Burge's Contextual Theory of Truth and the Super-Liar Paradox

Matt Leonard*

Abstract

One recently proposed solution to the Liar paradox is the contextual theory of truth. Tyler Burge (1979) argues that truth is an indexical notion and that the extension of the truth predicate shifts during Liar reasoning. A Liar sentence might be true in one context and false in another. To many, contextualism seems to capture our pre-theoretic intuitions about the semantic paradoxes; this is especially due to its reliance on the so-called Revenge phenomenon. I, however, show that Super-Liar sentences (where a Super-Liar sentence is a sentence which says of itself that it is not true in any context) generate a significant problem for Burge's contextual theory of truth.

Keywords: liar paradox, contextualism, super liar, revenge, truth

1 Introduction

The first sentence of this paper is false. Why is the first sentence such a problem? Well, suppose that the first sentence is indeed false. If we judge the first sentence false, then it seems to be true, because it (truly) says of itself that it is false! Now, suppose that the first sentence is true. If the first sentence is true, then what it claims is

*The author wishes to thank Michael Glanzberg, Adam Sennet, Terence Parsons, Mark Balaguer, David Pitt, Robbie Hirsch, Jacob Caton, Reuben Stern, Goncalo Santos, Ray Jennings, Tyrus Fisher, Michael Hatcher, Michael Anderson, Zeph Scotti and audiences at UC Berkeley, Georgia State University, Arizona State University, and the 2011 Logica Conference for helpful discussions and comments.

true. But it claims that it is false; so it must be false! The problem, then, is that we have a sentence, which seems to be both true and false. But (just about) all of us believe that contradictions cannot be true!

Given the following two rules of inference,

Semantic Ascent: $\alpha \vdash \mathrm{Tr}(\ulcorner\alpha\urcorner)$

Semantic Descent: $\mathrm{Tr}(\ulcorner\alpha\urcorner) \vdash \alpha$[1]

and the following instance of the Liar,

β $\ulcorner\beta$ is not true.$\urcorner$

the Liar's formal proof sometimes runs the following course:

(1)	$\beta = \ulcorner\beta$ is not true.$\urcorner$	[Given]
(2)	Assume $\ulcorner\beta$ is true.$\urcorner$	[For *Reductio*]
(3)	$\ulcorner\beta$ is not true$\urcorner$ is true.	[Substitutivity, (1) and (2)]
(4)	$\ulcorner\beta$ is not true.$\urcorner$	[Semantic Descent from (3)]
(5)	$\ulcorner\beta$ is not true.$\urcorner$	[*Reductio* (2)–(4)]
(6)	Assume $\ulcorner\beta$ is not true.$\urcorner$	[For *Reductio*]
(7)	$\ulcorner\beta$ is not true$\urcorner$ is true.	[Semantic Ascent from (6)]
(8)	$\ulcorner\beta$ is true.$\urcorner$	[Substitutivity, (1) and (7)]
(9)	$\ulcorner\beta$ is true.$\urcorner$	[*Reductio* (6)–(8)]
(10)	$\ulcorner\beta$ is true$\urcorner$ and $\ulcorner\beta$ is not true$\urcorner$.	[(5) and (9)]

One relatively recent response to the Liar is provided by those who endorse a contextual approach to the semantic paradoxes. In general, contextualism is the view that there is an indexical element involved in the reasoning process of the Liar paradox; given a token of the Liar sentence, the extension of 'true' is contingent upon the context of utterance, and in some theories, the intentions of the speaker. Truth is an indexical notion. If a Liar sentence is not true in some context $\Gamma 1$, then the same Liar sentence will be true in a context $\Gamma 2$, where

[1] Semantic Ascent should be read as "From α, you may validly infer that α is true." Semantic Descent should be read as "If α is true, then you may validly infer α."

$\Gamma 2 > \Gamma 1$. There are many different contextualist theories of truth; I will, however, be looking at only one. In section 3, I will explicate Burge's (1979) theory. He claims that the extension of the truth predicate varies with shifts in context. What I hope to do in this paper is present a modest and novel worry for Burge's theory of truth. In section 4, I will mention some worries I have with his theory (and in particular, how his theory deals with the so-called Super-Liar paradox). I'll end by trying to show that the theory as it stands falls into somewhat of a dilemma. But first, let me provide some context by looking at Tarski, Kripke, and the so-called Revenge phenomenon.

