A Dilemma for Dialetheism

Jason Zarri

[Note: This is a revised version of an article originally published in the Spring 2010 edition of the Stanford undergraduate philosophy journal The Dualist (vol. 15).

Also: This article deals with issues in formal logic and contains a fair amount of logical notation. If you don't have any experience with formal logic you might not get much out of it.]

0. Introduction

Dialetheism is the belief that some sentences are both true and false—or, depending on how one defines negation, that some sentences are both true and not true. As I will use the term 'dialetheism', dialetheism is opposed to *trivialism*, the belief that *every* sentence is both true and false. It is thus essential to the dialetheist position that a sentence may be true or false without being both true and false. Consequently, there must be a difference between being exclusively true, or exclusively false, and being both true and false. Given this, it is natural to want to have a way to express such claims. In spite of that, we shall see below that expressing them while avoiding trivialism is none too easy a thing to do.

Dialetheism is usually motivated by considerations involving logical and semantic paradoxes, most famously by the liar paradox. One of the most basic versions of the liar is (L):

(L): This sentence is false.

Is (L) true? If it is, then what it says is the case, and since what (L) says is that it's false, it's false. And if (L) is false then, since what (L) says is that (L) is false, (L) is as it says it is, and hence (L) is true. So if (L) is true, it's false, and if it's false, it's true. It appears that either way we have a contradiction. And as dialetheists see it, things are just as they appear: (L) is both true and false.

One problem that besets accounts of the liar paradox is that of "revenge liars". A revenge liar is a liar sentence which, while being of the same kind as the liar sentences an account attempts to handle, cannot be handled in the same way as they are without generating contradictions that the account in question is powerless to resolve. For instance, one might try to avoid the conclusion that (L) is both true and false by distinguishing between *being false* and *not being true*, and then saying that (L) is both *not true* and *not false*. One would then face the following revenge liar:

(L*): This sentence is not true.

(L*) would seem to be true if and only if it is not true, and so distinguishing between being false and not being true will not enable one to give a consistent account of it.

It might be thought that dialetheic accounts are immune to revenge problems. If one can intelligibly hold that standard liars are both true and false, or both true and not true, why not revenge liars as well? However, in this paper it is argued that dialetheism faces a dilemma: either it cannot express its core claim—that some, but not all, contradictions are true— or else it too suffers from revenge problems. I explore a few different ways in which a dialetheist might try to avoid this dilemma. First, I present a variant of the logic called LP, and show both that it is subject to revenge problems and that it is not well suited to a dialetheic interpretation. Second, I develop a means of expressing the exclusive truth or falsity of sentences which can be utilized by any language that has certain features. Unfortunately, it leads straight to trivialism. Finally, I examine the claim that dialetheists can express the exclusive truth or falsity of sentences in the same way a non-dialetheist can, and conclude that they cannot do so. My thesis is not entirely new[1], but nevertheless it is important and worth arguing for. At the least, my *arguments* for it will help advance the debate.

1. A Many-Valued Approach

One way in which one could try to express the exclusive truth or falsity of a sentence is to utilize a many-valued logic. I will adapt the three-valued logic called “LP”—Graham Priest’s “Logic of Paradox”— as expounded in Priest (1991) for this purpose. [2] In LP there are three truth values, {1}, {0}, and {1,0}; {1} being interpreted as *(exclusively) true*, {0} being interpreted as *(exclusively) false*, and {1,0} being interpreted as *both true and false*. LP is a first-order language, but I will omit the details pertaining to the quantifiers, predicates, etc., because the paradoxes we will consider concern only the propositional aspects of the language, identity being the sole exception. I think it is sufficient to say that predicates, including identity[3], have both a positive extension—the set of all objects of which the predicate is true—and a negative extension—the set of all objects of which the predicate is false—and that these extensions can overlap. Something similar goes for the connectives. On p. 324 of Priest (1991) Priest gives the following clauses for negation and conjunction, respectively:

$$1 \in v(\neg\alpha) \leftrightarrow 0 \in v(\alpha)$$

$$0 \in v(\neg\alpha) \leftrightarrow 1 \in v(\alpha)$$

$$1 \in v(\alpha \land \beta) \leftrightarrow 1 \in v(\alpha) \text{ and } 1 \in v(\beta)$$

$$0 \in v(\alpha \land \beta) \leftrightarrow 0 \in v(\alpha) \text{ or } 0 \in v(\beta)$$

