Abstract. In this paper I respond to Jacquette’s criticisms, in (Jacquette, 2008), of my (Barker, 2008). In so doing, I argue that the Liar paradox is in fact a problem about the disquotational schema, and that nothing in Jacquette’s paper undermines this diagnosis.

1.

In (Barker, 2008), I questioned Dale Jacquette’s treatment of the Liar paradox in (Jacquette, 2007), and hinted that contrary to that treatment, the Liar is in fact a certain logical instability in the familiar disquotational schema. Jacquette responded in turn, in (Jacquette, 2008). Suffice it to say that Jacquette was not at all convinced by my arguments; indeed, he characterizes my paper as “riddled from beginning to end with confusions and misunderstandings”(Jacquette, 2008, p. 143). I am equally unconvinced by Jacquette’s latest volley, and I consider the underlying issues sufficiently important to merit one more attempt to make my argument clear; hence the present response.

Jacquette has offered a veritable laundry-list of complaints about my original paper. It would be easy enough to respond point-by-point, but such a response would severely try my readers’ patience. Thus I have instead returned to my original argument, presented in somewhat more detail than before, and have also incorporated responses to the more pertinent of Jacquette’s criticisms. In so doing, I hope that I have also clarified my view on the relation between the Liar paradox and the disquotational truth schema, an important issue in its own right and, in my view, the most fundamental issue of the present dialog.

1 I thank the editors of this Journal for helpful comments on an earlier draft.
2.

Jacquette and I seem to agree that a Liar sentence is simply a sentence that self-ascribes falsehood.\(^2\) Perhaps the simplest example of such a sentence is

\[(1) \text{ Sentence (1) is not true.}\]

Sentences like (1) generate a certain puzzle, or family of puzzles, known as the Liar paradox. Informally, we can argue from the hypothesis that (1) is true to the conclusion that (1) is false, and vice versa, from the hypothesis that (1) is false to the conclusion that (1) is true. Needless to say, this is puzzling if not outright contradictory. Using quite ordinary reasoning, which I will refer to as Liar reasoning, we may apparently derive a contradiction out of thin air. Now one way to address this rather unsatisfactory situation is to identify a fallacy in Liar reasoning, and this is exactly what Jacquette believes he has done. This is also exactly where I take issue with Jacquette.

As a first step toward formalizing Liar reasoning, we must find a way of saying, formally, that a given sentence is in fact a Liar sentence. At issue here are two candidate formalizations of this claim, both of which Jacquette considers in his original article. First, there is a conditional:

\[(2) \ L \rightarrow \neg \text{TRUE}(\lceil L \rceil)\]

(2) says, or attempts to say, that \(L\) is a Liar sentence. That is, (2) represents an attempt to express formally the following idea: that \(L\) says that \(L\) is not true. It should be noted that (2) is not itself a Liar sentence; rather, it is \(L\) that is the Liar sentence, and (2) simply says (or attempts to say) that \(L\) is a Liar sentence. The other formal attempt to capture the idea of a Liar sentence is quite similar to (2), but it is a biconditional, not a conditional:

\[(3) \ L \leftrightarrow \neg \text{TRUE}(\lceil L \rceil)\]

Again, (3) is not actually a Liar sentence; instead, it attempts to say that \(L\) is a Liar sentence.

Jacquette makes the following set of claims, both in his original paper and in his response to me. First, he argues, (2) adequately captures the idea that \(L\) is a Liar sentence. Second, if the informal Liar reasoning is

\(^2\) For simplicity, I am ignoring other forms of circularity that also generate Liar-type paradoxes. For example, two or more sentences may form a paradoxical chain, e.g.: (a) sentence (b) is false; (b) sentence (a) is true. A more complicated example is Yablo’s paradox: there we have a sequence \(S_1, S_2, \ldots\), where each sentence \(S_i\) in the sequence says that the remaining sentences \(S_{i+1}, S_{i+2}, \ldots\), are not true; see (Yablo 1993). Such cases also generate paradoxes, but via more complicated reasoning than with the simple Liar sentence.
formalized with (2) as its basis, then that reasoning is unsound: a
contradictory conclusion simply fails to follow from (2). If this is so, then
the sought-after fallacy in Liar reasoning has been discovered. And finally,
(3) is simply a non-starter, since it is logically inconsistent in a sense to be
explained shortly. Thus, we may simply dismiss (3) and focus on (2).

By contrast, I have argued that it is (3), and not (2), that better
captures the idea of what it is to be a Liar sentence. More importantly,
I argue, the Liar paradox is simply not solved by the above
considerations, or for that matter by any other considerations that
Jacquette adduces. Since Jacquette takes me to task for characterizing
his work as an attempt to “solve” the Liar (Jacquette, 2008, p. 143), I
should perhaps add that the Liar is also not dissolved, blocked,
diagnosed, debunked, explained, or otherwise taken care of by
Jacquette’s treatment. To make it clear that this is the case, however, I
will need to examine the Liar paradox in somewhat more depth than I
have done thus far.

