Draft of June 2020—please do not cite without permission.

The Premise Paradox

T. Parent (Nazarbayev University)
nontology@gmail.com

1. Introduction

This paper argues that incautious talk of "premises" can lead to contradiction. This is so, even though a "premise" is here defined by a purely formal or syntactic feature: It is a sentence that is underived rather than derived in the context of a proof. More broadly, the contradiction can occur even if our object language is semantically open in Tarksi's (1944) sense[1].

A preliminary matter: In classical symbolic logic, expressions must be individuated as linguistic *types*. (To some, this will seem like a matter of course, but I worry that a quick acceptance of it now may lead to a hasty rejection of it later.) A key reason is that it is quite easy to generate paradox if an "expression" individuated as a single, non-repeatable occurrence.[2] Most basically, if the expression is a one-time affair, then the Law of Identity will have false instances:

- this very demonstrative phrase = this very demonstrative phrase
- 'dogs' = 'dogs'
- 'Socrates is mortal' = 'Socrates is mortal'

---

[1] Tarski tells us that a language L is "semantically open" iff: L does not contain its own semantic terms, or more precisely, L does not contain semantic terms that are defined on expressions of L.

[2] In lieu of "tokens," I ultimately prefer talk of "occurrences:" If tokens are concrete particulars, then occurrences are the abstract parts of a sentence-type or proof-type. (Here I follow Wetzel 2008.) Yet I speak of "tokens" since that is more familiar, and is often less awkward stylistically. But my remarks would be more precise if the talk of occurrences were used instead.

Such identity-statements would be false—for the terms flanking '=' in each case would denote different expression-tokens. (In the latter two examples, this holds assuming that a quotation denotes the "expression" inside the quotes.)

One response (raised by Bill Lycan, in correspondence) is that this shows merely that the expressions flanking '=' in each case would be equivocal. So these examples would not violate the Law of Identity, since the Law requires univocal terms on either side of '='. Semantically speaking, this is quite correct. However, I am here referring to the Law as a *formal* rule, i.e. a generalization whose instances are defined by their syntactic form. Thus construed, the Law says that whenever '=' is flanked by tokens of the same *formal* type, the resulting sentence is true. So the language must avoid any form whose tokens shift denotation. But if expressions are individuated as tokens, this requirement is violated. E.g., if a "demonstrative-phrase" is defined as a token, then different tokens of 'this very demonstrative-phrase' will be of the same formal type, yet fail to co-refer. Further, it is no help to use subscripts to distinguish 'this very demonstrative-phrase$_1$' from 'this very demonstrative-phrase$_2$'. Granted, one could not derive as a theorem 'this very demonstrative-phrase$_1$ = this very demonstrative-phrase$_2$'. Yet one could still derive 'this very demonstrative-phrase$_1$ = this very demonstrative-phrase$_1$'. Here too, if a "demonstrative-phrase" is numerically one token, it is a false instance of the Law.

So again, at least in a *formal* system for classical logic, expressions cannot be seen as individual tokens; they must instead be seen as types. [3]

[3] A footnotewothy consequence is that 'token' cannot be a term of a classical language. For it is false that this very token = this very token. Other problems also arise if expressions are referred to as tokens. Suppose line 1 of a proof is 'This very sentence-token starts this proof'. Then, since the same sentence-type can be derived later in the proof, yet is false in that case, the system is unsound. (Note that even the identity-predicate is unneeded for this latter

2. The paradox

The paradox I wish to present consists in a proof of a falsity from apparently true premises. Here, let 'MP' name modus ponens (seen as a purely syntactic transformation rule), and let "=Elim" be the formal rule corresponding to the indiscernability of identicals, i.e., $\forall x \forall y(x = y \supset [Fx \equiv Fy])$:

(1)	Socrates is mortal	[Premise]
(2)	(5) is not a premise in this proof	[Premise]
(3)	(1) = (5)	[Premise]
(4)	Socrates is mortal ⊃ Socrates is mortal	[Trivial]
(5)	Socrates is mortal	[(4), (1), MP]
(6)	(1) is not a premise in this proof	[(2), (3), =Elim]

Above, the premises seem true while (6) is false. So the derivational system looks unsound.[4]

In short, (1) and (5) have contrasting roles in the proof, and that leads to contradiction if (1) = (5). But here is an objection: If we take seriously the type-identity between (1) and (5), then since (1) is a premise in the proof, we might insist that (5) is a premise too. In which case, (2) is false. The problem, however, is that (5) is derived by *modus ponens*—yet a premise by our definition is underived in a proof.

In the end, this should prompt us to identify (1) and (5) not as numerically one type, but rather as distinct tokens of the type. Then, we can say that (1) is a premise in the proof and (5) is not. (Premise (3) on this interpretation is therefore false.) But although this is the correct thing to

---

problem.) Now as before, this would owe to an equivocation on 'This very sentence-token'. But again, as long as the two sentence-tokens are of the same formal type, the formal system will allow the equivocation.

[4] A shorter version of the proof exists if the reiteration rule is allowed. But I have encountered "purists" who look down on the reiteration rule, and the paradox by reiteration may seem just to confirm their scruples. I thus prefer the modus ponens paradox above.

say, it is not the sort of thing one can say in a classical formal system. Again, if (1) is a non-repeatable token, it is then false that ‘Socrates is mortal’ = ‘Socrates is mortal’, contra the (formal) Law of Identity.

So apparently, classical symbolic logic must put constraints on talk of “premises.” Without such constrains, the problem can be summed up as follows:

- If ‘Socrates is mortal’ is a token, then it falsifies the formal Law of Identity.
- If ‘Socrates is mortal’ is a type, then it is both a premise and not a premise in the above proof.