(where '↔' means 'if and only if'). That is, 1 (true) is a member of the value of “$\neg\alpha$” [4] if and only if 0 (false) is a member of the value of α, and 0 is a member of the value of “$\neg\alpha$” if and only if 1 is a member of the value of α. Similarly, 1 is a member of the value of “$\alpha \land \beta$” if and only if 1 is both a member of the value of α and the value of β, and 0 is a member of the value of “$\alpha \land \beta$” if and only if 0 is either a member of the value of α or the value of β (or both).

In the remainder of this section I will present my variant of LP in a somewhat more familiar notation. Instead of ‘α’ and ‘β’, I will use ‘P’ and ‘Q’ as schematic letters, and in place of ‘∧’ I will use ‘&’. Also, instead of ‘{1}’ and ‘{0}’ I will use ‘T’ and ‘F’. Besides ‘both true and false’ or simply ‘both’, Priest has no name for the third value that I am aware of. Since I think it is convenient for it to have one, I will christen it ‘verifalsity’.[5] I will write the third value as ‘V’, and I will use the symbol ‘>’ as a verifalsity operator, “>P” being read as “It is verifalse that P”.[6]

So how could we, e.g., express the exclusive falsity of some sentence P in a system such as this? An intuitive way would be to say that P is neither verifalse nor true; i.e., “¬>P & ¬P”. Whether or not this works depends on what the truth tables for ‘¬’, ‘&’, and ‘>’ look like. Given the clauses for negation and conjunction provided above, if we define “>P” so that it is equivalent to “P & ¬P” we get the following:

&	**T**	**V**	**F**
T	T	V	F
V	V	V	F
F	F	F	F

P	**¬P**	**>P**	**P & ¬P**	**¬>P**	**¬>P & ¬P**
T	F	F	F	T	F
F	T	F	F	T	T
V	V	V	V	V	V

As you can see, “¬P” and “¬>P & ¬P” are equivalent, so according to these truth tables being false is equivalent to being *exclusively* false, which is undesirable if we want to draw a distinction between a sentence’s being false and its being exclusively false. Furthermore, if P is verifalse then so is “¬>P & ¬P”, which means that “P is exclusively false” can itself be both true and false! Is there any way we can avoid this?

As it turns out, there is. We can re-define “>P” so that if P takes the value V “>P” takes the value T, making “>P” *exclusively* true. Doing this enables us to distinguish between two kinds of negation. I will call the first, ‘~’, *inverse negation* and the second, ‘¬’, *alternative negation*. We can read “~P” as “It is inversely not the case that P”, and “¬P” as “It is alternatively not the case that P”. The truth table for these three operators is as follows:

P	**~P**	**¬P**	**>P**
T	F	F	F
F	T	T	F
V	V	T	T

The idea behind our terminology is this: The *inverse negation* of P takes, for every possible value of P, a unique *inverse value*. It takes T to F, F to T, and V to V. The value of the inverse negation of P is a function of the value of P and *vice versa*. On the other hand, if we call the values which are distinct from T *alternative values*, we can say the *alternative negation* of P takes the value T just in case P takes one of these alternative values, and takes the value F otherwise. The value of the alternative negation of P is a function of the value of P, but the value of P is not a function of the value of the alternative negation of P: According to the above truth table, if the alternative negation of P takes the value T, P can take either the value F or the value V.

Given the above, one can define “¬P” as “~P ∨ >P”. However, having both inverse and alternative negation operators allows us to define disjunction in different ways. Fortunately, there is a natural choice in this case: If we were to define '∨' in terms of '¬' and then define '¬' in terms of '∨' we would have a circular definition. Accordingly, I will define "P ∨ Q" solely in terms of inverse negation, as "~(~P & ~Q)". (One can, of course, define other disjunction connectives in terms of alternative negation.) Thus, "~P ∨ >P" is equivalent to "~(~~P & ~>P)". Furthermore, ">P" is equivalent to "¬P & ¬~P".[7] More importantly, we can define "P is true only" as "¬>P & ¬~P", and "P is false only" as "¬>P & ¬P".[8] If P is verifalse, then it is neither true only nor false only, and we may say both that P is *weakly true* and that P is *weakly false*.

