A SIMPLE SOLUTION TO MORTENSEN AND PRIEST'S TRUTH TELLER PARADOX

Joseph WAYNE SMITH

SYNOPSIS

Mortensen and Priest in their paper "The Truth Teller Paradox", have maintained that the sentence:

(a) This very sentence is true

is neither true nor false. This being so, then clearly (a) is not true. But since it says of itself that it is true, it is false, thus contradicting the supposition that it is neither true nor false. Hence it is either true or false. After rejecting a number of proposed solutions, Mortensen and Priest propose that we should view disjunction in this situation in a non-standard (what they call "non-prime") fashion. While (a) fails to be true and fails to be false, nevertheless it is either true or false. They support this claim by plausibility considerations based upon Gödel's first Incompleteness Theorem.

In this paper I will reject the solution of Mortensen and Priest, arguing that their argument based upon Gödel's first Incompleteness Theorem fails to render their proposal of primeness plausible. Instead I suggest a solution which is I believe intuitively plausible and non-problematic: (a) is simply true.

Consider the following sentences:

(a) This very sentence is true;
(β) This very sentence is false.

The second of these, (β) is the well-known liar paradox. The first of these, (a) is equally paradoxical, or so Mortensen and Priest claim. (1) There is, Mortensen and Priest claim, nothing to choose between the hypotheses that it is true, and that it is false, and both hypotheses seem consistent unlike (β). For (β) there are at least *prima facie* proofs of the mutual truth and falsity of (β).

(1) C. MORTENSEN and G. PRIEST, "The Truth Teller Paradox", *Logique et Analyse*, vols. 95-96, 1981, pp. 381-388.

Nevertheless (a) has specific problems of its own. Suppose that (a) is neither true nor false. Then clearly (a) is not true. But since it says of itself that it is true, it is false, thus contradicting the supposition that it is neither true nor false. Hence it is either true or false. Thus there is a proof that (a) is *either* true or false. Yet there are apparently no proofs of its truth *or* its falsity. The proof that (a) is either true or false may be stated more formally letting «(a)» be an abbreviation for the sentence (a) and not the name of the sentence, and letting quasi-quotes be used as name forming functors:

Let $\ulcorner a \urcorner = \ulcorner T \ulcorner a \urcorner \urcorner$ (1)

Now $(\sim T \ulcorner a \urcorner \;\&\; \sim F \ulcorner a \urcorner) \rightarrow \sim T \ulcorner a \urcorner$ (2)

But $\ulcorner a \urcorner = \ulcorner T \ulcorner a \urcorner \urcorner \rightarrow (\sim T \ulcorner a \urcorner \rightarrow F \ulcorner a \urcorner)$ (3)

Then $(\sim T \ulcorner a \urcorner \;\&\; \sim F \ulcorner a \urcorner) \rightarrow F \ulcorner a \urcorner$ (4)

(4), from (1), (2), (3)

Then $(\sim T \ulcorner a \urcorner \;\&\; \sim F \ulcorner a \urcorner) \rightarrow (T \ulcorner a \urcorner \vee F \ulcorner a \urcorner)$ (5)

Since $T \ulcorner a \urcorner \vee F \ulcorner a \urcorner \vee (\sim T \ulcorner a \urcorner \;\&\; \sim F \ulcorner a \urcorner)$ (6)

Then $T \ulcorner a \urcorner \vee F \ulcorner a \urcorner$ (7)

(7), from (5) and (6)

Mortensen and Priest then consider the possibility of faulting one or more lines of the above argument. Since all of their considered solutions are inadequate, I turn immediately to Mortensen and Priest's own proposal. Their solution is that while (a) fails to be true and fails to be false, nevertheless it is either true or false. This involves taking a non-standard interpretation of disjunction. The intensional disjunction studied by Anderson and Belnap is such that if p is true and q is false, then pvq is not true, i.e. the truth of one of the disjuncts is not sufficient for the truth of the disjunction. ([2]) The disjunction of Mortensen and Priest is such that pvq holds even though p fails and q fails. A theory is said to be prime if and only if, whenever AVB is in the theory either A is in the theory or B is, and non-prime if this condition is not satisfied. The truth-teller paradox Mortensen and Priest maintain, is simply a case of the failure of the condition of primeness to hold.

