Reply to Slater

Graham Priest

In his 'Dialetheias are Mental Confusions' (200a), Slater argues that dialetheism may be dismissed very swiftly. People's beliefs can be inconsistent; the truth cannot. The negation symbol of a paraconsistent logic does express real negation. For such negation, A and $\neg A$ cannot be true together—by definition. End of story. Slater also essays a consistent account of the semantic paradoxes. In this note I will explain why I am unpersuaded about both matters.

1 Contradictory or Sub-Contrary?

Slater and I agree that negation, whatever it is, is a contradictory-forming operator (cfo). It is the relation that obtains between pairs such as 'Socrates is mortal', 'Socrates is not mortal' and 'Some person is mortal', 'No person is mortal'. The crucial question, then, is what exactly, this relationship amounts to.

Traditional logic—by which I mean logic in the Aristotelian tradition—characterises the relation in a familiar way. A and B are contradictories if you must have one or the other, but you can't have both. That is, $\Box(A \vee B)$ and $\neg\Diamond(A \wedge B)$. Hence, A and $\neg A$ are contradictories if we have $\Box(A \vee \neg A)$ and $\neg\Diamond(A \wedge \neg A)$, that is, $\Box\neg(A \wedge \neg A)$. Consider any propositional logic, the modal extension of which satisfies Necessitation (if $\vdash A$ then $\vdash \Box A$), as it should. Then if it is such that both:

1. $A \vee \neg A$
2. $\neg(A \wedge \neg A)$

are logical truths, $\neg$ is a cfo. Since LP satisfies these conditions, its negation symbol is a cfo. Note that this is not true of either intuitionistic logic or some paraconsistent logics, such as the da Costa C-systems. In the first of these, 1 is not a logical truth; in the second, 2 is not. This is why the charge that the negation operator in those logics is not really a cfo gets its bite.

Naturally, in a paraconsistent context 1 and 2 do not stop $A \wedge \neg A$ holding as well. But this does not show that $\neg$ is not a contradictory-forming operator. It just shows that there is *more* to it than one might have thought. Let us call this more, for want of a better phrase, its *surplus content*.[1]

In the paper where Slater first levelled his charge that the negation symbol of LP is not a cfo, he observed, correctly, that in the semantics for LP there can be interpretations, ρ, and sentences, A, such that $A\rho 1$ and $\neg A\,\rho 1$, i.e., $A\rho 0$. That is, sentences are both true and false in some interpretations.[2] The relevance of this observation is, however, less than transparent. The theory of interpretations is a twentieth (or late nineteenth) century construction aimed at giving an account of validity—an account that has somewhat tenuous links to Aristotle's own notion of syllogistic validity. Now, in the model theory of classical logic—by which I mean logical theory in the tradition of Frege and Russell—A cannot be both true-in-an-interpretation and false-in-an-interpretation. This delivers the logical truth of both 1 and 2, and so the fact that negation is a cfo. The semantics of LP do likewise, but also make room for negation to have surplus content.[3] They also show why the inference of Explosion:

$A \wedge \neg A \vdash B$

is invalid—which it is, incidentally, in syllogistic. (Contradictory premises do not suffice to make a syllogistic inference valid.)[4]

[1] And if it be retored that $\neg$ cannot have surplus content, the reply is 'Of course: $\neg\Diamond(A \wedge \neg A)$'!

[2] Slater (1995). I take semantic evaluations to be relations between formulas and truth values. Slater formulates LP as a three-valued logic, and takes evaluations, V, to be functions from formulas to the values $\{1, 0, -1\}$. Hence, what I write as $A\rho 1$, he writes as $V(A) \geq 0$. Nothing turns on this, in this context, I think.

[3] In section 3 of his paper, Slater points out that one may interpret the semantics of LP in intentional terms. This, however, shows nothing. One can interpret the semantics of classical logic in terms of switching-circuits. That does not show that classical logic is about electronics. All semantics are subject to multiple interpretations. And in the intended interpretation of LP semantics, $\neg$ is a purely extensional operator which simply operates on the truth(-in-an-interpretation) values of its inputs.

[4] For a discussion, see Priest (200b), 2.1.

2 Definition

One can, of course, contest the view that negation has a surplus content. For that matter, one can contest the view that a cfo really satisfies conditions 1 and 2, as do intuitionists.[5] The relation of being contradictories is one of a whole bunch of notions that play a central role in logic. Others include implication, modal status, generality. These are all things that are contentious—indeed, in the history of Western philosophy, logicians have frequently contended about them.[6] In logic, just as much as physics, people put forward theories of how the relevant notions behave (and we have to judge the theories by the usual criteria of theoretical adequacy).[7] The inferential behaviour of a cfo cannot, therefore, be settled by definition, as Slater thinks it can.