I will call the system that results from the above modifications "Exclusive LP", or "ELP" for short, because this system is designed to express exclusive truth and exclusive falsity. It is important to note that in ELP, unlike in LP, Disjunctive Syllogism holds for '∨' as I have defined it, which will have significance in what follows. This is true whether V is designated in ELP or not. A truth value is said to be *designated* if that value is preserved in valid inferences. In LP, both T and V are designated. Since the present system is a modification of LP, we may as well show some respect and begin by seeing what happens if V is designated.

Assume "P ∨ Q" has a designated value, "P ∨ Q" being defined as above. In order for Disjunctive Syllogism to be *invalid*, it would have to be the case that either "~(~P & ~Q)" and "~P" are both designated while "Q" receives an F, or that "~(~P & ~Q)" and "~Q" are both designated while P receives an F. Now, if V is designated, Disjunctive Syllogism is invalid when P and "~(~P & ~Q)" both take the value V and Q takes the value F, or conversely.[9] In spite of that, a restricted form of Disjunctive Syllogism *is* valid. Substituting "¬P" for Q, if "~(~P & ~¬P)" is designated then either P is designated or "¬P" is designated. So if "~(~P & ~¬P)" and "~P" are both designated then so is "¬P", since "¬P" is designated if "~P" is. And if "~(~P & ~¬P)" and "~¬P" are both designated then so is P, since P is designated if "¬P" is inversely not the case. So this restricted form of Disjunctive Syllogism is valid, which is all we will need in what follows.

But what happens if V is *not* designated? Assume, again, that "~(~P & ~Q)" is designated. Since T is now the only designated value, "~(~P & ~Q)" must have the value T. So, in order for Disjunctive Syllogism to be *invalid*, it would have to be the case that either "~(~P & ~Q)" and "~P" both take the value T while "Q" takes the value F, or for "~(~P & ~Q)" and "~Q" to both take the value T while P takes the value F. Assume the first disjunct of the previous sentence. In that case "~P" receives a T and Q receives an F. Since Q receives an F, "~Q" receives a T. By Conjunction Introduction we get "~P & ~Q". But in this case we also have "~(~P & ~Q)", which was included in the first disjunct. By Conjunction Introduction we get "(~P & ~Q) & ~(~P & ~Q)", which is a contradiction. So we have a *reductio* of the first disjunct. A parallel argument gives us a *reductio* of the second disjunct. So we have a *reductio* of both potential counterexamples to Disjunctive Syllogism, and hence this form of Disjunctive Syllogism must also be valid for '∨'.[10]

One thing to notice about ELP is that sentences of the form "¬P" and ">P" never take the value V, i.e., they are never verifalse. For if P takes either the value T or the value F ">P" takes the value F, and if P takes the value V then ">P" takes the value T. In order for ">P" to take the value V, either P would have to take a value which is distinct from T, F, and V, or we would be forced to conclude that the value of ">P" is not a function of the value of P. Similar reasoning holds for "¬P". These properties are attractive if one wants to express the truth or falsity of P in a way that rules out its being verifalse. But they come at a price: They are *ad hoc*, and give rise to a revenge liar that the system cannot handle, because the system has four features that are jointly inconsistent.

First, "P ∨ ¬P" is a logical truth in this system, for it always takes the value T, as can be checked with the truth table for '¬' and the above definition of "P ∨ Q". Second, by virtue of the truth table for '¬' and the truth table for '&' given above, it is impossible to have "P & ¬P".[11] Third, Disjunctive Syllogism is valid for '∨', as was shown above. Finally, there are instances in which P and "¬P" entail each other. This is easy enough to show. Consider the following sentence, a "revenge liar"[12]:

(1): It is alternatively not the case that (1).

(Formally: (1): "¬(1)".) If (1) takes the value T, then what it says is the case, and since what it says is that its alternative negation is the case, it follows that "¬(1)" takes the value T. So (1) entails "¬(1)". And if "¬(1)" takes the value T, it is alternatively not the case that (1), and since that is precisely what (1) says, (1) also takes the value T. So "¬(1)" entails (1). Thus there is at least one instance in which P and "¬P" entail each other.

