Can we solve the Liar paradox by simply declaring Liar sentences to be false? That is the proposal of Dale Jacquette in (Jacquette, 2007). The Liar paradox arises because apparently we can derive a contradiction from the assumption that the Liar sentence is true, and we can also derive a contradiction from the assumption that it is false. Jacquette argues that the former reasoning (i.e., the derivation of a contradiction from the assumption that the Liar is true) is sound, while the latter reasoning is unsound. Thus, he argues, we can consistently hold Liar sentences to be false, thereby solving the Liar paradox. Unfortunately, Jacquette’s reasoning is flawed. Specifically, it depends on misidentifying the Liar paradox. Jacquette’s proposed solution does indeed solve a puzzle, of sorts; that puzzle just happens not to be the Liar paradox.

As is well known, the Liar paradox is an apparent contradiction generated by (1) the Tarskian disquotational schema for the truth predicate, and (2) the fact that sentences can refer to themselves, and in particular, can ascribe falsehood to themselves. Jacquette chooses the following formalization for the disquotational schema:

\( (TS) \forall p [\text{TRUE}(\lceil p \rceil) \leftrightarrow p] \)

He also adds the following to “formalize commitment to bivalent logic”:

\( (CL) \forall p [\text{TRUE}(\lceil p \rceil) \leftrightarrow \neg \text{FALSE}(\lceil p \rceil)] \)

Finally, the idea of self-reference, and in particular of self-ascription of falsehood, is captured as follows:

\( (L) L \rightarrow \text{FALSE}(\lceil L \rceil) \)

Jacquette spends the main body of the paper arguing that Liar reasoning (i.e., the informal reasoning that we use to derive a contradiction when considering Liar sentences) breaks down when applied to (TS), (CL) and

---

(L). That is, he argues, such reasoning cannot be formalized so as to derive a contradiction from those three formal principles.

However, this is wholly unsurprising and is in no way a diagnosis or solution of the Liar paradox. The reason (L) fails to generate a contradiction is that it is simply too weak. In fact, (L) would be satisfied by substituting any false sentence for \( L \) whatsoever. In that case, (L)’s antecedent would be false, making (L) itself true. For example, we could substitute ‘Elephants are smaller than gerbils’ for \( L \) in (L) and get the true material conditional

\[
(1) \text{ If elephants are smaller than gerbils, then ‘Elephants are smaller than gerbils’ is false.}
\]

Since (1) has a false antecedent, it is a true sentence. So in effect, all that the condition (L) does is to take a sentence letter, \( L \), and declare it to be false; unsurprisingly this does not lead to any sort of trouble.

Clearly the intent of (L) was something more—(L) was intended to say not that \( L \) is false, but that \( L \) is a Liar sentence. But (L) falls short, saying at most that \( L \) is a sentence that implies its own falsity. To be a Liar sentence, a sentence must actually state its own falsity—or more exactly, for \( S \) to be a Liar sentence, \( S \) must actually be equivalent to the statement that \( S \) is false. Many sentences imply their own falsity. A logically inconsistent sentence implies its own falsity, for example, simply because it implies everything whatsoever. But such sentences are not Liars. (L) is simply too weak to capture the idea that \( L \) is a Liar sentence.

If (L) is too weak, what would be an appropriately strong condition? What condition would formally capture the idea that \( L \) is a Liar sentence? The following is just about universally\(^2\) recognized as an appropriate condition:

\[
(L^*) \quad L \leftrightarrow \text{FALSE}(\downarrow L^*)
\]

(That is, we have simply replaced the conditional in (L) with a biconditional.) And indeed, the conditions (TS), (CL) and (L*) are jointly inconsistent, as the following simple derivation shows:

\[
\begin{align*}
1. \text{TRUE}(\downarrow L^*) & \leftrightarrow L^* \quad \text{(from(TS))} \\
2. \text{TRUE}(\downarrow L^*) & \leftrightarrow \neg \text{FALSE}(\downarrow L^*) \quad \text{(from(CL))}
\end{align*}
\]

\( ^2 \) Jacquette has a different reading of the literature: he writes that „commentators on the liar have generally construed the liar sentence itself as a conditional rather than biconditional statement...” (p. 97). However, as I will argue below, this claim is simply inaccurate. In addition, by speaking of the Liar as a “conditional statement,” Jacquette seems to be confusing the Liar sentence \( L \) with the condition (L). The Liar sentence itself need not take the form of a conditional or biconditional sentence.
3. \( L^* \leftrightarrow \neg \text{FALSE}(\lceil L^* \rceil) \) (from steps 1 and 2)
4. \( \text{FALSE}(\lceil L^* \rceil) \leftrightarrow \neg \text{FALSE}(\lceil L^* \rceil) \) (from \((L^*)\) and 3)

And step 4 is (classically) inconsistent.