But maybe one can define some operator, let us call this $, which is a cfo and has no surplus content. If one can, then, in some ways, the behaviour of ¬ is beside the point: a classical logician can simply concede it to the paraconsistent logician, and make their point in terms of $. If we interpret Slater's remarks about definition in this way, they have a point. The crucial question then becomes: how is $ to be defined? Slater does not say exactly what sort of definition he has in mind, nor what, exactly, he takes the definition to be. Clearly, an explicit definition, of the form '*dialetheia* means *true contradiction*', is not going to get us very far. Such definitions are eliminable without loss—or if they are not, they are creative, and so objectionable. We must appeal therefore to some notion of implicit definition.

There are two plausible strategies here.[8] The first is proof-theoretic. One simply takes $ to be a connective that satisfies all the rules of inference governing negation in classical logic. In this case, we will have Explosion, which effectively rules out surplus content—on pain of any surplus turning into the total content of everything.[9]

[5]Slater claims (p. 2) that I do not think that intuitionist logic is a rival to classical logic. It certainly is. It disagrees with classical logic as to how a cfo should behave—as well as many other things.

[6]And unless one is a nominalist, these disputes are not simply about the way that words are used.

[7]The theories, of course, have to answer to certain data. Thus, theories of consequence have to answer to the particular inferences that strike us as valid or invalid. But as in all theoretical enterprises, the data is defeasible.

[8]These are discussed further in Priest (1990).

[9]Note that simply having a connective satisfying Explosion is not sufficient to damage

The problem with this strategy is that there is no guarantee that this specification determines a notion with any sense. As Prior pointed out,[10] an arbitrary set of rules may well not succeed in capturing any meaningful notion. Prior's example was a connective, $*$ (tonk), taken to be governed by the rules $A \vdash A * B$, $A * B \vdash B$. Clearly, if tonk were a legitimate notion, we could prove everything. But if \$ were a legitimate notion we could, similarly, prove everything, given only that we have the T-schema and some way of forming self-referential truth-bearers. (We simply formulate a liar sentence, L, of the form $\$T\langle L\rangle$, and establish $L \wedge \$L$ in the usual way.)[11]

The other strategy is to characterise \$ model-theoretically. The natural thought here is to specify a connective whose truth-in-an-interpretation conditions are:

0. $\$A\rho 1$ iff it is not the case that $A\rho 1$

(and, if you like: $\$A\rho 0$ iff $A\rho 1$). One might contest the claim that these truth conditions determine a meaningful connective. But let us grant that they do.[12]

Given that conjunction and disjunction behave in the usual fashion, it is straightforward to establish that, for all ρ, $(A \vee \$A)\rho 1$, and:

1. for no ρ, $(A \wedge \$A)\rho 1$

so, for all ρ, $\$(A \vee \$A)\rho 1$. Hence, given that modal operators behave in a natural way, we can establish that \$ is a cfo. But does \$ satisfy Explosion? According to the model-theoretic account of validity, $A \wedge \neg A \models B$ iff:

2. for all ρ, **if** $(A \wedge \neg A)\rho 1$, $B\rho 1$

Keep your eye on the boldfaced **if**. To establish 2, we have to infer it from 1. The inference is a quantified version of the inference:

3. it is not the case that C; so **if** C, D

dialetheism. In a logic appropriate for dialetheism one may have a logical constant, $\bot$, such that $\bot \vdash B$, for all B. (See Priest (1987), 8.5.) If one defines $-A$ as $A \to \bot$, then one has $A, -A \vdash B$.

[10] Prior (1960).

[11] For other arguments to the effect that an explosive \$ is, indeed, meaningless, see Priest (200a), ch. 5.

[12] There are at least semantics for relevant logics where a connective is given such truth conditions. See Meyer and Routley (1973).

If **if** satisfies *modus ponens*, if it is, say, the conditional of a relevant logic, then the inference 3 is not valid. If, on the other hand, **if**, does not satisfy *modus ponens*, say it is the material conditional, then 3 may well be valid. But now the validity of Explosion does not rule $ out from having surplus content. One cannot get from $(A \wedge \$A)\rho 1$ and $A \wedge \$A \models B$ to $B\rho 1$, since this would use an instance of the Disjunctive Syllogism, invalid in all paraconsistent logics:

$(A \wedge \$A)\rho 1$ and (it is not the case that $(A \wedge \$A)\rho 1$ or $B\rho 1$); so $B\rho 1$

Either way, then, $ fails to perform as required. The dialetheist can accept the implicit definition involved in 0. They just have no reason to suppose that the connective, so defined, rules out surplus content.