Now for the paradox. First, we take (1) and "¬(1)", which we have just seen to be sentences that entail each other. We can assume "(1) ∨ ¬(1)", which is guaranteed to be true because of the first feature of this system. Then we assume (1) for *reductio*.[13] By our first assumption, (1) entails "¬(1)". We have (1), so we can also obtain "¬(1)". By Conjunction Introduction, we obtain "(1) & ¬(1)". But by the second feature, this is impossible. Thus we can conclude "~(1)", and must accordingly reject (1). Since we have "(1) ∨ ¬(1)", Disjunctive Syllogism forces us to conclude "¬(1)". But since "¬(1)" entails (1) by our first assumption, we can obtain (1), and hence "(1) & ¬(1)" by Conjunction Introduction. As above, this is impossible. So we must accept "~¬(1)", and thus reject "¬(1)". So even though "(1) ∨ ¬(1)" is a logical truth in this system, we cannot accept either disjunct![14]

. This is not all. Another problem with ELP is how the notion of a unitary value of verifalsity and its attendant verifalsity operator are to be interpreted. For one thing, ">P" and "P & ~P" are not equivalent.[15] "P & ~P" is logically indeterminate in this system, and can take either the value F or the value V, but never T. ">P", on the other hand, while also logically indeterminate, can take on the value F or the value T, but never V. This can be checked with the following truth table:

P	**~P**	**>P**	**P & ~P**
T	F	F	F
F	T	F	F
V	V	T	V

">P", then, does not express the conjunction "P & ~P", and hence not the usual notion of a contradiction. And whatever "P & ~P" may express, it is not the claim that it is verifalse that P. So in addition to suffering from revenge problems, the system we have been examining is illsuited to a dialetheic interpretation. Given the failure of this approach, dialetheists must look for another way to express the claim that some sentences are not both true and false.

2. A More General Approach

In our first attempt we limited ourselves to a single language, but this attempt will apply to any language which has the following features: First, it allows self reference. Second, it contains at least two truth values which are interpreted as *true* and *false*. Third, a sentence may have both of these truth values. Fourth, it contains a sentence TT which is logically equivalent to the sentence "Every sentence is true". Finally, it has a connective '<==>', which I will call "aletheic equivalence"; "P <==> Q" being read as "P is aletheically equivalent to Q". We stipulate that "P <==> Q" is true if and only if every truth value which is assigned either to P or to Q is assigned to both, and that "P <==> Q" is false if and only if there is some truth value which is assigned to one but not the other. Alternatively, we could define "aletheic inclusion", i.e., "P => Q" as "Every truth value which is assigned to P is assigned to Q", and then define "P <==> Q" as "(P => Q) & (Q =>P)".

Aletheic equivalence is a different notion from that of material equivalence. If P is assigned the values T and F and Q is assigned only the value F on an interpretation, then "P is materially equivalent to Q" is also assigned both a T and an F. It must be assigned a T because both P and Q are assigned an F, and by the standard truth table for material implication any two false sentences materially imply each other. It must also be assigned the value F, because P is assigned a T and Q is assigned an F, and by the standard truth table for material implication a material implication is assigned an F if its antecedent is assigned a T and its consequent is assigned an F. So it is false that P materially implies Q, and hence false that P and Q are materially equivalent. It follows that "P is materially equivalent to Q" is both true and false on this interpretation. By contrast, if P is assigned the values T and F and Q is assigned only the value F on an interpretation, "P is *aletheically equivalent* to Q" is assigned an F, but not a T, since the value T is assigned to P but not to Q. In order for a sentence of the form "P <==> Q" to be assigned both a T and an F, it would have to be the case that every truth value which is assigned either to P or to Q is assigned to both, while at the same time some truth value is assigned to one but not the other. In short, we would have to have not merely an interpretation of the language which *represents* an inconsistent state of affairs, but rather an interpretation of the language which was *itself* inconsistent.

We now have the resources to express the claim that some arbitrary sentence P is true only, and the claim that some arbitrary sentence P is false only. In order to express the claim that P is false only, we say that P is aletheically equivalent to TT. TT is certainly false only, for it says that every sentence is true. Consequently, it must be assigned the value F, but not T. So if P is aletheically equivalent to TT, it too must receive the value F, but not T, and is accordingly false only. Similarly, to express the claim that P is true only, we say that "¬P" is aletheically equivalent to TT, or, to put it formally, "(¬P <==> TT)"; '¬' no longer being understood as alternative negation. (Saying "It is false that P is aletheically equivalent to TT," that is, "¬(P <==> TT)", is not enough, because it is also false that sentences which are assigned both a T and an F are aletheically equivalent to TT. But if "¬P" is aletheically equivalent to TT, "¬P" is false only, and hence P is true only.)

