Synthese (2017) 194:5023–5037
DOI 10.1007/s11229-016-1191-0

Self-referential propositions

Bruno Whittle[1]

Received: 11 January 2016 / Accepted: 9 August 2016 / Published online: 26 September 2016

Abstract Are there 'self-referential' propositions? That is, propositions that say of themselves that they have a certain property, such as that of being false. There can seem reason to doubt that there are. At the same time, there are a number of reasons why it matters. For suppose that there are indeed no such propositions. One might then hope that while paradoxes such as the Liar show that many plausible principles about sentences must be given up, no such fate will befall principles about propositions. But the existence of self-referential propositions would dash such hopes. Further, the existence of such propositions would also seem to challenge the widespread claim that Liar sentences fail to express propositions. The aim of this paper is thus to settle the question–at least given an assumption. In particular, I argue that if propositions are structured, then self-referential propositions exist.

Keywords Self-referential propositions · Paradoxes for propositions · Truth · The liar paradox

Are there 'self-referential' propositions? That is, propositions that say of themselves that they have a certain property, such as that of being false. There can seem reason to doubt that there are. There are of course self-referential sentences, such as 'the proposition expressed by this sentence is false'. But a standard response to these is to deny that they express propositions, in which case the existence of such sentences would not entail that of self-referential propositions.

At the same time, there are a number of reasons why the question of whether there are such propositions is significant. The first is as follows. Suppose that there

✉ Bruno Whittle
bruno.whittle@glasgow.ac.uk

1 Department of Philosophy, University of Glasgow, Glasgow G12 8QQ, UK

are indeed no such propositions. One might then hope that while paradoxes such as the Liar must be grappled with in giving an account of language, one can give an account of propositions—and of propositional attitudes and acts, such as belief and assertion—entirely untroubled by such things. That is, one might hope that although the Liar shows that many plausible principles about sentences must be given up, no such fate will befall principles about propositions.

Consider, for example, the truth-schema for sentences (i.e. '*A*' is true iff *A*, for a sentence *A*). The existence of Liar sentences (e.g. 'this sentence is not true') seems to give us strong reasons to reject this: since it shows that we cannot maintain it without being classically inconsistent.[1] On the other hand, if there are no Liar *propositions* (e.g. propositions that say of themselves that they are not true), then one might hope to maintain the truth-schema for these while remaining classically consistent.[2] One might even hope to keep both the truth-schema and classical logic for propositions. But the existence of self-referential propositions—in particular, Liar propositions—would frustrate such hopes.

Similarly, the existence of Liar sentences seems to provide strong reasons to reject bivalence for sentences (i.e. the principle that every sentence is either true or false). But one might hope to maintain this for propositions.[3] Again, however, the existence of self-referential propositions would seem to dash such hopes.

A further group of reasons why the question of whether there are self-referential propositions matters is as follows. As I have in effect noted, a standard response to Liar sentences is to deny that they express propositions.[4] But if there are self-referential propositions, i.e. Liar propositions, then we would have a version of the paradox whose solution would require something fundamentally different from this standard move. Indeed, the existence of such propositions would seem to challenge this traditional claim about Liar sentences. For standard arguments for this claim use either the truth-schema or bivalence for propositions (which would be challenged by the existence of Liar propositions). Further, since the sentential and propositional versions of the paradox would seem to be very similar, it would seem desirable to give similar solutions. But then, since the solution in the propositional case will not deny that something expresses a proposition, it seems that the solution in the sentential case shouldn't either.

All of this makes a suggestion of Saul Kripke's particularly intriguing. For in his celebrated paper on truth (in a footnote, no less) he suggests that as long as propositions are 'structured', it may be possible to apply Gödelian techniques for generating self-reference directly to them (1975, p. 713). The result would be a range of self-referential

1 Thus, most recent work on truth for sentences does indeed reject it: e.g. Kripke (1975), Gupta (1982), Herzberger (1982), McGee (1991), Gupta and Belnap (1993), Maudlin (2004) and Leitgeb (2005). (Although there are exceptions to this trend, such as Priest (1979, 1987/2006), Field (2008) and Beall (2009)).

2 For example, Sobel (1992) and Glanzberg (2001) maintain the truth-schema for propositions, and certainly do not mean to embrace classical inconsistency. Indeed, Glanzberg goes so far as to write: I doubt that anything that failed to validate [it] could count as a reasonable theory of propositions (2001, p. 228).

3 For example, Skyrms (1984), Sobel (1992) and Gaifman (2000) all give up bivalence for sentences but maintain it for propositions.

4 See, e.g., Skyrms (1984), Sobel (1992), Gaifman (2000) and Glanzberg (2001).

propositions, including Liar propositions. Thus, since propositions are indeed structured on many of the most popular—and apparently most plausible—accounts, this would seem to be highly significant. So it is surprising that Kripke's suggestion does not seem to have been pursued. The aim of the present paper, however, is to pursue it and show it to be correct.

The structure of the paper is as follows. Section 1 contains preliminaries. Section 2 outlines the construction of self-referential propositions. Section 3 considers objections. Section 4 spells out why the existence of such propositions matters. And Sect. 5 goes through the construction of such propositions in full.