All this assumes that the 'not' in 0 behaves as paraconsistent logic says that it does. It might be thought that we would fare better if we take it to be $ itself. But in fact, in that case, we get nowhere. Given that the truth conditions for $ themselves employ $, we have to know what inferences govern $ before we can infer anything about how sentences containing $ behave. This cannot be taken for granted: the whole point was precisely to justify various inferences concerning $. It might be suggested that we can simply *assume* that $ is a connective that satisfies the principles of classical negation; but the problem with this is obvious enough. We were supposed to be in the process of showing that $, as characterised by its truth conditions, satisfies Explosion (and so rules out surplus content). So such an assumption is clearly question-begging.

Though the route has been a long one, the point is simple. The phrasing of an implicit definition, on its own, does not get us very far. One needs to show that this succeeds in characterising a notion. One then has to show that the notion characterised has the properties that one claims it does. To establish this, one needs to make inferences from the defining conditions. What we have just seen is that one can make the case that $, characterised semantically, rules out surplus content only by begging the question in one way or another.[13]

[13] Before we leave the notion of negation, a quick comment on two other points. Slater claims (p. 6) that dialetheists have a problem expressing denial, and that one can have inconsistent beliefs only by being muddled. Both claims are false. Denial is a speach act that is not the same as the assertion of a negation; and the "preface paradox" shows that rational people can have inconsistent beliefs. On both of these matters see, e.g., Priest (1993), and (200a), ch. 6.

3 Self-Reference

Of course, none of these considerations show that negation does have surplus content. That is the onus of arguments for dialetheism—which brings us to one of these, the paradoxes of self-reference. Slater finds flaws in the arguments involved in the liar paradox and the heterological paradox.

For the liar paradox, he claims (p. 4) that truth is not a property of sentences, but of propositions, and that once one sees this, it is impossible to come up with the sort of self-referential proposition required for the liar. Now, for myself, I see nothing wrong with taking sentences to be truth bearers, provided that we are talking about *interpreted* sentences here (and not simply grammatical strings), and that the sentences do not contain indexical phrases, such as 'I' and 'now'. But let us grant that it is propositions which are to be regarded, strictly speaking, as truth bearers. The claim that one cannot come up with the appropriate self-referential propositions is false. One can do this with appropriate demonstratives, as in 'this is false', where 'this' refers to the proposition the sentence expresses.

Slater claims (p. 4) that there cannot be such propositions, since they would have to be part of themselves. But propositions are abstract objects, and it is not at all clear that they cannot contain themselves as parts. Indeed, using representations of propositions employing non-well-founded set theory, one can show that there are self-referential propositions of exactly the kind employed in the liar paradox—as Barwise and Etchemendy have demonstrated.[14]

Another way to obtain an appropriate self-referential sentence is as follows. Those who take sentences to express propositions make use of those very words. But now we can employ these words to formulate the sentence: the proposition expressed by this sentence is false. Reasoning as usual leads to contradiction.[15] Putting matters this way invites the objection that the description 'the proposition expressed by this sentence' fails to refer. This may be because the sentence expresses no proposition or because it expresses more than one proposition. Of course, merely to moot this possibility is not to solve the paradox (more of this in a moment): one needs to give reasons to suppose the claim to be true. Using the Barwise-Etchemendy account of propositions it is demonstrably false.

[14]Barwise and Etchemendy (1987), esp. chs. 3, 4.

[15]See Asher andr Kamp (1986).

In any case, the move is of no avail, since we can formulate an extended version of the paradox:

(*) either this sentence expresses no unique proposition or it expresses a false one.

If it expresses a unique proposition, we have the usual contradiction. And now if we claim that (*) does not express a unique proposition, since, presumably, we must be taking ourselves to be expressing a true proposition in saying this, (*) would appear to express a true disjunctive proposition.

Slater's account of the heterological paradox is somewhat different. He diagnoses a suppressed premise, concerning ambiguity, in the argument. The argument can then be taken as a *reductio* of that premise. Actually, this is a quite general strategy for attempting to solve the self-referential paradoxes. Take some premise, A, in the argument involved, claim that this is true only on condition that B, and then deny B. We have, in fact, just looked at a version of this strategy in connection with the liar paradox (where B was a claim to the effect that a certain sentence does not express a unique proposition). But just because this strategy is available to virtually *any* supposed solution to the paradoxes, it is worth very little unless there is independent reason for supposing that B fails—and even then, as we have already seen, this strategy may still fail, due to extended paradoxes.