Unfortunately, this approach has problems of its own. Again we face a revenge liar paradox, one that not even a dialetheist can accept. Consider (2):

(2): This sentence is aletheically equivalent to TT.

(Formally: (2): (2) <==> TT.) Is (2) true? If so, (2) is aletheically equivalent to TT, for that (2) and TT are aletheically equivalent is precisely what (2) says. By the definition of "being aletheically equivalent", any interpretation which assigns the truth value T to (2) must assign the truth value T to TT as well. Thus if (2) is true, TT must be true as well. But if TT is true, every sentence is true. On pain of trivialism, (2) cannot be true. Suppose now that (2) is false. In that case, it is false that (2) is aletheically equivalent to TT. TT is false only, so if it is false that (2) is aletheically equivalent to TT, (2) must be true[16], and in virtue of what (2) says we must conclude that (2) is also aletheically equivalent to TT, in which case, since (2) is true, TT is also true. Again we get the result that every sentence is true.

Let us see how this works in a particular case. Returning to LP, we can set up the aletheic equivalency paradox as follows. The clauses for aletheic equivalence are (to revert to Priest's symbolization):

$$1 \in v(\alpha <==> \beta) \leftrightarrow v(\alpha) = v(\beta)$$

$$0 \in v(\alpha <==> \beta) \leftrightarrow v(\alpha) \neq v(\beta)$$

The justification for these clauses is that, as in the original definition of aletheic equivalence, P and Q are aletheically equivalent just in case every truth value assigned to one is assigned to the other. In terms of Priest's set-theoretic symbolization, two formulas α and β are aletheically equivalent just in case the *set* of truth values assigned to α is a subset of the *set* of truth values assigned to β, and vice versa. But if two sets are subsets of each other, they are the same set. Hence the occurrence of identity and distinctness on the right hand sides of the clauses. In what follows, I will call a set of truth values a *truth status*.[17]

Now for the paradoxical sentence. Consider (3):

(3): (3) <==> TT

Obviously, it cannot be the case that $1 \in v(\text{TT})$, for then everything is true. So $v(\text{TT}) = \{0\}$. Now, what is the truth status of (3)? If $v((3)) = \{1\}$, then by the definition of '<==>' $v((3)) = v(\text{TT})$, so $v(\text{TT})=\{1\}$. That can't be, so $v((3)) \neq \{1\}$. It follows that either $v((3)) = \{0\}$, or that it is also the case that $1 \in v((3))$. Could it be that $1 \in v((3))$, so that the status of (3) is {1,0}, and hence $v((3)) \neq v(\text{TT})$? No, for if $1 \in v((3))$ then (3) is true (even if false as well), and so $v((3)) = v(\text{TT})$ (even if it is also the case that $v((3)) \neq v(\text{TT})$). So it is not the case that $1 \in v((3))$, and $v((3)) = \{0\}$. Hence the truth statuses of (3) and TT are both {0}, and $v((3)) = v(\text{TT})$. By the clauses for aletheic equivalence, we have "(3) <==> TT", because their truth statuses are the same. But if "(3) <==>TT" is true then in virtue of what (3) says $1 \in v((3))$. And as we established above, if $1 \in v((3))$ then $1 \in v(\text{TT})$. So $1 \in v(\text{TT})$, and everything is true.

The notion that everything is true is anathema to dialetheists and non-dialetheists alike. How we are to avoid it is unclear, but whatever we may say, we cannot say that sentences like

(2) and (3) are both true and false. Thus they cannot be given the standard dialetheic treatment. Of course, whatever solution a non-dialetheist adopts, a dialetheist could adopt as well. The question is whether they can do so in a way that is not arbitrary. Let us take Grelling's Paradox as an example. We stipulate that a predicate is *autological* if and only if it is in its own extension, and that a predicate is *heterological* if and only if it is false that it is autological. From these stipulations it would appear to follow that 'heterological' is both heterological and autological. It would be arbitrary to say that the standard liars are both true and false, but that 'heterological' is meaningless—as opposed to being both heterological and autological—because its definition is impredicative. If (2) and (3) are in the liar family dialetheists should treat them just as they treat other liars. And if we can resolve (2) and (3) without declaring them to be both true and false, or both true and not true, could not other liars be dealt with in the same way? If they could then dialetheism, while not refuted, would lose much of its motivation.