1 Preliminaries

1.1 Propositions

For the purposes of this paper, then, I assume that propositions are 'structured', i.e. in a way that mirrors the structures of the sentences that express them.[5] This is true on the traditional Fregean and Russellian accounts, and I would argue that it is likely to be true on any adequate account—but that is not of course a case that I will make here.[6] Indeed, for definiteness, I will assume a broadly Russellian approach (unless otherwise stated). That is, I will assume that the proposition that John is tall, for example, is a structured entity built out of John together with the property of being tall. (I will say much more about how I propose to think about propositions in Sect. 5). However, everything that I will say could easily be made compatible with a Fregean approach, or any other on which propositions have a sentence-like structure.

1.2 Self-reference

A *self-referential* proposition is one that says of itself that it has a certain property and that does not say anything else. Thus, an atomic proposition $F(p)$ (i.e. the proposition that p is F) such that $p = F(p)$ would be self-referential. That is, if there really are such propositions, then they would be self-referential (although that there are is something that one might doubt: see below). Similarly, a negated atomic proposition $\neg G(q)$ such that $q = \neg G(q)$ would be self-referential. As would be a proposition r of the form $\forall x(H(x) \rightarrow J(x))$ such that H applies precisely to r. And so on. This is not a completely precise characterization, but it will suffice for the purposes of this paper.

It is hard to deny that there are self-referential sentences. As Kripke pointed out, we can produce one simply by baptizing the string 'Jack is short': Kripke (1975, p. 693). In contrast, it seems very far from obvious that there are self-referential propositions. For, assuming that propositions are Russellian, one analogous to 'Jack is short' would be of the form $S(p)$, where S is the property of being short and $p = S(p)$. But then

[5] However, if there is a mismatch between the surface and the logical form of the sentence, then it is the structure of the latter that is mirrored.

[6] On structured propositions see, e.g., Salmon (1986), Soames (1987, 2010), Kaplan (1989), Fine (2007) and King (2007).

p would have *itself* as a constituent: something that is plausibly impossible (just as it is plausibly impossible for a set to contain itself). If propositions are instead Fregean, then an atomic proposition about itself would not have itself as a constituent. But it would be made out of a mode of presentation of a proposition made out of that very mode, which, again, is plausibly impossible.

One might try rather to produce a self-referential proposition via a sentence such as 'the proposition expressed by this sentence is false'. But, as I noted, a standard response to such sentences is to deny that they express propositions.

1.3 An alternative approach

An approach to the question of whether there are self-referential propositions that is very different from that which I will pursue is that of Barwise and Etchemendy (1987). That work gives an account of truth focused on propositions, and self-referential propositions play a central role. However, this account does not in fact seem well-suited to establishing the existence of such propositions.

It uses the non-wellfounded set theory of Aczel (1988) to provide models of such propositions. For example, a proposition $\neg T(p)$ that says of itself that it is untrue (i.e. such that $p = \neg T(p)$) is modelled by something like a set that contains itself (more precisely: a set that belongs to its own transitive closure[7]). But if one is unsure whether there is a proposition of this form about itself—for example, on the grounds that such a proposition would have to have itself as a constituent—then one is unlikely to be convinced by the existence of such models. After all, the existence of such a model no more establishes that there really is such a proposition than the existence of non-standard models of arithmetic establishes that there really is a natural number with infinitely many predecessors. A similar point can be made about any of the other models of self-referential propositions given in that work, all of which employ non-wellfounded sets in a similar way. In contrast, the approach pursued below does not use anything like these non-wellfounded models and, in part because of this, would seem much better suited to establishing that there really are self-referential propositions.

2 Self-referential propositions: outline

The idea is thus to construct self-referential propositions using a version of Gödel's 'diagonal' function. In this section I will outline the construction, and show how the resulting propositions give rise to propositional versions of paradoxes such as the Liar, i.e. versions that do not involve expressions, mental states or similar. The construction will be given in full in Sect. 5.

Thus, let Z be the proposition $0 = 0$. And let d be a function such that for any proposition p, $d(p)$ is the result of replacing all occurrences of Z in p with p itself (if p *is* Z, then $d(p)$ is simply p). For example, if p is the proposition $\neg Z$, then $d(p)$ is $\neg\neg Z$, whereas if p is the proposition $T(Z)$ (i.e. the proposition that Z is true), then

[7] The transitive closure of a set x is the set whose members are the members of x, the members of the members of x, the members of the members of the members of x, etc.

$d(p)$ is $T(T(Z))$. How d behaves with arguments that are not propositions will not matter, but for definiteness let's assume that d sends any such thing to 0.

This function seems straightforward, and so prima facie it seems hard to deny that it exists (although I will consider attempts in Sect. 3). However, given such a function, it is straightforward to construct self-referential propositions, such as one that says of itself that it is untrue.

Before giving the construction, I should make clear how I will think about functions in connection with propositions, that is, how functions are constituents of propositions. Specifically, I will take it that just as formulas of standard formal languages can contain function symbols in addition to names and predicate symbols, so propositions can contain functions in addition to objects and properties. For example, the proposition that the successor of 0 is a number, i.e. $N(s(0))$, is built from N, s and 0, and is thus distinct from $N(1)$, which is built simply from N and 1.