In the case of the Heterological paradox, the suppressed premise that Slater claims to find is that the heterological predicate, H—that is, in the notation he uses, $\exists Y(xRY \wedge \neg Yx)$—is univocal. Now the predicate H certainly does not appear to be ambiguous, as Slater seems to concede (p. 8). He offers an argument to the effect that it is; but I must confess that I am at a loss to make anything sensible of it. The thought appears to be that the denotation of 'self' is context-dependent. Hence, the meaning of 'self-applicable' depends on the context. Whether or not 'self' is generally context-dependent, in 'x is self-applicable', 'x is self-pitying', 'x is self-aggrandizing', etc., the 'self' is just an anaphoric back reference to x. No context is needed to determine this.

Worse, even if H is ambiguous, we can just disambiguate in an appropriate fashion and run the argument for the relevant sense. Note that if H is ambiguous, this is presumably due to an ambiguity in the predicate R ('refers to'). But if this is ambiguous, so is Slater's claim that H is ambiguous: $\exists X \exists Y(\langle H \rangle\, RX) \wedge \langle H \rangle\, RY \wedge X \neq Y$. He, I assume, had a particular sense of 'refer' in mind.

Finally, if one is prepared to talk of properties, one can formulate the Heterological paradox simply in terms of them. H, now, is the property of not applying to itself, $\lambda P(\neg PP)$. We then have:

$$HH \leftrightarrow \lambda P(\neg PP)H \leftrightarrow \neg HH$$

This invites a reply (which Slater, in effect, makes) to the effect that the lambda term fails to pick out a unique property. Let us call such a term 'undefined', and write Ux for 'x is undefined'.

We now have to face the issue of the truth conditions of sentences of the form $\lambda P(A)Q$ when $\lambda P(A)$ is undefined. There would seem to be (at least) three options here, in accord with standard policies of defective reference:

$\lambda P(A)Q$ is false

$\lambda P(A)Q$ is true if *some* denotation of '$\lambda P(A)$' applies to some (or every) denotation of 'Q' (and false otherwise)

$\lambda P(A)Q$ is true if *every* denotation of '$\lambda P(A)$' applies to some (or every) denotation of 'Q' (and false otherwise)

Each would seem to make perfectly good sense, giving rise to a different notion of predication. We will operate in terms of the first (in line with Russell's theory of definite descriptions).

Having got this straight, we can now formulate an extended version of the paradox. By standard fixed-point constructions, we can generate a predicate, H^*, of the form $\lambda P(\neg PP \vee U\langle H^*\rangle)$ (the property of not applying to itself or of this specification being defined). We then have:

$$H^*H^* \leftrightarrow \lambda P(\neg PP \vee U\langle H^*\rangle)H^* \leftrightarrow (\neg H^*H^* \vee U\langle H^*\rangle)$$

If H^* is defined, we have the usual contradiction. If not, then the right hand side is true. So, then, is H^*H^*. So H^* is defined. We are back with contradiction.

References

[1] N. Asher and J. Kamp (1986), 'The Knower's Paradox and Representational Theories of Attitudes', pp. 131-47 of J. Y. Halpern (ed.), *Theoretical Aspects of Reasoning about Knowledge (Proceedings of the 1986 Conference)*, Los Altos, CA: Morgan Kaufmann.

[2] J. Barwise and J. Etchemendy (1987), *The Liar: an Essay on Truth and Circularity*, Oxford: Oxford University Press.

[3] R. K. Meyer and R. Routley (1973), 'Classical Relevant Logics, I', *Studia Logica* 32, 51-68.

[4] G. Priest (1987), *In Contradiction*, Dordrecht: Kluwer Academic Publishers; second edition, Oxford: Oxford University Press, forthcoming.

[5] G. Priest (1990), 'Boolean Negation and All That', *Journal of Philosophical Logic* 19, 201-15; a revised version appears as ch. 5 of Priest (200a).

[6] G. Priest (1993), 'Can Contradictions be True? II', *Proceedings of the Aristotelian Society, Supplementary Volume* 67, 35-54.

[7] G. Priest (200a), *Doubt Truth to be a Liar*, Oxford: Oxford University Press, forthcoming.

[8] G. Priest (200b), 'Paraconsistency and Dialetheism', Ch. 3, Vol. 8 of D. Gabbay and J. Woods (eds.), *Handbook of the History of Logic*, Amsterdam: Elsevier, to appear.

[9] A. Prior (1960), 'The Runabout Inference-Ticket', *Analysis* 21, 38-9; reprinted as pp. 129-31 of P. Strawson (ed.), *Philosophical Logic*, Oxford: Oxford University Press, 1967.

[10] B. H. Slater (1995), 'Paraconsistent Logic?', *Journal of Philosophical Logic* 24, 451-54.

[11] B. H. Slater (200a), 'Dialetheias are Mental Confusions', this volume.