3. First Objection

On pp. 287-90 of Priest (2006a), Priest discusses a paradox which, at first sight, is very similar to (2) and (3) above. He considers a sentence ξ of the form "$F<\xi> \land \neg T<\xi>$"[18], i.e., which says of itself that it is false and not true. Suppose we define a valuation function, Val, like so:

1. Val $<\alpha> = \{1\}$ if and only if $T<\alpha> \land \neg F<\alpha>$

2. Val $<\alpha> = \{1, 0\}$ if and only if $T<\alpha> \land F<\alpha>$

3. Val $<\alpha> = \{0\}$ if and only if $\neg T<\alpha> \land F<\alpha>$

Formulated in these terms, ξ becomes a sentence of the form Val $<\xi> = \{0\}$. From this it follows pretty quickly that $0 = 1$, which means that everything is true. However, Priest points out that cases 2 and 3 of the definition of Val overlap, and as he rightly notes, "This is not, therefore, a good definition (any more than is a definition of a numerical function, f, such that $f(n) = 1$ if $n \leq 5$ and $f(n) = 0$ if $n \geq 5$)" (Priest 2006a, p.288).

To avoid such problems, Priest proposes to use relational evaluations in place of functions. So instead of saying things like $\text{Val} <\alpha> = \{1\}$; or $\text{Val} <\alpha> = \{1, 0\}$; or $\text{Val} <\alpha> = \{0\}$; he would say $\text{Rel}(<\alpha>, 1) \wedge \neg\text{Rel}(<\alpha>, 0)$; or $\text{Rel}(<\alpha>, 1) \wedge \text{Rel}(<\alpha>, 0)$; or $\neg\text{Rel}(<\alpha>,1) \wedge \text{Rel}(<\alpha>, 0)$. Thus, on p. 289 of Priest (2006a) he says of ξ: "ξ now becomes a sentence of the form $\text{Rel}(<\xi>, 0) \wedge \neg\text{Rel}(<\xi>, 1)$, from which we can infer $\text{Rel}(<\xi>, 1) \wedge \text{Rel}(<\xi>, 0) \wedge \neg\text{Rel}(<\xi>, 1)$—as one would expect—but go no further" [footnote omitted]. Priest also points out (Priest 2006a, p. 289) that one could also define "$\text{Val} <\alpha>$" in terms of the relational evaluations by using set abstracts, as: $\{x : \text{Rel}(<\alpha>, x)\}$. ξ would then be of the form: $\{x : \text{Rel}(<\xi>, \text{x})\} = 0$. But Priest goes on to show that ξ, so construed, does not entail triviality. That being so, Priest could urge that my aletheic equivalency paradox can be dealt with in the same way.

In response, I would agree with Priest that ξ does not entail triviality when relational evaluations are being used. Nevertheless, I think a variant of the aletheic equivalency paradox does go through. Using set-abstracts, we can re-define '<==>' in this way:

$$\text{Rel}(<\alpha <==> \beta>, 1) \leftrightarrow \{x : \text{Rel}(<\alpha>, x)\} = \{x : \text{Rel}(<\beta>, x)\}$$

$$\text{Rel}(<\alpha <==> \beta>, 0) \leftrightarrow \{x : \text{Rel}(<\alpha>, x)\} \neq \{x : \text{Rel}(<\beta>, x)\}$$

That is, "$\alpha <==> \beta$" relates to 1 just in case the set of all values to which α relates is the same set as the set of all values to which β relates, and "$\alpha <==> \beta$" relates to 0 otherwise. If aletheic equivalence is defined like this, the argument for triviality goes through almost exactly as it did with sentence (3), so I will spare the reader the details. The important thing to note is that, in contrast to ξ, the argument does not depend on proving that $\{x : \text{Rel}(<(3) <==> \text{TT}>, x)\} = 0$ in order to show that trivialism follows. Instead, we need to show that "(3) <==> TT" relates to 1, and that if it does then so does TT; whether "(3) <==> TT" also relates to 0 is immaterial.