Jacquette does consider a version of the Liar paradox based on \((L^*)\), but confines his discussion to an addendum. He admits there that given \((L^*)\), “an outright logical antinomy is unavoidable.” (p. 96) However, he does not think this fact is particularly interesting. He writes:

The liar sentence cannot be defined biconditionally [i.e., in terms of \((L^*)\) instead of \((L)\)] if the purpose is to produce an interesting derivation of the liar paradox. If we introduce the liar as \((L^*)\), then, in light of the standard truth schema, we in effect assume that \(\text{TRUE}(\lceil L^* \rceil) \leftrightarrow \text{FALSE}(\lceil L^* \rceil)\). The biconditional proposition, and from it the derivation of \(L^* \leftrightarrow \neg L^*\), is no more paradoxical or in need of solution, given its blatantly logical antinomical form, than if we were to try to embarrass classical logic by baldly asserting any other explicit contradiction, such as \(A \leftrightarrow \neg A\), of which the biconditional liar is merely a substitution instance. (p. 97)

However, Jacquette’s reasoning is flawed. Granted, it is quite true that if we posit \((L^*)\), then in light of the the standard truth schema, we are effectively assuming a contradiction. That is exactly the point: \((L^*)\) is contradictory in light of the standard truth schema, which demonstrates that there is either something wrong with the standard truth schema, or something wrong with \((L^*)\). The Liar paradox, as standardly understood, is precisely this conflict between the standard truth schema and \((L^*)\). Either one has to go, or the other does.

So why not simply deny \((L^*)\)? Wouldn’t that be better than denying \((TS)\)? It would be, except that we are stuck with \((L^*)\): it is provable. More specifically, \((L^*)\) has substitution instances that are mathematically provable. It also has substitution instances that are empirically verifiable. These facts about \((L^*)\) are quite well known, but let us briefly review them here.

First, we can construct a substitution instance of \((L^*)\) mathematically. For any formula \(F(\alpha)\) of the language of arithmetic, we can find a sentence \(A\) such that the following is a theorem of Peano Arithmetic\(^3\):

\[
(A) \quad A \leftrightarrow F(\lceil A \rceil)
\]

This is Gödel’s famous Diagonal Lemma, and it does not essentially depend on what formal language we work in; it continues to hold if we add extra predicates (say) to the language of arithmetic, and in that case \(F\) may be any formula of the extended language. In particular, if we add the unary predicates \(\text{TRUE}\) and \(\text{FALSE}\) to the language of arithmetic, then as a special case of \((A)\) we have, for some sentence \(L\),

---

\(^3\) I am assuming a fixed Gödel numbering as given, and treating \(\lceil A \rceil\) as shorthand for the standard numeral of \(A\)’s Gödel number.
\[(A') L \leftrightarrow \text{FALSE}([L])\]

I emphasize that the sentence \((A')\) is a theorem of Peano Arithmetic, and is in no way an extra posit. \((A')\) is also, of course, formally identical to \((L^*)\), and so Peano Arithmetic would be rendered inconsistent by the addition of the premises (CL) and (TS).

The other way to find a true substitution instance of \((L^*)\) is to construct a sentence which, as a matter of empirical fact, says of itself that it is false. For example, suppose we introduce a new constant \(c\) into our language and decide to use it as a name for the sentence \(\text{FALSE}(c)\) — we can use \(c\) as a name for anything we like, including the string of symbols \(\text{FALSE}(c)\). Thus, we are entitled to assert

\[(SR) \ c = [\text{FALSE}(c)]\]

\((SR)\) is perfectly consistent, provided we do not also assume the disquotational schema (TS). However, (SR) logically implies

\[
\text{FALSE}(c) \leftrightarrow \text{FALSE}([\text{FALSE}(c)])
\]

which is simply a substitution instance of \((L^*)\). So again, we are stuck with \((L^*)\) and will have to look somewhere else if we want to reject a premise of the Liar argument. In case it is suspected that something illicit is going on with (SR), note that we can get the same result using a definite description instead of a name. For example, consider the following:

\((B)\) The boldfaced sentence in the article “Undeniably Paradoxical” is false.