The alternative—to identify such propositions—would seem unnatural. For consider the proposition that s is an injection, for example: $\forall x \forall y(s(x) = s(y) \rightarrow x = y)$. There does not seem to be any way of conceiving of this, except as containing s. But then it is hard to see why one should deny that the instances of this proposition—or $N(s(0))$—also contain this function.[8]

I should add, however, that nothing essential turns on this stance about functions. If one objects to it, one could construct self-referential propositions in a similar way, but using relations rather than functions (see Sect. 3). However, it is simplest to use functions, and so that is the construction that I focus on.

Thus, given d we can construct a self-referential proposition as follows:

$$\neg T(d(\neg T(d(Z)))),$$

i.e. the proposition that $d(\neg T(d(Z)))$ is untrue. Call this LP (for 'Liar proposition'). LP says of itself that it is untrue: for $d(\neg T(d(Z)))$ is just LP (this is what you get if you replace every occurrence of Z in $\neg T(d(Z))$ with $\neg T(d(Z))$ itself).

We are then led, via plausible steps, to contradiction. For, as we have seen:

(1) $d(\neg T(d(Z))) = \text{LP}$.

But, for any proposition p, $\neg T(d(p))$ is the proposition that $d(p)$ is not true. So it seems that $\neg T(d(p))$ is true iff $d(p)$ is not. In particular:

(2) LP is true iff $d(\neg T(d(Z)))$ is not.

But then by (2) and (1) we have:

(3) LP is true iff it is not.

Further, one can construct a whole range of self-referential propositions in a similar way, giving rise to propositional versions of every other paradox (or puzzle) that results

[8] In support of the identification of $N(s(0))$ and $N(1)$ one might note that it is natural to describe the sentences '$s(0)$ is a number' and '1 is a number' as being 'about the same thing' (i.e. 1). However, there are many other cases where we would give a comparable description, but where we would certainly *not* want to say that the things in question are constituents of the propositions expressed. For example, it is natural to describe 'all odd primes are φ' and 'all primes greater than two are φ' as being 'about the same things', but we would not want to say that these infinitely many numbers are constituents of the propositions in question.

from a self-referential sentence. For example, a proposition that says of itself that it is true will give a propositional version of the truth-teller; one that says of itself that if it is true, then $0 = 1$, will give a version of Curry's paradox; and so on.

Let M be the proposition Mont Blanc = Mont Blanc. Then, for any proposition p, we can construct a proposition p^* that says of itself exactly what p says of M. Thus, if p is $T(M)$, then p^* will say of itself that it is true, whereas if p is $T(M) \to 0 = 1$, p^* will say that if it is true, then $0 = 1$. Here is how to do this: assuming, for simplicity, that Z does not occur in p. First, let p' be the result of replacing all occurrences of M in p with $d(Z)$. Then, to obtain p^*, replace all of the occurrences of Z in p' with p' itself (i.e. $p^* = d(p')$). This proposition will say of itself whatever p said of M. For p^* clearly says of $d(p')$ whatever p said of M (since the occurrences of $d(p')$ in p^* are precisely those that replaced occurrences of M in p). But $d(p')$ is p^*.

Indeed, we can use a similar method to produce propositions that form more complex networks. For example, propositions p and q such that p says that q is true, while q says that p is not, or an infinite sequence of propositions with the structure of Yablo's paradox.

For example, to produce p and q consider a function e as follows, using I_n for the proposition $n = n$: for propositions r and t, $e(r, t)$ is the result of replacing, in r, all occurrences of I_1 with r, and all occurrences of I_2 with t. Then, if

$p' = T(e(I_2, I_1))$, and
$q' = \neg T(e(I_2, I_1))$,

we are done by letting

$p = T(e(q', p'))$, and
$q = \neg T(e(p', q'))$

(it is easy to see that $e(q', p') = q$ and $e(p', q') = p$).

To generate an infinite sequence of propositions, where each member talks about later ones, we proceed as follows. Consider a function f such that, for any propositions $p_0, p_1 \ldots, p_n, \ldots, f(p_0, p_1, \ldots, p_n, \ldots)$ is the result of replacing, in p_0, all occurrences of I_i with p_i, for each $i \geq 1$.[9] We can then produce an infinite sequence of propositions, where each says that the next is true, for example, as follows. First, for each $n \geq 1$, let

$$q'_n = T(f(I_{n+1}, I_1, I_2, \ldots)).$$

Then if

$$q_n = T(f(q'_{n+1}, q'_1, q'_2, \ldots)),$$

$q_1, \ldots, q_n, \ldots$are as required.[10]

[9] That is, f is an ω-ary function, with a place for each natural number. If desired, one could give a similar example using a unary function from sequences of propositions.

[10] The more complicated case, where each proposition says something about *all* subsequent ones, is handled as follows. Here is an example where each proposition says that all later ones are untrue(giving a

I will give these constructions more carefully in Sect. 5. That is, I will show how they can be carried out within a natural, but more precisely stated, account of propositions. First, however, I will consider objections to the claim that propositions along the lines constructed above exist (Sect. 3), and then—having answered these—I will explain why the existence of such propositions would seem to be significant (Sect. 4). (The outline above will suffice for these discussions.) I will focus on the example of LP, but similar points apply to the whole range of propositions constructed above.