2 Tarski's hierarchical theory and Kripke's paracomplete theory

Tarski maintained that the threat of paradox emerges when the truth predicate for a language L_1 resides in L_1 itself. This is why natural language generates paradox. Thus, he proposed that the truth predicate for a language L_1 must be placed in a metalanguage L_2. If we start with an interpreted language L_1, which excludes a truth predicate, we can then add a truth predicate to form L_2 and make claims regarding the veracity of sentences in L_1. For instance, for a sentence $\ulcorner \phi \urcorner$ in L_1, we can claim in L_2 that $\ulcorner \phi$ is true.$\urcorner$ The hierarchy is infinite. For any sentence $\ulcorner \phi \urcorner$ and any level n, we can only claim that $\ulcorner \phi$ is true$\urcorner$ in L_{n+1}. How does this solve the Liar paradox? It solves the Liar paradox because it blocks the formulation of a Liar sentence. Since there are no Liar sentences, there is no paradox.

Though Tarski (1933) did successfully block the Liar in giving his definition of truth, most people (I think rightly) want to say that while Tarski's definitions of truth and denotation are fruitful for metalogic, they are too restrictive for our ordinary notions of truth and meaning. One's theory of the semantic paradoxes should match our pre-theoretic intuitions about natural language, rather than block paradoxical sentences in an artificial language. Hence, Tarski's solution to the Liar is too restrictive.

A more recent (and popular) theory is the paracomplete solution to the semantic paradoxes. Paracomplete solutions maintain that Liar sentences do not have a truth-value (they lack truth conditions). Perhaps the most influential response endorsing truth-value gaps is

Kripke's (1975) theory of truth. Kripke begins with a classical language that lacks a truth predicate.

We should think of an interpreted language L as an ordered triple $\langle \mathcal{L}, \mathcal{M}, \sigma \rangle$, where $\mathcal{L}$ is the syntax, $\mathcal{M}$ is a model that provides an interpretation to the nonlogical vocabulary, and σ is a valuation scheme. Classical languages are characterized as having the following set $\mathcal{V}$ as their 'semantic values': $\{1, 0\}$.[2] Let L_0 be a classical language. In L_0, $\mathcal{M} = \langle \mathcal{D}, \mathcal{I} \rangle$, where $\mathcal{D}$ is a non-empty domain and $\mathcal{I}$ is an 'interpretation-function' which assigns to each name of L_0 an object from $\mathcal{D}$ and assigns to each n-ary predicate an element of $\mathcal{D}^n \rightarrow \mathcal{V}$, in other words, a function taking n-tuples of $\mathcal{D}$ and yielding a truth value, i.e., a semantic value 1 or 0.[3] The extension of an n-ary predicate F contains all n-tuples $\langle a_1, \ldots, a_n \rangle$ of $\mathcal{D}$ such that $\mathcal{I}(F)(\langle a_1, \ldots, a_n \rangle) = 1$, or colloquially, the set of things of which F is true. The valuation scheme for classical languages is τ (dubbed τ for Tarski), where a disjunction is true iff one of its disjuncts is true, a conjunction is true iff both of the conjuncts are true, and so on.

Kripke constructs a non-classical language using Strong Kleene logic in the following way.[4] $\mathcal{V}$ is expanded to $\{1, \frac{1}{2}, 0\}$ and so our new language L_1, $\langle \mathcal{L}, \mathcal{M}, \kappa \rangle$, where κ is the new valuation scheme, is now a three-valued non-classical language. For L_1, model $\mathcal{M} = \{\mathcal{D}, \mathcal{I}\}$, where $\mathcal{I}$ does the very same thing in L_1 as it did in L_0, except now it assigns to n-ary predicates elements of $\mathcal{D}^n \rightarrow \{1, \frac{1}{2}, 0\}$. We want to conceive of predicates in terms of extensions and antiextensions. As in the classical language, the extension of an n-ary predicate F contains all n-tuples $\langle a_1, \ldots, a_n \rangle$ of $\mathcal{D}$ such that $\mathcal{I}(F)(\langle a_1, \ldots, a_n \rangle) = 1$, or colloquially, the set of things of which F is true. The antiextension of an n-ary predicate F contains all n-tuples $\{a_1, \ldots, a_n\}$ of $\mathcal{D}$ such that $\mathcal{I}(F)(\langle a_1, \ldots, a_n \rangle) = 0$, or colloquially, the set of things of which F is false. L_1 leaves open the possibility of some n-tuples not falling in either the extension or antiextension of F; in this case, we say that F is *undefined* for some n-tuple. Let F^+ and F^- be the extension and antiextension of F. L_0 and L_1 both agree that nothing exists in both the extension and antiextension; or, $F^+ \cap F^- = \emptyset$. They differ in the