4. Second Objection

My thesis, to recall, is that dialetheists either cannot express the exclusive truth or falsity of sentences, or that if they can they face revenge liars for which they cannot give a dialetheic solution. If this is correct, it had better be the case that there is a non-dialetheic approach to liar paradoxes which is not subject to these same problems, otherwise *everyone* is in trouble. It follows that there must be some way a non-dialetheist can express the exclusive truth or falsity of sentences which does *not* generate irresoluble revenge liar paradoxes. So one objection a dialetheist could raise against arguments like the foregoing—and which has in fact been raised by Priest[19]—is that expressing the exclusive truth or falsity of sentences requires no special logical machinery on the part of the dialetheist, because they can express such claims in precisely the same way a non-dialetheist can.

One way for a non-dialetheist to express the exclusive falsity of some sentence P is to say "P is false". Since they think that falsity excludes truth, this is all they need to say. However, dialetheists think that falsity does not exclude truth, so for them saying this will not do the trick. But the non-dialetheist could also express the exclusive falsity of P by saying "P is false only", or "P is false and not also true", or something similar. Whatever phrasing they use, a dialetheist can also express the exclusive falsity of P by using the same words. If these words are unproblematic when a non-dialetheist uses them, they should be equally unproblematic when a dialetheist uses them.

Or so it might seem. But even if a dialetheist says "P is false only" and means it—or tries to mean it—this will not by their lights exclude the truth of P. One can easily generate liars such as "This sentence is false only", which a dialetheist will naturally want to regard as being both true and false only. It appears that on closer inspection a dialetheist is prepared to use 'false only' and similar locutions in ways no non-dialetheist would.

What should we conclude from this? I think a non-dialetheist should insist that if the truth of some such sentence as "P is false only" does not exclude its falsity on a dialetheist's usage then 'false only' cannot mean the same as it normally does. The meaning of an expression is determined by its *whole* use, so even if both dialetheists and non-dialetheists use ‘false only’ in the same way in contexts which do not involve liar paradoxes this does not guarantee sameness of meaning. In order for it to have the same meaning it must be used the same way in paradoxical contexts too. But this condition is not met in the present case. In a dialetheist’s usage “P is false only” has different truth conditions from what it has in a non-dialetheist’s usage, for on the former usage it does not exclude the truth of P and on the latter usage it does. So ‘false only’ must have different conditions of application on the two usages, and hence a different meaning. If a dialetheist is prepared to treat some sentences as being in both the extension and anti-extension of a predicate *F* in a dialetheic logical system, *F* cannot express exclusive falsity in that system. It does not necessarily follow from this that dialetheism is not true, but it does follow that a dialetheist cannot express the exclusive truth or falsity of sentences in the same way a non-dialetheist can.

5. Conclusion

The foregoing considerations show that dialetheists are in a precarious position. To distinguish themselves from trivialists, dialetheists must be able to make sense of the idea that some but not all contradictions are true. Yet we have seen that either they cannot express this idea or else that in doing so they are faced with paradoxes that their accounts cannot handle. While dialetheic accounts of the logical and semantic paradoxes certainly have their attractions, it seems that in the end dialetheism’s dilemma is inescapable.[20]

References

Beall, J. C. (ed.), (*2007), Revenge of the Liar: New Essays on the Paradox*. Oxford University Press, Oxford.

Burgis, B. (2008), 'Can Dialetheists Make Sense of Monaletheias?', *(Blog & ~Blog)*. Available at URL = http://blogandnot-blog.blogspot.com/2008/08/can-dialetheists-make-sense-of.html (Accessed: 19 May 2009).

Littmann, G. and Simmons, K. (2004), 'A Critique of Dialetheism', in *G. Priest, J.C. Beall & B. Armour-Garb (eds.) The Law of Non-Contradiction: New Philosophical Essays,* Oxford University Press, Oxford, pp. 314-35.

Priest, G. (1991), 'Minimally Inconsistent LP', *Studia Logica: An International Journal for Symbolic Logic*, 50, pp. 321-31.