3 Objections

The most obvious possible objection is as follows. It is essential to the above construction that d can apply to propositions that contain this function. What, then, about simply denying that a function can ever apply to a proposition that contains it?

To be effective as a means of denying that there are Liar propositions, one would similarly have to deny that a property can apply to a proposition that contains it. For although the construction above was in terms of a function, one could just as well use a 2-place property D as follows: for any propositions p and q, $D(p, q)$ iff q is the result of replacing every occurrence of Z in p with p itself. Now consider:

$$\forall x(D(r, x) \rightarrow \neg T(x)),$$

where r is: $\forall x(D(Z, x) \rightarrow \neg T(x))$. It is easy to see that this says of itself that it is untrue, just as LP does.

What, then, about denying that a function or property can apply to a proposition that contains it? This would seem to be unacceptably restrictive. For example, there could be no property of being known K that could apply to the proposition that Z is known in this sense (i.e. $K(Z)$). Rather, one would need a hierarchy of properties of being known—leading to problems similar to those faced by Tarski's approach to truth.[11] Such a blanket prohibition on functions and properties applying to propositions that contain them would thus seem unacceptable.

If properties, for example, *can* apply to propositions that contain them, then they cannot be modelled by sets. More precisely, given the standard account of sets (i.e. Zermelo-Fraenkel set theory with urelements, ZFU), one cannot *both* model an n-place property with the set of ordered n-tuples that it applies to, *and* model a proposition with a set-theoretic construction of its constituents (i.e. a set whose transitive closure

Footnote 10 continued
propositional version of Yablo's paradox).

$$\begin{aligned} q'_n &= \neg T(f(I_{n+1}, I_1, I_2, \ldots), f(I_{n+2}, I_1, I_2, \ldots), \ldots) \\ q_n &= \neg T(f(q'_{n+1}, q'_1, q'_2, \ldots), f(q'_{n+2}, q'_1, q'_2, \ldots), \ldots) \end{aligned}$$

Here $\neg T(p_1, p_2, \ldots)$ is shorthand for: $\neg T(p_1) \wedge \neg T(p_2) \wedge \ldots$. The sequence $q_1, \ldots, q_n, \ldots$ is then as required. One could give a similar example without infinite conjunctions, but for reasons of space I omit the details.

[11] For this approach, see his (1935). For its problems, see, e.g., Kripke (1975) and Soames (1999).

contains these constituents). Since in ZFU no set can belong to its own transitive closure, whereas the members of an ordered n-tuple do belong to the transitive closure of that n-tuple, and thus to the transitive closure of any set whose transitive closure contains the n-tuple. Similarly, if a function can apply to a proposition that contains it, then—given standard set theory—one cannot *both* model an n-place function with a set of ordered $n + 1$-tuples (in the familiar way), *and* model a proposition with a set-theoretic construction of its constituents. However, as the above example with the property of being known makes clear, even if not being able to model properties and functions in ZFU in this way is a cost, it is one that can be avoided only at the apparently much higher one of a severely restrictive account of propositions.

But one might at this point have the following concern. As I noted in Sect. 1, one can apparently quite reasonably reject as impossible a proposition of the form $F(p)$ such that $p = F(p)$, on the grounds that such a proposition would have to have itself as a constituent. Am I now arguing that we *should* accept that a property or function can in some sense 'contain' itself? Absolutely not. I am arguing that we should accept that a property or function can apply to a proposition that contains it. But this entails that the property or function 'contains' itself (in some sense) only if we think of a property or function as 'containing' the things that they apply to, and that seems unmotivated once we recognize the limits of standard set theory when it comes to modelling properties and functions.

What—as an alternative objection—about denying simply that this particular function d exists (and similarly that the property D does)? After all, I have shown that d can be used to generate paradoxes. Isn't that already ground for denying its existence? No—for the following reason. To determine the value of d for a given proposition p, all one needs to know is where one particular entity (i.e. Z) occurs in p. One does not need to know anything about which things the properties in p apply to, or about which values the functions in p take. Thus, given some straightforward notation for propositions (such as that which I am using, or that of Sect. 5), one could easily write a computer program that, given the notation for a proposition q as input, gives that of $d(q)$ as output. It would thus seem incredible to respond to the paradox that LP give rise to—not by revising our naive theory of T—but rather our naive theory of d. After all, when we consider LP, and other propositions involved in the paradox, it may not be clear whether T applies to them, but it is completely clear what value d takes when applied to them.

The claim that d exists is also supported by the fact that it would be extremely difficult, if not impossible, to give an adequate account of propositions without resources that would enable one to express a function that is at least coextensive with d (and which would thus generate paradox in just the way that d does). This is just a propositional version of the familiar point that an adequate account of the syntax of a standard formal language would seem to require resources sufficient to express (a function coextensive with) a sentential version of d.

Thus, d exists. But then—given that $\neg$, T and Z of course also exist—it is hard to see what could prevent a proposition from resulting from their straightforward combination as in LP.

4 Significance

So self-referential propositions such as LP exist. Why does that matter?

The first group of reasons concern prima facie plausible principles about truth, for sentences and propositions. Thus, Liar sentences show that one cannot maintain the truth-schema for sentences while remaining classically consistent. But one might hope to maintain this for propositions while so remaining.[12] Indeed, one might hope to maintain both this schema and classical logic for propositions. But LP dashes such hopes.