---

[2]Let '1' represent 'is determinately true', '0' represent 'is determinately false', and when I mention it shortly, let '$\frac{1}{2}$' represent 'is undefined'.

[3]And likewise assigns to each n-ary function-symbol an element of $\mathcal{D}^n \rightarrow \mathcal{D}$, i.e., an n-ary function from $\mathcal{D}^n$ to $\mathcal{D}$.

[4]Here I'll rely on a nice summary of Kripke in (Beall, 2007).

following way. As opposed to L_0, L_1 holds that there can be an x such that x is undefined for F; L_1 denies that, necessarily, $F^+ \cup F^- = \mathcal{D}$. In other words, L_1 denies that every sentence is in the extension or antiextension of the truth predicate.

Kripke then constructs his so-called fixed-point language. Kripke begins with (the classical) L_0, which lacks a truth predicate and extends it to (the non-classical) L_1, which contains a truth predicate. Unlike Tarski's theory, the truth predicate can be applied to every sentence of L_1 (including all of the sentences of L_0). In the above paragraph, I mentioned that L_0 is to be interpreted with a classical model $\mathcal{M}_0$. Kripke proposes to build up a model $\mathcal{M}_1$ for the expanded L_1. Kripke employs an inductive method here. Start with an empty extension and an empty antiextension. Start throwing in true sentences to the extension and false sentences into the antiextension. Eventually, Kripke shows, we will arrive at a level (this is going to be a transfinite level) where adding any more sentences to the extension and antiextension will cease to be 'productive,' i.e., it reaches a *least fixed point*. Liar sentences do not appear in either set, and thus are viewed as 'gappy.' What Kripke seems to have shown is that (i) a language can contain its own truth predicate and (ii) Liar sentences come out lacking a truth-value.

This common sort of response to the Liar, however, has been met with a serious problem. It is often referred to as the 'Revenge of the Liar,' or 'Strengthened Liar reasoning.' The revenge problem is not really a new problem; it is simply another instance of the Liar masked for truth-value gap responses to the original Liar. Technically, the original Liar is of the following form,

β_{OL} $\quad$ β_{OL} is false.

When met with the original Liar, one can just claim that β_{OL} cannot be true and it cannot be false; 'No problem, β_{OL} lacks a truth value'. However, consider again an instance of the Strengthened Liar,

β_{SL} $\quad$ β_{SL} is not true.

The Strengthened Liar is supposed to show that β_{SL} cannot be true, cannot not be true, and cannot not have a truth-value. So what is the revenge problem? If β (from now on, just take β to have the strengthened form) is neither true nor not true, then in particular it is not true. But if it is not true, then it seems that β is not true (since that is what β seems to tell us). Therefore, β seems to be true in an important sense; β is true "after all"! The Strengthened Liar presents the problem in a more intuitive way than the original Liar. So there seems to be a reformulated paradox for 'gappy' theories.

3 Burge and contextualism

Burge wants to distance himself from both the Tarskian and Kripkean solutions to the Liar. He rejects the former for the same reasons that many people do, as I've mentioned above (i.e., it is too restrictive for our ordinary notion of truth, and so on). He rejects the latter, i.e., truth-value gap theories, because of the revenge problem. As a result, he posits a hierarchical theory that, though similar in some respects to Tarski's, differs by attempting to meet some of the pre-theoretic semantic intuitions Tarski's theory did not account for. In particular, he does this by claiming that the truth predicate is indexical and that its extension shifts from context to context. Now, in "The Concept of Truth in Formalized Languages," Tarski sought to block the Liar by assuming that he was dealing with some purely extensional concept of truth, not our ordinary notion of truth. In fact, he argued that natural language was inconsistent and inevitably generated the Liar. Burge, on the other hand, is interested in our ordinary/natural notion of truth. He wants to give a theory concerning our ordinary notion of truth which can block the Liar paradox *as it occurs in natural language*. He frowns on theories which simply block liar sentences in artificial languages with fancy technical ingenuities.