Before proceeding, however, there is an important clarification that
needs to be made. The sentences (2) and (3) above are not Liar sentences.
The sentence (2) represents an attempt to say that the sentence \( L \) is a Liar
sentence; but (2) itself is not a Liar sentence. Likewise, sentence (3)
represents an attempt to say that \( L \) is a Liar sentence; but (3) itself is not a
Liar sentence. This ought to be fairly clear and straightforward; indeed, in
my original paper I considered it too obvious to be worth mentioning. Yet
Jacquette is not entirely consistent on this point. He spends a good deal of
time in both papers, for example, speaking of “conditional liars” and
“biconditional liars.” Both terms are misnomers, however, since it is the
sentences (2) and (3) that are conditionals and biconditionals, respectively,
whereas it is the sentence \( L \) itself that is a Liar sentence. For all we know,
the sentence \( L \) might itself be a conditional, biconditional, conjunction,
atomic sentence, or anything else.

Moreover, even if it can somehow be shown that the Liar sentence is
best viewed as a conditional, that would be irrelevant to the current
discussion. The dispute between (2) and (3) is not a dispute over whether
Liar sentences are conditionals, biconditionals, or anything else. It is a
dispute about whether the statement that a given sentence is a Liar is best
seen as a conditional or a biconditional. One might, for example, agree
that the sentence \( L \) is a conditional, and still insist that it is (3), and not (2),
that says that \( L \) is a Liar. Alternatively, one might agree with Jacquette
that (2) says that \( L \) is a Liar sentence, while still doubting that \( L \) itself is a
conditional.

Now I would dismiss the terms “conditional liar” and “biconditional
liar” as merely infelicitous choices of words if Jacquette did not make
remarks such as the following: “I maintain that the liar sentence is a
conditional that says of itself that it is false: (L) $L \rightarrow \text{FALSE}[L]$” (Jacquette, 2008, p. 144). Here Jacquette is plainly referring to the conditional $L \rightarrow \text{FALSE}[L]$, and not the sentence $L$ itself, as a Liar sentence. In other words, he appears to be claiming that (2), and not the sentence $L$ that (2) contains, is a Liar sentence. He goes on in the same paragraph to explain “why the liar sentence cannot be a biconditional,” (Jacquette, 2008, p. 144), evidently referring to the biconditional (3); again, however, no one is claiming that (3) is a Liar sentence, the pertinent claim being that (3) says that $L$ is a Liar sentence. Later, in a discussion of Tarski, he writes: “The liar paradox in Tarski’s ($\beta$) is biconditional, just as we would expect it to be, but the liar sentence, in Tarski’s ($\alpha$), and as I would also insist, is conditional rather than biconditional”³ (Jacquette, 2008, p. 148).

I do not personally think that Jacquette has shown, either in connection with Tarski or otherwise, that Liar sentences are or need to be conditionals. However, surely it does not matter whether he has shown this or not. Again, even if one were to show that the Liar sentence $L$, is a conditional, that would in no way establish that (2), rather than (3), is the proper way to express the idea that the sentence $L$ is a Liar. (I am actually quite happy to allow that some Liar sentences are conditionals. Liar sentences can take a variety of forms; what makes them Liars is not their form but the fact that they all self-ascibe falsehood.)

Jacquette himself appears to be of two minds about this issue, for he also writes: “[T]he sentence $L$ is self-referentially defined by (L) as a sentence that says of itself that it is false” (Jacquette, 2008, p. 145). (Recall that Jacquette’s (L) is essentially my (2).) Here we find a clear distinction between the Liar sentence $L$ on the one hand and the conditional (2) on the other. Later he writes, “I […] take a liar sentence to state self-referentially of itself that it is false. The real question here is whether or not sentence $L$ as defined by (L) accomplishes this purpose and lives up to this intent” (Jacquette, 2008, p. 146). I could not agree more.

I have discussed this matter at such length because everything that follows depends on it. The arguments to follow will simply be incomprehensible if one fails to distinguish Liar sentences on the one hand from the sentences (2) and (3) on the other.

³ It should be noted that neither Jacquette nor I have been sufficiently careful in reading ($\beta$), for that sentence is simply an instance of the disquotational schema.
Jacquette states that the biconditional (3) entails a logical inconsistency, and in so doing he claims to agree with me (see Jacquette, 2008, p. 151, specifically section 8). However, it is not strictly true that (3) entails a logical inconsistency, if by that one means that an overt contradiction may be derived from (3) using only the first-order predicate calculus. To get a formal contradiction out of (3), one needs a supplementary, non-logical assumption. And the natural such assumption is the familiar disquotational schema:

\[(4) \text{ TRUE}(\overline{A}) \leftrightarrow A\]

(4) and (3) together formally imply a contradiction, while (4) and (2) together do not, a fact that Jacquette and I fortunately seem to agree on. Indeed, it is this incompatibility between (3) and (4) that leads Jacquette to reject the biconditional (3) as a non-starter. Let us briefly recall why there is a conflict between (3) and (4). Substituting \(L\) for \(A\) in (4), we get

\[(5) \text{ TRUE}(\overline{L}) \leftrightarrow L\]

And (5), taken together with the biconditional (3), or \(L \leftrightarrow \neg \text{TRUE}(\overline{L})\), immediately yields the following formal contradiction:

\[(6) \quad L \leftrightarrow \neg L\]

By contrast, this argument breaks down if (2) is substituted for (3). In that case, instead of (6) we simply get

\[(7) \quad L \to \neg L\]

which is logically equivalent to \(\neg L\). This should all be quite uncontroversial, and indeed Jacquette and I have both made these exact points in our respective papers.