For to maintain the truth-schema for propositions is to accept every proposition of the following form, for a proposition p.

(TP) $T(p) \leftrightarrow p$

But we have seen that the following proposition is true.

(P1) $d(\neg T(d(Z))) = \neg T(d(\neg T(d(Z))))$

While the following is an instance of (TP).

(P2) $T(\neg T(d(\neg T(d(Z))))) \leftrightarrow \neg T(d(\neg T(d(Z))))$

But (P1) and (P2) are classically inconsistent. So one cannot maintain (TP) while being classically consistent, and one cannot maintain it together with classical logic (since it entails everything in that logic).

Consider next bivalence. Many approaches to truth give this up for sentences—on the basis of considerations about Liar sentences—but hold on to it for propositions.[13] However, LP gives us reasons for rejecting bivalence for propositions that seem every bit as strong as those that Liar sentences give us for rejecting it for sentences.

For example, one argument against bivalence for sentences uses the following rules, where Tr and Fa mean sentential truth and falsity, respectively.

(TR) Tr(‘A’) / A
(FR) Fa(‘A’) / $\neg A$

For suppose that $c =$ ‘$\neg$Tr(c)’. We can then derive $\bot$ from Tr(c): we have Tr(‘$\neg$Tr(c)’) (by $c = \neg$Tr(c)), and then $\neg$Tr(c) (by (TR)). And we can also derive $\bot$ from Fa(c): we get Fa(‘$\neg$Tr(c)’), then $\neg\neg$Tr(c) (by (FR)), and then Tr(c); and then $\bot$ as before. But then it seems that we should reject Tr(c) $\vee$ Fa(c)—an instance of bivalence for sentences. However, LP of course allows us to give just the same argument against bivalence for propositions, using propositional versions of (TR) and (FR). More generally, it seems that any argument that uses Liar sentences to make trouble for bivalence

12 See, e.g., Sobel (1992) and Glanzberg (2001). I should note that in that work Glanzberg represents propositions as sets of possible worlds. But he is quite clear that this is merely a simplifying assumption, and that the claims of the work are not supposed to make essential use of this. He writes:

The Liar paradox…is insensitive to issues of how finely structured propositions must be. Thus, we may take the possible worlds view of propositions as at least a simplifying assumption, regardless of whether the familiar arguments, such as those of (Soames 1987), ultimately show propositions to be structured entities (2001, p. 245).

13 For example, Skyrms (1984), Sobel (1992) and Gaifman (2000).

for sentences, will correspond to an equally convincing one that uses LP to make trouble for bivalence for propositions. Thus, if Liar sentences should lead us to give up the sentential principle, then it seems that LP should lead us to give up the propositional one too.

The second group of reasons that the existence of LP matters concern the traditional claim that Liar sentences fail to express propositions. For LP of course gives rise to a paradox that cannot be solved by anything like this standard move. But, further, it in fact seems to challenge this traditional claim. The first reason that it does this is simply that standard arguments for this claim use propositional versions of either the truth-schema or bivalence. But, as we have just seen, LP challenges these.

For example, one such argument is as follows.[14] Consider a Liar sentence of the form 'this sentence does not express a true proposition'. That is, suppose that $b =$ '$\neg\exists p(\text{Exp}(b, p) \wedge T(p))$' (where $\text{Exp}(x, q)$ means that x expresses q). It is surely the case that *if* b expresses a proposition, than that proposition is $\neg\exists p(\text{Exp}(b, p) \wedge T(p))$. That is, $\forall q(\text{Exp}(b, q) \rightarrow q = \neg\exists p(\text{Exp}(b, p) \wedge T(p)))$. But now suppose $\text{Exp}(b, r)$. By (TP) and what we have just seen, $T(r) \leftrightarrow \neg\exists p(\text{Exp}(b, p) \wedge T(p))$, and thus $\neg T(r)$. That is, $\forall p(\text{Exp}(b, p) \rightarrow \neg T(p))$, giving $\neg\exists p(\text{Exp}(b, p) \wedge T(p))$; which, by (TP) again, gives $T(r)$—contradiction. Thus, $\neg\exists p\text{Exp}(b, p)$. But of course this argument is called into question once (TP) is.

Alternatively, one might argue for the claim that Liar sentences do not express propositions using bivalence for propositions, together with (a) the claim that Liar sentences are neither true nor false, and (b) the claim that if x expresses p, then x is true (false) if p is true (false). But this argument is also called into question once bivalence for propositions is.

Finally, the existence of LP also challenges the claim that Liar sentences fail to express propositions for the following reason. The paradox that LP gives rise to is obviously very similar to that which these sentences give rise to. It would thus seem desirable to give similar solutions. But the solution in the propositional case will not involve anything like the claim that something fails to express a proposition—so it seems that the solution in the sentential case shouldn't either.

5 Self-referential propositions: in full

I will now fill in the outline of Sect. 2. More precisely, I will show how the constructions of that section can be carried out in a natural, but more precisely stated, account of propositions. This is essential, because a certain difficulty emerges as soon as one tries to think clearly about the nature of propositions.