Burge argues that there is a hidden conversational implicature and a shift in extension (parallel with a shift of context) that occurs in Strengthened Liar reasoning. According to Burge, Strengthened Liar reasoning runs the following course:

Step 1: An occurrence of a Liar like sentence.

Step 2: The Liar sentence is not true.

Step 3: The Liar sentence is true after all.

Most solutions to the Liar have either ignored such reasoning or attempted to block it by formal means. Burge, on the other hand, thinks a more satisfying approach is to interpret the reasoning so as to justify it. He thus takes the Strengthened Liar as a model for how we *should* think when confronted with the semantic paradoxes.

Consider the following very plausible scenario (and notice the corresponding Steps 1–3):

> Suppose I see a fake university professor enter a room and begin writing falsehoods on the blackboard. Suppose also that I think that I am in Room 398 and that the fake professor, at this moment, is in 399. So I write on the board at 11:30 A.M. on 6/24/11, (Step 1) "There is no sentence written on the board in Room 399 at 11:30 A.M. on 6/24/11 which is true as stan- dardly construed." However, unbeknownst to me, I am in fact the one in Room 399, and this is the only sentence written on the board. The usual Kripkean (or gappy) reasoning shows that this cannot have truth conditions; thus, it is not true. (Step 2) So there is no sentence written on the board in Room 399 at 11:30 A.M. on 6/24/1 which is true as standardly construed. But we have just stated the sentence in question. (Step 3) Thus, it is true after all.

The truth predicate used in this scenario is not some technical notion of truth (like, say, Tarski-truth). It is our ordinary notion of truth. This is the sort of paradox to which Burge is interested in providing a solution.

Burge wants to stipulate a formal system that defines a *pathological* sentence, as interpreted in a context. Burge stipulates that *pathologicality* is a disposition to produce disease for certain semantical evaluations. Thus, the Liar comes out pathological.[5] *Rootedness* is defined as the lack of pathologicality, i.e., a formula's being rooted means that it is nonpathological, and (roughly) that it has a truth-value.[6]

Burge then distinguishes extensions of 'true' by marking occurrences of them with subscripts beginning with i. In the Strengthened Liar case, for Step 2 Burge claims that the Liar sentence is not true. He marks this initial context of utterance 'true$_i$'. In Step 3, and from a broader application of truth, he claims that the Liar is true.

[5] The Truth-Teller (a sentence which says of itself that it is true) will also come out pathological.

[6] Rootedness is essentially the same notion as groundedness in Kripke's theory.

Burge dubs this context of utterance 'true$_k$,' where $k > i$. He argues that though pathological$_i$ sentences are not true$_i$, pathological$_i$ sentences are nonpathological$_k$, and thus true$_k$. All rootless$_i$ sentences are not true$_i$. So a sentence and its negation may both be not true$_i$, though one or the other will be true$_k$. Burge doesn't offer the following restricted Tarskian truth schema anywhere, but I presume this is a T-schema he would accept (which I'll subscript 'B' for Burge):

(T_B): $(\forall i)$ If a sentence $\ulcorner\phi\urcorner$ is rooted$_i$, then $\ulcorner\phi\urcorner$ is true$_i$ iff p.

where $\ulcorner\phi\urcorner$ names any well-formed sentence in Burge's system and $\ulcorner p\urcorner$ is the sentence itself. Notice how Burge's theory interprets the Strengthened Liar:

Step 1:	β $\ulcorner\beta$ is not true.$\urcorner$	[i.e., a Liar token.]
Step 2:	The Liar is not true.	[i.e., $\ulcorner\beta\urcorner$ is not true$_i$.]
Step 3:	The Liar is true after all.	[i.e., Step 2 is true$_k$.]