And with that, we finally have enough machinery in place to squarely address the proposal of (Jacquette, 2007) and (Jacquette, 2008). That proposal essentially regards the failure of (2) and (4) to jointly imply a contradiction as a solution (or dissolution, or diagnosis, or ...) of the Liar paradox. Specifically, the Liar paradox involves the informal derivation of a contradiction; and since (2), not (3), is the proper way of saying that \(L\) is a Liar sentence, any formalization of Liar reasoning would have to involve the derivation of a formal contradiction from (2) and (4). Since no such derivation is possible, Liar reasoning is thereby blocked. At any rate, this seems to me to capture the gist of Jacquette’s proposal.

My objection, both in (Barker, 2008) and here, is two-fold. First, I would argue that it is (3), not (2), that formally expresses the idea that \(L\) is
I will argue this point at length in a later section. But second, regardless of how we resolve this controversy about (2) and (3), the disquotational schema (4) itself is under serious threat. There are numerous hard-to-deny claims from which a formal contradiction may nevertheless be derived with help from (4). In other words, there are various hard-to-deny claims that are formally inconsistent with (4). That by itself is rather puzzling, since (4) is also hard to deny. Moreover, these hard-to-deny claims do involve sentences that self-ascribe falsehood, i.e., Liar sentences. Finally, the derivation of a formal contradiction from at least some of these claims (in conjunction with the disquotational schema) involves, as an intermediate step, a sentence formally identical to the biconditional (3), which speaks against the idea of simply rejecting (3). This situation is, again, a puzzling one, and we have a name for the puzzle: it is called the Liar paradox. I will now develop these points in more detail.

4.

A Liar sentence is a sentence that says of itself that it is not true. Now the conditional (2) and the biconditional (3) are two ways to attempt to capture this idea formally, but there is also a third way, which Jacquette does not consider. Specifically, consider the sentence

\[ \neg \text{TRUE}(a) \]

where \( a \) is a name that denotes the sentence (8) itself. This fact, that \( a \) denotes (8), may be expressed formally as follows:

\[ a = \left[ \neg \text{TRUE}(a) \right] \]

Thus, the condition (9) is yet another way, alongside (2) and (3), by which we might try to express the idea that a given sentence is a Liar. Specifically, we may regard (9) as a way of saying that (8) is a Liar sentence. Now from (9) and the disquotational schema (4), we may derive a contradiction using only standard classical first-order logic. We start with the following instance of (4):

\[ \text{TRUE}(\left[ \neg \text{TRUE}(a) \right]) \leftrightarrow \neg \text{TRUE}(a) \]

(10) is simply the result of substituting the sentence \( \neg \text{TRUE}(a) \) for the schematic letter \( A \) in (4). Now from (10) and (9) we may derive the following, using the substitutivity property of identity:

\[ \text{TRUE}(a) \leftrightarrow \neg \text{TRUE}(a) \]
Disquotation, Conditionals, and the Liar

(We have simply replaced the quotation name \[ \neg \text{TRUE}(a) \] in (10) by the constant \( a \), a move that is justified by (9).) And (11) is directly self-contradictory. Thus, an overt contradiction is derivable from (9) and the disquotational schema (4), using only the first-order predicate calculus with identity.

Now the derivation of a contradiction from (9) and (4) may certainly be regarded as (a formalization of) Liar reasoning. After all, (9) is about as direct a statement as one could desire for the idea that the sentence \( \neg \text{TRUE}(a) \) ascribes non-truth to itself, and the derivation of a contradiction from (9) and (4) directly parallels the derivation of a contradiction from (3) and (4). In any case, whether we choose to call sentence (8) a Liar sentence or not, the fact that we can derive a contradiction from (9) and (4) shows that on pain of contradiction, we must reject either (9) or (4).

Just as Jacquette recommends that we reject the biconditional (3), we may be tempted similarly to deny (9) on the grounds that it runs afoul of the disquotational schema (4). However, it is far from clear that we can simply deny (9). The constant \( a \) is a name, and we may use a name to refer to any object we please. Why then can we not use the name \( a \) to denote the string of symbols \( \neg \text{TRUE}(a) \)? Of course, once we do this, we are immediately committed to (9), which is logically inconsistent with (4). Yet there is nothing clearly illicit in declaring that the name \( a \) shall denote the expression \( \neg \text{TRUE}(a) \).\(^4\) Thus, the fact that (9) conflicts with the disquotational schema (4) constitutes something of a puzzle in its own right. An adequate treatment of the Liar, in its present manifestation, would have to explain why, contrary to all appearances, we may not use the name \( a \) to denote the sentence \( \neg \text{TRUE}(a) \): either that, or explain why we may deny at least some instances of (4), which is logically the only alternative to denying (9).