Specifically, the following.[15] On the usual way of thinking about things, the proposition that 1 is a (natural) number, for example, is something like the ordered pair of the property of being a number N and 1: $\langle N, 1\rangle$. What, then, about the proposition that the successor of 0 is a number, which (as explained in Sect. 2) one wants to distinguish

[14] See, e.g., Sobel (1992) and Glanzberg (2001).

[15] This difficulty is mentioned in Kaplan (1989, p. 496). Kaplan does not say in any detail how it should be solved, but the solution below is in the general direction that he suggests.

from this? It seems natural to think of this as (something like) another ordered pair, but this time of N together with a complex of the successor function s and 0, i.e. something like $\langle s, 0\rangle$. So the proposition would be $\langle N, \langle s, 0\rangle\rangle$. But the problem is now easy to see. For if the result of combining 1 with N to form $\langle N, 1\rangle$ is the proposition that 1 is a number, then shouldn't the result of combining $\langle s, 0\rangle$ with N *in just the same way* be the proposition that $\langle s, 0\rangle$—i.e. this complex—is a number? And, if not, then what is the proposition that this complex is a number?

The most natural way of solving this difficulty would seem to be as follows. Give up on the idea that the proposition that 1 is a number is anything like $\langle N, 1\rangle$. Rather, it is (something like) the pair of N and (something like) the 'ordered single' of 1. I use [1] for the latter component. I will state the resulting account of propositions more precisely below. But the way in which it solves the difficulty—i.e. allows us to distinguish propositions that *use* complexes from those that *mention* them—is this. The proposition that 1 is a number is $\langle N, [1]\rangle$. The proposition that the successor of 0 is a number is $\langle N, [s, [0]]\rangle$, where $[s, [0]]$ is a complex of s and [0] (here the complex is used). Finally, the proposition that this complex is a number is $\langle N, [[s, [0]]]\rangle$ (where, of course, $[[s, [0]]]$ is the ordered single of $[s, [0]]$; here the complex is mentioned). And these last two propositions are distinct because $[s, [0]] \neq [[s, [0]]]$—solving the difficulty.

More generally, propositions are as follows.

5.1 Simple terms

I assume that for every object x, there is an object $[x]$, called a *simple* (*propositional*) *term*. I call x the *constituent* of $[x]$.

What exactly is an object? All that I will assume is that numbers are objects, as are simple terms, and complex terms and propositions (to be introduced below). But if one does not think that terms or propositions are *really* objects, then one can simply read my uses of 'object' as 'object, term or proposition'.

In ZFU, one can easily define (ordered) n-tuples, for $n \geq 1$, such that the following is satisfied.[16] ('X' is for 'extensionality'.)

(X) If the m-tuple of $x_1, \ldots, x_m$ (in that order) is identical to the l-tuple of $y_1, \ldots y_l$ (in that order), then $m = l$ and $x_1 = y_1, \ldots, x_m = y_l$.

In the following, I assume that we have settled on one such definition, and use n-tuple (and similar) to mean n-tuple (and similar) so defined.

The ordered single of x is at least a natural model of $[x]$. I will give similar models of complex terms and propositions below. Now, on one version of the account being proposed, terms and propositions would in fact be identified with these models. However, this version would seem to face a problem similar to that which Benacerraf (1965) raises for identifications of numbers with sets. That problem is simply that since there are multiple, apparently equally good, ways of modelling numbers with sets, any such identification would appear arbitrary. But any attempt to identify terms

[16] For example, one can define the n-tuple of $x_1, \ldots, x_n$ as $\{[\![1, x_1]\!], \ldots, [\![n, x_n]\!]\}$, where $[\![i, x_i]\!]$ is $\{\{i\}, \{i, x_i\}\}$.

and propositions with sets would seem to face a similar issue: why use one definition of n-tuples rather than another, for example?[17]

In the number case, the natural lesson would seem to be that these are sui generis objects, and so not reducible to sets or anything else. And the natural lesson in the propositional case would seem to be similar. For this reason, I will not assume that the models of this section tell us what terms and propositions really are. Rather, they are meant simply to convey the sort of way in which these are constructed from their basic constituents, such as objects, functions and properties.

Indeed, all that one needs to assume about simple terms is the following, which of course holds in the suggested set-theoretic model by (X).

(XS) If $[x] = [y]$, then $x = y$.

5.2 Complex terms

I assume that there are also *complex* (*propositional*) *terms* generated recursively as follows: if $n \geq 1$, f is an n-place function, and $t_1, \ldots, t_n$ are simple or *complex terms*, then there is a *complex term* $[f, t_1, \ldots, t_n]$. Once again, $f, t_1, \ldots, t_n$ are the *constituents* of $[f, t_1, \ldots, t_n]$. And a natural model of $[f, t_1, \ldots, t_n]$ is the $n+1$-tuple of $f, t_1, \ldots, t_n$ (in that order).

To be clear, I am not suggesting that set theory can be used to give natural models of *functions*, at least not in the standard way (see Sect. 3). Rather, I am suggesting that it can be used to give a natural model of $[f, t_1, \ldots, t_n]$—but in this model f is an urelement. What then are functions? It is beyond the scope of this paper to say in any detail. All that is required for my purposes here is that certain straightforward ones exist (i.e. d and the variants of it considered in Sect. 2). But I would suggest thinking of functions as sui generis objects—as ways of going from one object to another (or from a number of objects to another)—in the same way that it is plausible to think of numbers or sets as sui generis objects.