4 A dilemma for Burge's contextual theory of truth

The initial appeal of contextual approaches to the semantic paradoxes is that they accord with some of our intuitions about truth, and in particular, how to interpret the Strengthened Liar dialectic. As enticing as this appeal might be, there is a worry with Burge's contextual theory that throws doubt on whether this type of response to the Liar is, in fact, the right type of response. I want to mention both a general worry for all contextualist solutions to the Liar, and a specific worry with Burge's theory. The specific worry is just a problem with Burge's response to the general worry; so first, let me mention the general worry.

The general threat to contextualism emerges when the Strengthened Liar is reformulated in a way that explicitly refers to hierarchical contexts; this formulation is sometimes referred to as the Super-Liar. What type of response can the contextualist provide for sentences like 'This sentence is not true at any level, or in any context,' or sentences like $\ulcorner\psi\urcorner$?

ψ $\qquad$ $(\forall i)$ $\ulcorner\psi$ is not true$_i$.$\urcorner$

It seems that contextualism faces the same sort of paradox Tarski and Kripke face. Either $\ulcorner\psi\urcorner$ is not true$_i$ at any level i or $\ulcorner\psi\urcorner$ is true at some level n. Suppose it is not true$_i$ at any level. But that is just what $\ulcorner\psi\urcorner$ says of itself. Hence, $\ulcorner\psi\urcorner$ is true$_k$, where $k > i$ (i.e., $\ulcorner\psi\urcorner$ is true 'after all'). On the other hand, suppose $\ulcorner\psi\urcorner$ is true at some level n. If that is the case, then $\ulcorner\psi\urcorner$ should come out false at n, because it says of itself that it is not true at any level. In both cases, contextualism seems to be unable to account for $\ulcorner\psi\urcorner$.

Let me show even more explicitly the problem with $\ulcorner\psi\urcorner$ (using Burge's notation). Here I'll universally quantify over the extensions that can be applied to the truth predicate:

(1)	$\ulcorner\psi\urcorner : (\forall i)\ulcorner\psi$ is not true$_i\urcorner$	[Given]
(2)	Assume $(\forall i)\mathrm{Tr}_i\ulcorner\psi\urcorner$	[*Reductio*]
(3)	$\mathrm{Tr}_i[(\forall i)\neg\mathrm{Tr}_i\ulcorner\psi\urcorner]$	[Substitution (2)]
(4)	$(\forall i)\neg\mathrm{Tr}_i\ulcorner\psi\urcorner$	[Semantic Descent (3)]
(5)	$(\forall i)\neg\mathrm{Tr}_i\ulcorner\psi\urcorner$	[*Reductio* (2)–(4)]
(6)	Assume $(\forall i)\neg\mathrm{Tr}_i\ulcorner\psi\urcorner$	[*Reductio*]
(7)	$(\forall i)[\mathrm{Tr}_{i+1}(\neg\mathrm{Tr}_i\ulcorner\psi\urcorner)]$	[$\ulcorner\psi\urcorner$, Burge's Theory]
(8)	$(\forall i)\mathrm{Tr}_{i+1}\ulcorner\psi\urcorner$	[Substitution (7)]
(9)	$(\forall i)\mathrm{Tr}_i\ulcorner\psi\urcorner$	[*Reductio* (6)–(8)]
(10)	$(\forall i)[\neg\mathrm{Tr}_i\ulcorner\psi\urcorner \wedge \mathrm{Tr}_i\ulcorner\psi\urcorner]$	[(5), (9)]

Contextualism seemed most plausible when it was allegedly able to circumvent the revenge problem; Super-Liars, at least prima facie, seem to immediately force contextualist truth theories back into paradox. Some philosophers think that this is a knock-down argument against contextualism. I do not intend to settle this difficult question, in this paper. However, I should note that Burge already knows that this version of the Liar can be generated against contextualism; in fact, he provides a response in advance in his original paper. What is puzzling is that no one seems to address his response. In what follows, I'll argue that the most devastating problem emerges when we put pressure on Burge's response.

Foreseeing the potential problem, Burge writes,

> Attempts to produce a 'Super Liar' parasitic on our symbolism tend to betray a misunderstanding of the point of our account. For example, one might suggest a sentence like (a), '(a) is not true at any level'. But this is not an English reading of any sentence in our formalization. Our theory is a theory of 'true', not 'true at a level'.
> (Burge, 1979, p. 192)

Burge wants to allow the schematic variables on the truth predicate to be contextually determined (by some Gricean process). But he doesn't want to allow quantification on them, in something like the way that type-theoretic levels, in type theories, do not allow quantification. They do not mark a quantifiable argument place on the truth predicate. It's not as if there is *really* a parameter there. The 'parameter' is really just being used as a label to indicate that there is some Gricean process going on. So if you think of it this way (where the truth predicate is immune from quantification) then you can't really formulate the Super-Liar because you can't formally quantify over contexts (or, extensions which are generated from contexts).