Of course, names per se do raise a number of philosophical questions in their own right, so perhaps there is some legitimate uncertainty about whether we are really entitled to assert (9). That turns out not to matter, however, since there is yet another version of the Liar that we need to consider. Logically, much of the work of names can be performed by definite descriptions. A definite description is simply an expression “the \( F \),” where \( F \) is a predicate that is satisfied by exactly one object. What if we construct a Liar sentence that refers to itself not by name, but by description?

Specifically, let \( D(x) \) be a formula whose only free variable is \( x \), and consider the sentence

\(^4\) This argument was essentially made in (Kripke, 1975); see in particular his "Jack" example.
(12) \( \exists x \ (D(x) \land \neg \text{TRUE}(x)) \)

Now suppose further that \( D(x) \) happens to be satisfied by exactly one object. Then the sentence (12) says, more or less, that this unique object satisfying \( D(x) \) is not true. (I say “more or less” because (12) does not explicitly say that \( D(x) \) is satisfied by a unique object; this will turn out not to matter for our purposes, however.) Suppose further that this unique object happens to be the sentence (12) itself. Then (12) is yet another sentence that refers to itself and says, of itself, that it is not true. The fact that \( D(x) \) is satisfied uniquely by (12) may be expressed formally as follows:

(13) \( \forall x \ (D(x) \leftrightarrow x = \left[ \exists x \ (D(x) \land \neg \text{TRUE}(x)) \right]) \)

(13) says, in effect, that “the \( x \) such that \( D(x) \)” is a definite description denoting the sentence (12). Since (13) says that \( D(x) \) is uniquely satisfied by the sentence (12), and since (12) in turn says that the object that satisfies \( D(x) \) is not true, we may regard (12) as a Liar sentence; and it is fair to ask whether (13), which says that (12) is a Liar sentence, comes into conflict with the disquotational schema (4) in the same way that (9) and (3) both do. And indeed, (13) does conflict with the disquotational schema in just this way. Let us abbreviate sentence (12) as \( L \); then (13) may be written more compactly as

(13’) \( \forall x \ (D(x) \leftrightarrow x = \left[ L \right]) \)

(13’) in turn logically implies

(14) \( L \leftrightarrow \neg \text{TRUE}(\left[ L \right]) \)

To see that (13’) implies (14), let us first rewrite (14) by expanding the left hand side of the biconditional:

(14’) \( \exists x \ (D(x) \land \neg \text{TRUE}(x)) \leftrightarrow \neg \text{TRUE}(\left[ L \right]) \)

To show that (13) implies (14), we need only show that (13’) implies (14’). Now (13’) simply says that the sentence \( L \) is the unique object satisfying \( D(x) \). Thus, both the left-to-right and the right-to-left direction of (14’) follow directly from (13’). (Left-to-right: if \( \exists x \ (D(x) \land \neg \text{TRUE}(x)) \), then since \( L \) is the unique object satisfying \( D(x) \), it follows that \( \neg \text{TRUE}(\left[ L \right]) \). Right-to-left: if \( \neg \text{TRUE}(\left[ L \right]) \), then since \( L \) satisfies \( D(x) \), it certainly follows that \( \exists x \ (D(x) \land \neg \text{TRUE}(x)) \).) Thus, (13) logically implies (14).

Now (14), in turn, is simply a substitution instance of the biconditional (3), and it runs afoul of the disquotational schema (4) in exactly the same
way that (3) does. Namely: (14), together with the $L$-instance $L \leftrightarrow \text{TRUE}(\L)\text{ of (4), jointly imply } L \leftrightarrow \neg L\text{, a contradiction. Thus, we can derive a contradiction from (13) together with the disquotational schema using Liar-type reasoning. (14) conflicts with the disquotational schema in exactly the same way that (3) does, yet we cannot simply deny (14). We cannot simply deny (14), because (14) is not a premise of the argument: it is rather a consequence of the premise (13). Thus, the only way to reject (14) is to reject (13) as well.

Now if we decide to reject (13), we face essentially the same problem we faced when we contemplated rejecting (9). Namely, if the formula $D(x)$ happens, as a matter of empirical fact or otherwise, to be uniquely satisfied by the string of symbols $\exists x (D(x) \& \neg \text{TRUE}(x))$, then we are automatically committed to (13). To reject (13), we must somehow show that a formula $D(x)$ can never uniquely describe the sentence $\exists x (D(x) \& \neg \text{TRUE}(x))$. Yet $D(x)$ could easily turn out to do so. For example, let us stipulate that $D_1(x)$ shall abbreviate (some formalization of) the following: “$x$ appears in Barker’s second reply to Jacquette immediately to the right of the first occurrence in that article of the string ‘(15).’” Now consider the sentence

$$\exists x (D_1(x) \& \neg \text{TRUE}(x))$$

It should be obvious that $D_1(x)$ is satisfied by the sentence (15), and only by sentence (15). Have I not thereby created a sentence, namely (15), that refers to itself, and says of itself that it is not true, via the formula $D_1(x)$?