Since (B) is the only boldfaced sentence of this article, we have the following:

\[(SR2) \text{The boldfaced sentence in the article “Undeniably Paradoxical”} = \text{The boldfaced sentence in the article “Undeniably Paradoxical” is false.}\]

When formalized, (SR2) implies an instance of \((L^*)\) in exactly the same way that (SR) does. But (SR2) holds as a matter of empirical fact, so denying it is not an option.

Incidentally, we also cannot save (TS) by denying (CL), as the latter plays no essential role here. To see this, simply drop (CL) and replace \((L^*)\) with

\[(L') L \leftrightarrow \text{TRUE}([L])\]

\[\text{[4]}\] The displayed sentence follows from (SR) and the tautology \(\text{FALSE}(c) \leftrightarrow \text{FALSE}(c)\) by substitution.
A contradiction may now be derived very easily. As the $L$-instance of (TS), we have

$$L \leftrightarrow \text{TRUE}(L)$$

which, combined with ($L^\prime$), gives us the (classical) contradiction

$$L \leftrightarrow \neg L$$

Of course, we do have to assume classical logic in this derivation; we made the same assumption in deriving a contradiction from (TS), (CL) and ($L^*$). Jacquette writes as though positing (CL) were tantamount to assuming classical logic, but it is not. To assume classical logic is to help oneself to the classical rules of inference. We could assume (CL) but confine ourselves to a weaker set of inference rules; or we could work within classical logic but refrain from assuming (CL). In fact, (CL) is basically just a definition of FALSE in terms of TRUE, which helps explain why it was so easy for us to dispense with it.

In short, then, to get the Liar going, it is not necessary to explicitly assume ($L^*$), ($L^\prime$), or even ($L$). One simply constructs self-referential sentences that ascribe falsehood (or untruth) to themselves, at which point the relevant instances of ($L^*$) or ($L^\prime$) are provable. And this is the standard understanding of the Liar that one finds in the literature, notwithstanding Jacquette’s claims to the contrary. In (Tarski, 1956), for example, Tarski develops the Liar as follows (p. 158):

[W]e shall use the symbol ‘$c$’ as a typographical abbreviation for the expression ‘the sentence printed on this page, line 5 from the top.’ Consider now the following sentence:

$c$ is not a true sentence.

Having regard to the meaning of the symbol ‘$c$’, we can establish empirically:

(a) ‘$c$ is not a true sentence’ is identical with $c$.

For the quotation-mark name of the sentence $c$ (or for any of its other names) we set up an explanation of type (2):

(β) ‘$c$ is not a true sentence’ is a true sentence if and only if $c$ is not a true sentence.

Tarski then goes on to explain how (β) yields a contradiction when combined with an unrestricted version of (TS). (β) is, of course, simply a substitution instance of ($L^\prime$). Similarly, in (Kripke 1975), Kripke shows how to construct self-referential sentences in several different ways, using the techniques outlined above. ($L^\prime$) is never derived explicitly, because there is no need: the existence of such sentences is enough to produce a contradiction in the presence of (TS) and classical logic. Overall, I am
aware of nothing in the literature that could be considered a “conditional Liar” in Jacquette’s sense of the term.

In conclusion, the Liar paradox can be stated very simply: the conditions (TS), (CL) and (L*) (or alternatively, the conditions (TS) and (L*')) are jointly inconsistent, assuming classical logic. This fact is problematic, because it is hard to see how it can be open to us to reject any of these three conditions. This is an entirely standard understanding of the Liar paradox, despite Jacquette’s repeated and rather puzzling assertions to the contrary. The Liar paradox is simply not a puzzle about the weaker condition (L). Denying the sentence L solves a puzzle, of sorts: it explains why (TS), (CL) and (L) are mutually consistent. But it does not solve the Liar paradox.5

References


5 I would like to thank Peter Boltuc and Roxanne Kurtz for their helpful comments on an earlier draft of this paper.