In one way or another, then, ‘Socrates is mortal’ could be used to show that such a derivational system unsound.[5]

3. A proof-theoretic objection

A different objection has been independently raised (in conversation) by Graham Priest and Peter Woodruff. The basic thought is that we should not speak of a sentence-type as a premise, but rather say this of a sentence-type *relative to a line*—or what amounts to the same, *relative to its position in the sequence* of sentences comprising the proof. In the background is the idea that a proof is an ordered *n*-tuple of sentence-types, and a single type can have more than one position in the sequence—just in the way that the number 1 has more than one position in the Fibbonaci sequence $\langle 1, 1, 2, 3, 5, 8, 13\ldots\rangle$. If so, then in the proof above, we should first

[5] It may be wondered how this is possible, given the soundness proof for classical logic. Well, as is familiar from Gödel’s second incompleteness theorem, the existence of a soundness proof is not sufficient for soundness. Rather, one needs a soundness proof that starts with true premises. Yet the soundness proof for classical logic assumes that one cannot derive a falsity from true literals. So if our language contains literals like (2), then the assumptions of the soundness proof will not be correct.

say that the type ‘Socrates is mortal’ is assigned to the first index of the proof-sequence (for short, the sentence is “at index-1”), where it has the role of a premise. We can then add that the same sentence-type is also at index-5, where it does not have the role of a premise. In this manner, we may avoid contradiction while avoiding talk of sentence-tokens.

There is more than one way to construe this suggestion, and on one construal, it indeed seems promising. (See the next section.) However, on another construal, the proposal only changes the formulation and not the substance of the problem. Consider that the following proof still appears unsound:

(1)	Socrates is mortal	[Premise]
(2)	The sentence at index-5 is not a premise in this proof	[Premise]
(3)	The sentence at index-5 = the sentence at index-1	[Premise]
(4)	Socrates is mortal $\supset$ Socrates is mortal	[Trivial]
(5)	Socrates is mortal	[(4), (1), MP]
(6)	The sentence at index-1 is not a premise in this proof	[(2), (3), =Elim]

In this latest proof, (6) is false. Yet on the present construal, the objector would be committed to the truth of (2), along with (1) and (3). The objection thus fails, for ‘the sentence at index-1’ and ‘the sentence at index-5’ still refer to numerically the same sentence-type. So if the sentence at index-5 is not a premise, then so too with the sentence at index-1.

This is not sleight of hand. We can make the paradox vivid by asking: Is the sentence-type at index-5 a premise in the proof? The present line is committed to answering “no.” But since the sentence at index-1 = the sentence at index-5, it then must be equally true that the sentence at index-1 is a not premise in the proof. Yet the sentence at index-1 indeed has the role of a premise.

And just to be clear: If, instead, it is claimed that the sentence at index-5 is a premise, then by our definition of "premise," the sentence at index-5 is underived in the proof. But of course, the sentence at index-5 is derived by MP.

4. Closing remarks

How should the paradox be resolved? The promising suggestion would be to relativize to a line the very idea of a "premise." So we would never talk of 'Socrates as mortal' as a premise *tout court.*[6] We would instead speak of it as a "premise-at-(1)," while simultaneously saying it is not a "premise-at-(5)." But refusing (1) the status of a "premise" *tout court* is a revisionary proposal, and seems to have no precedent in standard discussions of proof calculi (natural deduction, truth-trees/tableaux, etc.). To say the proposal is revisionary is not to say that it is *revolutionary.* Its implications may end up being quite minimal. Be that as it may, I am not prepared to defend it here.[7]

It is not unprecedented, of course, to recognize that a premise may occur later in a proof. But when it does, the custom has been to say that it is still a premise (without qualification). Yet

---

[6] Or rather, we would not say it is a premise *tout court* unless the notion of being "derived in a proof" is also relativized to a line. Conceivably, one could say that the sentence-type 'Socrates is mortal' in the previous proof is a premise *tout court* since it is underived-at-some-line, viz., at line (1). (And its being derived-at-some-line as well, viz., at line (5), would not detract from that.) But hiding the line-relativity under the label "premise *tout court*" may well obscure rather than clarify the status of 'Socrates is mortal' in the proof.

[7] Tim Button suggests (in conversation) a different solution for the paradox. We might restrict the predicate 'x is a premise in y' so that values for 'y' do not include the very proof in which the predicate is used. But this is a constraint on self-reference—the predicate cannot be composed (at the y-position) with a term for its own containing proof. Such a constraint would also require further elaboration and defense before it could be affirmed.

I suspect that custom assumes here that if a premise *occurs* later in the proof, then one occurrence of the sentence is underived, while another occurrence is derived. And this assumption may well be correct. But as we saw, a classical formal system must always individuate sentences as types, never as tokens or occurrences.

My main purpose here was negative only; it was just to show that under present assumptions, incautious talk of "premises" leads to contradiction.[8]

References

Tarski, A. 1944. The Semantic Conception of Truth and the Foundations of Semantics. *Philosophy and Phenomenological Research* 4(3): 341–76.

Wetzel, L. 2008. *Types and Tokens: An Essay on Universals*. Cambridge, MA: MIT Press.

[8] Thanks to Axel Barceló, Tim Button, Cory Johnson, Gary Kemp, Dan Linford, William G. Lycan, Michaelis Michael, Yves Peraire, Martin, Pleitz, Graham Priest, Roger Wertheimer, Peter Woodruff, and Min Xu, for feedback on earlier drafts.