(15) is self-referential because the formula $D_1(x)$ turns out, as a matter of empirical fact, to denote the sentence (15) itself. It is also possible to create self-referential sentences in a different way, using only pure mathematics, by the famous trick due to Gödel in (Gödel 1931). This trick is somewhat intricate, but it is worth describing it in detail to establish that it really does apply in the current situation. We begin with the syntactic notion of substitution. Consider a formula $F(x)$ with one free variable, and consider also a closed term $t$. Then the substitution of $t$ into $F(x)$, denoted $F(t)$, is simply the result of replacing each free occurrence of $x$ in $F(x)$ by $t$. In particular, $t$ might be the quotation name $\L{G}$ of some formula $G$, in which case the resulting substitution is $F(\L{G})$. Now the notion of substitution, being a syntactic notion, may well turn out to be expressible in our formal language itself. Thus, let us suppose that there is a formula $\text{Sub}(x, y, z)$ which says that $x, y$ and $z$ are formulas, and that $z$ is obtained by substituting the quotation name of $y$ into the formula $x$. Then the following will certainly hold, for all formulas $F(x)$ and $G$:

$$\forall z (\text{Sub}([F(x)]_\L, [G]_\L, z) \leftrightarrow z = [F([G]_\L)])$$
(16) appears to be rather innocuous. Indeed, (16) appears to be definitionally true, given the meaning of ‘Sub.’ It may be somewhat surprising, then, that (16) is logically incompatible with the disquotational schema (4).

Proving this fact about (16) takes a bit of work, however. As a first step, we define the formula $H(x)$ as follows:

$$(17) \quad H(x) \equiv \exists z \ (\text{Sub}(x, x, z) \& \neg \text{TRUE}(z))$$

(Here the symbol ‘$\equiv$’ indicates that the expression on the right is the definition of the expression on the left.) Next, we use $H(x)$ to construct a sentence $L$, thus:

$$(18) \quad L \equiv H[\hat{H(x)}]$$

Observe that as a special case of (16), we have

$$\forall z \ (\text{Sub}[\hat{H(x)}, \hat{H(x)}, z] \leftrightarrow z = \hat{H(H(x))})$$

Now expanding $H[\hat{H(x)}]$ via (17), we may rewrite (19) as follows:

$$\forall z \ (\text{Sub}[\hat{H(x)}, \hat{H(x)}, z] \leftrightarrow z = \exists z \ (\text{Sub}[\hat{H(x)}, \hat{H(x)}, z] \& \neg \text{TRUE}(z))$$

And abbreviating $\text{Sub}[\hat{H(x)}, \hat{H(x)}, z] \& \neg \text{TRUE}(z)$ by $D(z)$, we can further rewrite (19) as follows:

$$\forall z \ (D(z) \leftrightarrow z = \exists z \ (D(z) \& \neg \text{TRUE}(z)))$$

Now the reader will no doubt notice that (21) is formally identical to (13). Recall that (13) is logically incompatible with the disquotational schema (4); by exactly the same reasoning, so is (21), and therefore so is (19), of which (21) is a logical consequence. But (19) is just a special case of (16), so (16) itself is logically incompatible with the disquotational schema. Finally, I leave it to the reader to verify that when $L$ is defined as in (16), the statement (16) logically implies the biconditional $L \leftrightarrow \neg \text{TRUE}(\hat{L})$.

The foregoing discussion was necessarily rather technical, but here is its upshot. As long as we have, in our formal language, a formula for substitution that obeys (16), we will be able to construct a formula $D(x)$ which uniquely denotes a sentence $L$, which in turn says that the object denoted by $D(x)$ — i.e., the sentence $L$ itself — is not true. Thus, as long as we have (16), we can construct Liars in the strong, biconditional sense of the term, i.e., sentences satisfying the biconditional (3). Since that biconditional is logically incompatible with the disquotational schema (4), so is (16).
To sum up, we have found three rather innocuous-looking statements — (9), (13) and (16) — any one of which immediately yields a contradiction when it is paired with the disquotational schema (4). This is certainly a puzzle, since it is hard to see how we could be justified in rejecting (4), and it is also hard to see how we could be justified in rejecting (9), (13) and (16). This is a puzzle that Jacquette has really not solved, regardless of how one comes down on the issue of (2) vs. (3). Finally, this paradox is indeed a Liar paradox, since in each of (9), (13) and (16), a Liar sentence played an essential role.

5.

The considerations of the last section can actually be used to show that the disquotational schema is logically incompatible with Peano Arithmetic, subject to a certain convention about quotation names. I made this point all too briefly in (Barker, 2008), and the point was greeted with incredulity: “[Barker] actually claims to be able to deduce (**L**) as a substitution instance for a theorem of Peano Arithmetic!” (Jacquette, 2008, p. 151). Jacquette’s (**L**) is simply my (3), and yes, I do claim to be able to derive (an instance of) (3) from Peano Arithmetic. That this can be done is in fact quite well known, though evidently not quite well known enough. The basic technique can be found in (Gödel 1931), though Gödel does not actually derive an instance of (3) there.