Recall that Burge is interested in giving a theory of our ordinary notion of truth. Suppose Ralph is walking down the street and shouts the following to a crowd of people,"To the sentence immediately following this one, I stipulate the name $\ulcorner\phi\urcorner$. $\ulcorner\phi\urcorner$ is not Tarski-true in L, in any of the transfinite metalanguages of L." The crowd would completely ignore Ralph. Why? Because his utterance is not ordinary English. Super-Liars, you might think, utilize some technical notion of truth and hence are immune from the relevant considerations.

I don't think that this is the case, however. I'll briefly argue that it is not inconceivable to construe a Super-Liar in ordinary language (using an ordinary truth predicate). Suppose we have a situation similar to the one I described earlier (with the professor who was a fraud). Suppose again that I walk into Room 399 but think that I am in 398 (and I also think that the fake professor is in 399). Suppose I write on the board at 11:30 A.M. on 6/24/11 one the following:

1. "There is no sentence written on the board in Room 399 at 11:30 A.M. on 6/24/11 which will ever be true."

2. "The sentence written on the board in Room 399 at 11:30 A.M. on 6/24/11 is not true in any context."

3. "The sentence written on the board in Room 399 at 11:30 A.M. on 6/24/11 is not true, no matter how you judge it (or, no matter how you look at it)."

I am not claiming that speakers understand the technical notion of contexts. All the speaker must do is utter one of the sentences above; and when she does, it seems obvious that we have some sort of natural language Super-Liar. But this is a problem for Burge. These above sentences seem to obviously include our ordinary truth predicate (in the same way that the example Burge provided regarding the fake professor includes our ordinary truth predicate). Burge's theory turns out inconsistent for such sentences, however. Such natural sentences land in paradox for the same reasons the more formal $\ulcorner\psi\urcorner$ landed in paradox,

$$\psi \qquad (\forall i)\ \ulcorner\psi \text{ is not true}_i.\urcorner$$

Namely, they are either not true$_i$ at any level i or true$_n$ at some level n. Assume the former, and then they will be true at level k, where $k > i$ (i.e., they are true 'after all'). Assume the latter, and they should come out false at n. Either way, we have a paradox.

What I've attempted to show, then, is that just as the Strengthened Liar can be uttered in natural language, so too can (some) instances of the Super-Liar. Burge is interested in providing a theory of our ordinary notion of truth. Since I've just shown that there are legitimate Super-Liar candidates in natural language, Burge's theory should be able to apply to them as well.

Unlike many other people working on truth theory, Burge specifically is interested in giving a theory of our ordinary notion of truth. He wants to account for liar sentences as they occur in natural language. Thus, he is left with somewhat of a dilemma. He can either give a comprehensive theory of our ordinary notion of truth or not. If he does, then he needs to be able to give an account of natural Super-Liars like the ones I mentioned above. If he doesn't he still needs to be able to give an account of more formal Super-Liars (like $\ulcorner\psi\urcorner$ above). Either way, the problem of the Super-Liar seems to remain for Burge's contextual theory of truth.

References

Beall, J. (2007). Truth and paradox: A philosophical sketch. In D. Jacquette (Ed.), *The handbook of the philosophy of science: Philosophy of logic* (Vol. 10, pp. 325–410). Elsevier B.V.

Burge, T. (1979). Semantical paradox. *The Journal of Philosophy*, *76*(4), 169–198.

Kripke, S. (1975). Outline of a theory of truth. *The Journal of Philosophy*, *72*(19), 690–716.

Tarski, A. (1933). The concept of truth in formalized languages. In *Logic, semantics, metamathematics* (pp. 152–278). Oxford University Press.

Matt Leonard
Department of Philosophy
University of California, Davis
e-mail: mjleonard@ucdavis.edu
URL: http://philosophy.ucdavis.edu/people/mjleonar