The natural assumptions about complex terms are as follows.

(XC) If $[f, t_1, \ldots, t_n] = [g, u_1, \ldots u_m]$, then $n = m$ and $f = g, t_1 = u_1, \ldots, t_n = u_m$.

(SC) No complex term is a simple term.

Again, these hold in the suggested set-theoretic model by (X). It is (SC) that ensures that the difficulty raised at the start of this section is solved.

5.3 Atomic propositions

Next, I assume that if $n \geq 1$, H is an n-place property, and $t_1, \ldots, t_n$ are (simple or complex) terms, then there is an object $\langle H, t_1, \ldots, t_n \rangle$, called an *atomic proposition*. $H, t_1, \ldots, t_n$ are the *constituents* of $\langle H, t_1, \ldots, t_n \rangle$.[18] A natural model of $\langle H, t_1, \ldots, t_n \rangle$ is

[17] For discussions of versions of this problem for accounts of propositions, see Jubien (2001), King (2007, pp. 47–50, 127–36) and Soames (2010, pp. 91–94).

[18] Note that the defined use of ‘constituent’ is slightly narrower than that of previous sections. For on the former 0 is not a constituent of the proposition $\langle N, [0] \rangle$; rather, it is a constituent of a constituent of this proposition.

again the corresponding $n+1$-tuple. I use different brackets for terms and propositions just for readability.

The natural assumptions are: (XA), i.e. the analogue of (XC) for atomic propositions, and that no atomic proposition is a term. These hold in the suggested model by (X), as long as no property is a function (which I assume).

$\langle H, [a]\rangle$ is the proposition that a is H, $\langle H, [f, [a]]\rangle$ is the proposition that $f(a)$ is H, and so on.

Thus, $\langle N, [1]\rangle$ is the proposition that 1 is a number; $\langle N, [s, [0]]\rangle$ is the proposition that $s(0)$ is a number; and $\langle N, [[s, [0]]]\rangle$ is that to the effect that $[s, [0]]$ (i.e. this complex term) is a number. The first and second are distinct by (XA), together with the fact that no simple term is a complex one (i.e. (SC)); and the second and third are distinct for the same reason. So the difficulty raised at the start of this section is solved as anticipated.

(What about the distinctness or otherwise of the first and third propositions? These will indeed be distinct as long as 1 is distinct from the complex term $[s, [0]]$. This is highly plausible. However, it does not follow from the explicit assumptions above, since none of these have any bearing on what numbers are.)

Here is a further example that will help make clear the way in which the account works—and the way in which it allows us to distinguish propositions that are about (i.e. mention) terms from those that have them as constituents (i.e. use them). Thus, let i be the identity function. Then the proposition that $i(0)$ is 0, that is, $\langle =, [i, [0]], [0]\rangle$, is of course true. —Despite the fact that the two *constituents* of this proposition are distinct (by (SC)). For the proposition is not *about* the constituents. What (SC) commits us to is rather the falsity of $\langle =, [[i, [0]]], [[0]]\rangle$ (the identity proposition that mentions the terms that the last proposition used).

5.4 Compound propositions

Finally, I assume that there are *negated* and *conjoined propositions* generated recursively as follows: if p and q are atomic, *negated* or *conjoined propositions*, then there is a *negated proposition* $(\neg, p)$ and a *conjoined proposition* $(\wedge, p, q)$. I also call *negated* and *conjoined propositions negations* and *conjunctions* (respectively); $\neg$ and p are the *constituents* of $(\neg, p)$, and $\wedge$, p and q are those of $(\wedge, p, q)$. The idea is that $\neg$ and $\wedge$ are abstract entities corresponding to the English words 'not' and 'and' (respectively), but nothing will turn on what exactly these are. Negations and conjunctions are once again naturally modelled by pairs and triples (respectively). The natural assumptions are similar to those in the atomic case: i.e. $(\neg, p) = (\neg, q)$ entails $p = q$; no negation is a conjunction, atomic proposition or term; and similarly for conjunctions. These will hold in the model given (X), together with the assumption that no function is a property.

If H and J are 1-place properties and a and b are objects, then $(\neg, \langle H, [a]\rangle)$ is the proposition that it is not the case that a is H; and $(\wedge, \langle H, [a]\rangle, \langle J, [b]\rangle)$ is that to the effect that a is H and b is J. And so on.

One could straightforwardly extend this account to quantified propositions, but for reasons of space I will not do this here.

5.5 Self-referential propositions

We now have a precisely stated framework in which to construct self-referential propositions, following the outline of Sect. 2.

Thus, let Z be the proposition $\langle =, [0], [0] \rangle$. And let d be a function such that for any proposition p, $d(p)$ is the result of replacing all occurrences of Z in p with p itself (as before, $d(Z) = Z$). For example, if p is $(\neg, Z)$, then $d(p)$ is $(\neg, (\neg, Z))$; and if p is $\langle T, [Z] \rangle$, then $d(p)$ is $\langle T, [\langle T, [Z] \rangle] \rangle$. Once again, for definiteness, let $d(x) = 0$ for any x that is not a proposition. As we saw in Sect. 3, it seems hard to deny that such a function exists.

