, a vehicle for making representations of the world; but it is an imperfect vehicle.

BIBLIOGRAPHY


¹⁵ Proper names represent an arguable exception, since some argue that they are directly referential and lack meanings altogether – see especially (Salmon 1982) and (Soames 2002).