The failure of primeness is illustrated by the following example: let

([2]) A. R. ANDERSON and N. BELNAP, *Entailment*, (Princeton University Press, Princeton, 1975), pp. 176-177.

PA be Peano arithmetic formulated with a base of classical logic, and let G be its Gödel sentence. Now it is the case that:

(1) $\vdash_{\overline{PA}} (G \vee \sim G)$

but by Gödel's first Incompleteness Theorem:

(2) $(\sim \vdash_{\overline{PA}} G \,\&\, \sim \vdash_{\overline{PA}} \sim G)$

That is to say, the set of theorems of Peano arithmetic is non-prime if it is consistent. Whilst this is true, this example does not give Mortensen and Priest the justification which they need for their non-standard disjunction. Many have taken G to be true by informal reasoning even though $\sim \vdash_{\overline{PA}} G$, that is to say, mathematical truth is not simply a matter of formal proof. Indeed this must be so many have said, under pains of either vicious infinite regression or circularity. But if this is so, then the Gödel example simply does not support Mortensen and Priest's proposals.

The solution which I propose to this problem is simply that sentence (a) is true. We have a proof that (a) must be either true or false, so if (a) is to have a truth-value at all, our range of options are restricted to either the truth-value 'true' or the truth-value 'false'. There seems to be more reason to take (a) to be true, rather than false, even if neither ascription *prima facie* leads to contradiction: (a) asserts of itself that it is true, not that it is false. If one is prepared to take the self-referential ascriptions of falsity of a sentence such as the liar sentence (β) at all seriously, it is just a *petitio principii* not to take the self-referential ascription of truth of sentence (a) seriously.

Mortensen and Priest both maintain that the supposition that (a) is false does not raise any problems with regard to consistency at all. This claim can be questioned. Let us assume here the principles '$T(p) \equiv p$' and '$F(p) \equiv p$'. Second, the truth-teller sentence can be represented by this sentence:

(1) $T(p)$.

The claim that (a) is false is the sentence:

(2) $F(T(p))$.

Further to claim that (a) is false entails that (a) is not true:

(3) $F(T(p)) \rightarrow \sim T(p)$.

But from (2) and (3) by *modus ponens* we infer:

(4) $\sim T(p)$

which contradicts (1). Sentence (1) cannot be simply eliminated from this proof, since it is a correct statement of the truth-teller sentence (a) and its occurrence in our argument is thus not an arbitrary one. Hence assuming that sentence (a) is false, despite one's first intuitions, does lead to contradiction. So what other truth value have we left to assign to (a) but the value of 'true'? Note that in doing this, we *do not* obtain a corresponding contradiction, i.e.

(1′) $T(p)$ (assumption)
(2′) $T(T(p)) \rightarrow \sim F(p)$ (classical entailment)
(3′) $T(T(p))$
(4′) $\sim F(p)$ ((2′), (3′) (modus ponens)
(5′) $\sim F(p)) \equiv \sim\sim(T(p)$ (classical equivalence)
(6′) $T(p)$

Second, all other solutions reviewed by Mortensen and Priest are inadequate. Whilst Mortensen and Priest's proposal is interesting and stands open to refinement, it lacks the simplicity of my own solution. Why see tragedy and paradox where none need be seen at all? The simple solution to the Truth Teller Paradox is that (a) is true. (3)

Department of Philosophy
The University of Adelaide

Joseph WAYNE SMITH

(3) I am grateful to Eddy Hughes and Chris Mortensen for a discussion of this proposed solution.