In addition to the arguments of Sect. 3, one can *also* provide set-theoretic models along the lines suggested above in which such a function exists. That is, Sects. 5.1–5.4 describe a family of set-theoretic models of propositions, in which functions, properties, $\neg$ and $\wedge$ are urelements. But it is straightforward to extend any member of this family to one in which d exists. (For reasons of space I leave this as an exercise for the reader.)

The construction of self-referential propositions can then proceed as in Sect. 2. Thus, let LP be the following.

$$(\neg, \langle T, [d, [(\neg, \langle T, [d, [Z]] \rangle)]] \rangle)$$

As before, we have:

(1*) $d((\neg, \langle T, [d, [Z]] \rangle)) = \text{LP}$.

That is, LP says of itself that it is untrue.

We then have a purely propositional version of the Liar. For if p is a proposition, then $(\neg, \langle T, [d, [p]] \rangle)$ is that to the effect that $d(p)$ is not true. So it seems that this should be true iff $d(p)$ is not. In particular:

(2*) LP is true iff $d((\neg, \langle T, [d, [Z]] \rangle))$ is not.

Giving:

(3*) LP is true iff it is not.

Similarly for the other propositions constructed in Sect. 2 (i.e. other self-referential propositions and those that form more complex networks).

Therefore, given a precise and apparently natural account of propositions, the constructions of Sect. 2 can be straightforwardly carried out. The above should also make plausible, however, that this will similarly be possible on any alternative such account (at least on which propositions are structured).

Thus, if propositions are structured, they can be self-referential, and a range of apparently plausible claims about truth will have to be rethought.

Acknowledgements For comments and discussion, I am grateful to Andrew Bacon, Agustín Rayo, Ian Rumfitt, Zoltán Szabó, an audience at the 2014 Eastern APA, and two referees for this journal.

References

Aczel, P. (1988). *Non-well-founded sets*. Stanford: CSLI Publications.

Barwise, J., & Etchemendy, J. (1987). *The liar: An essay on truth and circularity*. New York: Oxford University Press.

Beall, J. C. (2009). *Spandrels of truth*. Oxford: Clarendon Press.

Benacerraf, P. (1965). What numbers could not be. *Philosophical Review*, *74*, 47–73.

Field, H. (2008). *Saving truth from paradox*. Oxford: Oxford University Press.

Fine, K. (2007). *Semantic relationism*. Malden: Blackwell.

Gaifman, H. (2000). Pointers to propositions. In A. Chapuis & A. Gupta (Eds.), *Circularity, definition, and truth* (pp. 79–121). New Delhi: Indian Council of Philosophical Research.

Glanzberg, M. (2001). The liar in context. *Philosophical Studies*, *103*, 217–251.

Gupta, A. (1982). Truth and paradox. *Journal of Philosophical Logic*, *11*, 1–60.

Gupta, A., & Belnap, N. (1993). *The revision theory of truth*. Cambridge, MA: MIT Press.

Herzberger, H. G. (1982). Notes on naive semantics. *Journal of Philosophical Logic*, *11*, 61–102.

Jubien, M. (2001). Propositions and the objects of thought. *Philosophical Studies*, *104*, 47–62.

Kaplan, D. (1989). Demonstratives: An essay on the semantics, logic, metaphysics, and epistemology of demonstratives and other indexicals. In J. Almog, J. Perry, & H. Wettstein (Eds.), *Themes from Kaplan* (pp. 481–563). New York: Oxford University Press.

King, J. C. (2007). *The nature and structure of content*. Oxford: Oxford University Press.

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

Leitgeb, H. (2005). What truth depends on. *Journal of Philosophical Logic*, *34*, 155–192.

Maudlin, T. (2004). *Truth and paradox: Solving the riddles*. Oxford: Clarendon Press.

McGee, V. (1991). *Truth, vagueness, and paradox: An essay on the logic of truth*. Indianapolis: Hackett.

Priest, G. (1979). The logic of paradox. *Journal of Philosophical Logic*, *8*, 219–241.

Priest, G. (1987/2006). *In Contradiction* (2nd ed.). Oxford: Clarendon Press.

Salmon, N. (1986). *Frege's puzzle*. Cambridge, MA: MIT Press.

Skyrms, B. (1984). Intensional aspects of self-reference. In R. L. Martin (Ed.), *Recent essays on truth and the liar paradox* (pp. 119–131). Oxford: Clarendon Press.

Soames, S. (1987). Direct reference, propositional attitudes, and semantic content. *Philosophical Topics*, *15*, 47–87.

Soames, S. (1999). *Understanding truth*. New York: Oxford University Press.

Soames, S. (2010). *What is meaning?* Princeton: Princeton University Press.

Sobel, J. H. (1992). Lies, lies, and more lies: A plea for propositions. *Philosophical Studies*, *67*, 51–69.

Tarski, A. (1935). The concept of truth in formalized languages In his J. Corcoran (Ed.) *Logic, semantics, metamathematics: Papers from 1923 to 1938*, 1983. 2nd edn. (pp. 152–267) (J. H. Woodger trans.). Indianapolis: Hackett.