To prove this, we must first get a few technical preliminaries out of the way. Let **L** be the language of arithmetic, and let **L’** be any countable language extending **L**. Let us assume an effective Gödel numbering of the formulas of **L’**. Finally, for any formula **F** of **L’**, let $$[F]$$ be the numeral for **F**’s Gödel number. That is, if $$n$$ is the Gödel number of **F**, then $$[F]$$ is simply the following term of **L**: $$s...s(0)$$, where ‘s’ denotes the successor function and where ‘s’ occurs $$n$$ times. Thus, $$[F]$$ is a term of **L** that denotes (under the standard interpretation of **L**) the Gödel number of **F**. Or more simply, $$[F]$$ is a quotation name of **F**, or rather, the closest thing to a quotation name of **F** that we can have in the language **L**.

Let us now specify that **L’** shall be the language that results from adding to **L** a single unary predicate **TRUE**. Then there is indeed a sentence $$L$$ of the language **L’** such that the biconditional $$L \leftrightarrow \neg \text{TRUE}([L])$$ is a theorem of Peano Arithmetic, Jacquette’s skepticism in this matter notwithstanding. This is all the more remarkable given that the predicate **TRUE** does not even belong to the language **L** in which Peano Arithmetic is formulated. The key to proving this remarkable result is the following, even more remarkable result, from (Gödel 1931), which is variously known as the “diagonal lemma” and the “self-reference lemma.”
Letting $\mathbf{L}'$ again be any countable extension of $\mathbf{L}$, and letting $F(x)$ be any formula of the extended language $\mathbf{L}'$, there is a sentence $A$, also of $\mathbf{L}'$, such that the following is a theorem of Peano Arithmetic:

\[(22) \ A \leftrightarrow F(\overline{A})\]

Finally, choosing the formula $\neg \text{TRUE}(x)$ for $F(x)$, we see that there is a sentence $L$ of $\mathbf{L}'$ such that the biconditional $L \leftrightarrow \neg \text{TRUE}(\overline{L})$ is derivable from Peano Arithmetic: this is simply a direct consequence of the diagonal lemma.

I used the diagonal lemma in just this way in (Barker, 2008). However, I did not bother to prove the diagonal lemma or provide a full citation, because frankly I thought it was something every logician knew. However, Jacquette seems to take issue with it, if I am correctly understanding the argument of section 10 of his paper. Be that as it may, the diagonal lemma is an entirely standard result, though different authors present the result in slightly different ways. The following quote from (Boolos and Jeffrey 1980) is fairly typical:

Here’s the diagonal lemma:

**Lemma 2**

Let $T$ be a theory in which $\text{diag}$ is representable. Then for any formula $B(y)$ (of the language of $T$, containing just the variable $y$ free), there is a sentence $G$ such that

\[\vdash_T G \leftrightarrow B(\overline{G}).\] (Boolos, & Jeffrey, 1980, p. 173)

Here, the notation $\vdash_T G \leftrightarrow B(\overline{G})$ means that the biconditional $G \leftrightarrow B(\overline{G})$ is a theorem of the theory $T$. Moreover, Peano Arithmetic is indeed a theory in which the function $\text{diag}$ is representable (as Boolos and Jeffrey prove elsewhere in their text). Finally, my convention for corner quotes is identical to that of Boolos and Jeffrey. Thus, the diagonal lemma in the form (22) does indeed follow from the diagonal lemma as proved by Boolos and Jeffrey.

Now Jacquette suspects trickery in my corner-quote convention; he suggests that by adopting this convention and thereupon deriving his biconditional (L*) (which is essentially the same as my (3)), I am “already committing a very irregular, and, I would say, fallaciously invalidating, equivocation” (Jacquette, 2008, p. 153). However, the difference between the corner quote convention that I am adopting in this section, and the convention adopted by Jacquette (and by me elsewhere in this paper), is actually quite minimal. The only difference is that under the present convention, $\overline{F}$ denotes the Gödel number of the formula $F$ instead of denoting $F$ itself. This, in turn, is done simply to accommodate the fact that the language of arithmetic, under its standard interpretation, is a
language for describing numbers, not formulas. However, given the ability to assign numerical codes (i.e., Gödel numbers) to formulas, this difference is more apparent than real. Now in dealing with Gödel numbers instead of dealing directly with sentences, we are also implicitly reinterpreting TRUE(x) to mean “x is the Gödel number of a true sentence” instead of its usual meaning, “x is a true sentence.” And with this in mind, we can see that the disquotational schema $A \leftrightarrow \text{TRUE}[\lceil A \rceil]$ should now be understood to be making the following statement: $A$ if and only if the Gödel number of $A$ is the Gödel number of a true sentence. This is obviously equivalent to the more usual interpretation of the disquotational schema, since clearly $A$ is true just in case its Gödel number is the Gödel number of a true sentence.

In any event, the disquotational schema (with the present convention about corner quotes) is logically incompatible with Peano Arithmetic, even though Jacquette apparently finds this fact hard to believe. Using the diagonal lemma, we may show that there is a sentence $L$ of $L'$ such that the biconditional $L \leftrightarrow \neg \text{TRUE}[\lceil L \rceil]$ is a theorem of Peano Arithmetic: this is simply an immediate consequence of the diagonal lemma as stated above. The biconditional $L \leftrightarrow \neg \text{TRUE}[\lceil L \rceil]$ is, in turn, logically incompatible with the sentence $L \leftrightarrow \text{TRUE}[\lceil L \rceil]$, which of course is an instance of the disquotational schema. Thus, on pain of contradiction, we cannot add a new predicate TRUE to the language of arithmetic, interpret it to mean “is the Gödel number of a true sentence,” and then express this by adopting the biconditional schema $A \leftrightarrow \text{TRUE}[\lceil A \rceil]$, unless we are also prepared to give up much of Peano Arithmetic.

This entire construction is essentially just a dramatic way of showing how easily the disquotational schema can come to grief. In particular, it shows that the biconditional (3) which Jacquette rejects, or rather a certain arithmetic version of (3), is actually a mathematical theorem, and is therefore not a good candidate for rejection. (Also, the astute reader will realize that the proof of the diagonal lemma is actually contained in section 4 above, and that the Liar sentence $L$ we just constructed here is essentially just the sentence $L$ defined in (18).)

6.

The last two sections demonstrate a certain logical instability in the disquotational schema. Namely, while the disquotational schema is logically consistent, it is inconsistent with various hard-to-deny propositions, which in turn involve sentences that self-attribute falsehood. Even if we were to agree that the conditional (2) properly expresses the idea that $L$ is a Liar sentence, it is hard to see how that would address the
issues raised in the last two sections. After all, the logical incompatibility of the disquotational schema with these hard-to-deny propositions does not turn on whether it is (2) or (3) that says that \(L\) is a Liar sentence. That said, it is still a fair question whether (2) or (3) says that \(L\) is a Liar. I will now argue that while (3) says that \(L\) is a Liar sentence, (2) does not say this at all.

In asking whether (2) says that \(L\) is a Liar sentence, a good place to start is to ask the question: could (2) be true even if \(L\) were not a Liar sentence? If the answer is yes, then it is hardly surprising that (2) fails to generate a contradiction. And indeed, the answer is yes: (2) can hold even if \(L\) is not a Liar sentence. This is actually quite easy to see if we recall what it takes to make a material conditional true: it is enough for its consequent to be true, or for its antecedent to be false. Now let \(S\) be any false sentence whatsoever. \(S\) might be the sentence “Gerbils are reptiles,” for example. Consider the conditional

\[(23) \; S \rightarrow \neg \text{TRUE}(\overline{S})\]

(23) is true, for the simple reason that \(S\) is false. We do not need to inquire into the truth of \(S\)’s consequent to determine this. Now clearly \(S\) is not a Liar sentence: it is simply an arbitrarily chosen false sentence. Yet (23) is true. Therefore, (23) does not say that \(S\) is a Liar sentence. And by precisely the same reasoning, (2) does not say that \(L\) is a Liar sentence. Even if we assume (2), \(L\) might, for all we know, simply say that Gerbils are reptiles.

Now admittedly, \(L\) could be a Liar sentence, it just need not be one. Any Liar sentence will satisfy (2): if \(L\) is interpreted as a Liar sentence, (2) comes out true. However, as we just saw, (2) also comes out true if \(L\) is simply interpreted as the statement that Gerbils are reptiles, or that the Statue of Liberty is made of foam rubber, or indeed as any false statement whatsoever. (2) is satisfied both by Liars and by non-Liars, and this is a feature that (2) shares with many other sentences. Consider, for example:

\[(24) \; L \leftrightarrow L\]

The sentence \(L\) in (24) could be a Liar sentence. Or it could be a false sentence, or it could be a true sentence. But I think everyone will agree that (24) does not say that \(L\) is a Liar sentence.

In defense of (2), Jacquette argues that while “a logical equivalence must be formulated as a biconditional […] not every sentence that is logically equivalent to this or that meaning or to this or that other sentence must itself be formulated as a biconditional” (Jacquette, 2008, p. 145). Thus, for example, even though the Liar sentence \(L\) is equivalent to the statement \(\neg \text{TRUE}(\overline{L})\), it does not follow that \(L\) must itself be formulated as a biconditional. I entirely agree, and I am in no way arguing that the
Liar sentence is, or is equivalent to, or must be formulated as a biconditional. It is Jacquette, not me, who insists on speaking of conditional and biconditional Liars. My point is that the statement *that* *L* is a Liar, not the statement *L* itself, must be formulated as a biconditional (3). Indeed, this conclusion would seem to follow from Jacquette’s own assertion that “a logical equivalence must be formulated as a biconditional.”

Jacquette continues: “Thus, the sentence ‘All triangles have three sides’ is logically equivalent to the sentence ‘All squares have four sides,’ since both sentences are logically, or, perhaps, analytically true; yet neither of the sentences is itself biconditional in logical form” (Jacquette, 2008, p. 145). Quite true; yet the statement *that* these two statements are equivalent must surely take the form of a biconditional, rendered semi-formally as: “All triangles have three sides ↔ all squares have four sides.” Likewise, the statement *that* *L* is a Liar must surely take the biconditional form *L* ↔ −TRUE(⌜L⌝), irrespective of how *L* itself is formalized. Again, I am not arguing that *L* itself is a biconditional.

All things considered, Jacquette has surprisingly little to offer by way of an argument that (2) says that *L* is a Liar sentence. His primary argument appears to be that (3) is the only other plausible candidate, and that (3) must be rejected because it conflicts with the disquotational schema. I would urge a different interpretation of this state of affairs, however. Namely, the disquotational schema has the following remarkable property: although it appears innocuous, and indeed analytic, it is in fact quite fragile, in the sense that it is rendered inconsistent by the simple construction of a sentence that self-ascribes falsehood. This is naturally a most unexpected feature of the disquotational schema. But it is also a rather well-established feature in light of the preceding sections. Indeed, I would argue that this surprising feature is the whole point of the Liar paradox, a point I will discuss in the final section of this paper.

In the course of this discussion, we have discovered a rather remarkable fact, against which the present dispute between Jacquette and myself pales in comparison. Namely, we have found that the disquotational truth schema is, in a certain sense, fragile: while not outright inconsistent, it is inconsistent with various statements, such as (9) or (13), or indeed (3) itself, that do nothing more than state the existence of Liar sentences. Since the mere existence of Liar sentences is hardly in dispute, it seems that we must, on pain of inconsistency, deny at least some instances of the disquotational schema. Yet it is very difficult, to put it mildly, to see how
it could ever be permissible to do this. After all, is not the disquotation
schema simply a definition of the truth predicate, or at the very least a
straightforwardly analytic statement about truth? How, then, can we deny
it?

This is indeed a real puzzle, and it is entirely understandable that
Jacquette would prefer to deny just about any statement rather than deny
the disquotation schema. (However, if Jacquette should choose to deny
(3), he ought at least to acknowledge that the sentence he is denying is
simply a statement to the effect that \(L\) is a Liar sentence.) Indeed, this is
such an interesting and perplexing puzzle that I wish I could say I
discovered it; unfortunately, Alfred Tarski beat me to it. The real
question, of course, is what to do about it. The following is a non-
exhaustive list of possible approaches, all of which are well-represented in
the literature.

First, the disquotation schema might be restricted in some way. This
need not be quite as radical as it sounds. In fact, it is essentially one of the
oldest approaches to the Liar, as it derives from the work of Tarski, and
before him, Russell (see, for example, Russell, 1908). The idea here is that
the predicate TRUE should apply only to those sentences that do not
themselves contain TRUE. This can be understood as a restriction on the
disquotation schema: namely, that \(\text{TRUE}(\![\text{A}]\!)[\leftrightarrow \text{A}]\) should only be
asserted when \(\text{TRUE}\) does not appear in \(\text{A}\). With some care, a
contradiction may be avoided in this manner, but at a price. It is pretty
clear that ordinary language imposes no such restriction; thus, this
approach might serve better as a revision of ordinary language than as a
description of it.

Second, classical logic itself may be questioned. Again, this is less
radical than it may sound. Classical logic assumes that every sentence has
a definite truth value, true or false. However, in reality this assumption is
questionable. Thus, in denying bivalence and thereby moving away from
classical logic, we may simply be extending classical logic, adapting it to a
new domain in which not every sentence has a definite truth value. In
particular, Liar sentences have long been felt to be good candidates for
such treatment. That is, they strike many as lacking definite truth values.
There is a great body of literature that explores the truth-value-gap
approach to the Liar, but by far the most important work in that literature
is (Kripke, 1975).

Finally, we may choose to live with the contradiction that the
disquotation schema generates instead of trying to avoid it. This is the
most radical approach of all, and its leading proponent is Graham Priest

---

5 See (Tarski, 1956), which clearly demonstrates this feature of the disquotation
schema, notwithstanding any alternative reading Jacquette may have of it.
Disquotation, Conditionals, and the Liar

(see, for example, Priest, 1979). Priest forthrightly believes that Liar sentences are simultaneously true and untrue, and he can therefore easily accommodate the full disquotational schema. A related view refrains from actually classifying the Liar as both true and untrue, but nonetheless regards the disquotational schema as in some sense analytic; I develop a view along these lines in “The Liar and the Inconsistency of Language” (under submission).

While they differ from each other in substantial ways, these and other approaches to the Liar all acknowledge that the disquotational schema, at least in the setting of classical logic, harbors inconsistencies. That is the fundamental problem about the Liar, the problem that the various approaches are responding to. And that is the problem that is dodged, not answered, by putting a conditional where a biconditional should be.

References