Priest, G., J.C. Beall and B. Armour-Garb (eds.), *(2004), The Law of Non-Contradiction: New Philosophical Essays,* Oxford University Press, Oxford.

Priest, G. (2006a), *In Contradiction, 2nd edition*, Oxford University Press, Oxford

Priest, G. (2006*b), Doubt Truth to be a Liar*, Oxford University Press, Oxford.

Priest, G. (2008), *An Introduction to Non-Classical Logic: From If to Is, 2nd edition,* Cambridge University Press, New York.

Shapiro, S. (2004), 'Simple Truth, Contradiction, and Consistency'*, in G. Priest, J.C. Beall & B. Armour-Garb (eds.) The Law of Non-Contradiction: New Philosophical Essays,* Oxford University Press, Oxford, pp. 336-54.

[1] See Beall (2007). See also Littmann and Simmons (2004), and Shapiro (2004). Finally, see Burgis (2008).

[2] See also Priest (2008), pp. 124-5

[3] Priest requires that the positive extension of the identity predicate is the set $\{<x,x>; x \in D\}$, where D is the (non-empty) domain of quantification (Priest (1991), 323).

[4] Throughout this paper I will adopt the convention of enclosing formal sentences such as “¬α” in quotation marks, except when they contain no logical connectives.

[5] ‘Veri’ being derived from the Latin word ‘veritas’, which means *truth*, and ‘falsity’ (of course!) being “derived” from the English word ‘falsity’.

[6] Thus “>P” is not the same as the metalinguistic sentence “‘P’ is verifalse”, where ‘is verifalse’ functions as a predicate.

[7] (To avoid ambiguities of scope, “~P ∨ >P” is to be read as, “Either it is inversely not the case that P or it is verifalse that P,” and “¬P & ¬~P” is to be read as “It is alternatively not the case that P and it is alternatively not the case that it is inversely not the case that P.”)

[8] (Again, to avoid ambiguities of scope, “¬>P & ¬~P” is to be read as, “It is alternatively not the case that it is verifalse that P and it is alternatively not the case that it is inversely not the case that P,” and “¬>P & ¬P” is to be read as “It is alternatively not the case that it is verifalse that P and it is alternatively not the case that P.”)

[9] Thus, since Disjunctive Syllogism is not valid in general, one cannot use it to derive arbitrary sentences from a contradiction.

[10] Even if V is *not* designated, can one use Disjunctive Syllogism to derive arbitrary sentences from a contradiction? The answer is no. Although Disjunctive Syllogism is generally valid if V is not designated, one could only use it to derive an arbitrary sentence if a sentence of the form “P & ~P” took the value T. In that case one could use Conjunction Elimination on “P & ~P” to get P and go on to infer “P ∨ Q”, and then use Conjunction Elimination again to get “~P” and finally use Disjunctive Syllogism to infer Q. But in ELP sentences of the form “P & ~P” cannot take the value T, because P and “~P” can never both take the value T. They could both take the value V, in which case we have a sentence of the form “>(P & ~P)”. So “>(P & ~P)”can take the value T, but from “>(P & ~P)”one cannot infer P, only “>P”. One could then obtain “>P ∨ Q”, but in order to get Q one would have to first get “~>P” which one cannot do here. One *could* get “>~P” from “>(P & ~P)”, but “>~P” and “~>P” are two *very* different schemas.

[11] That is, “P & ¬P” never receives the value T. But if P receives the value V then “¬P” receives the value T, and “P & ¬P” receives the value V. This does not affect the argument in the main text because if we had both (1) and “¬(1)”—that is, if (1) and “¬(1)” both took the value T—then “(1) & ¬(1)” would receive both a T and an F, not a V. It is important to remember that for now we are supposing that T represents *exclusive* truth and F *exclusive* falsity, while V represents the conjunction of what I have called ‘weak truth’ and ‘weak falsity’. So in ELP V is an admissible evaluation while T,F is not.

[12] A “standard liar”, e.g.,

(L**): “~(L**)”

would be assigned the value V, but our revenge liar could not take that value because it has the form of an alternative negation.

[13] Since we have two negation operators, one may wonder whether the conclusion of a *reductio* of P ought to be “~P” or “¬P”. My answer is that, since “~P” entails “¬P” but not conversely, we ought to conclude “~P”. “¬P” is consistent with “>P”, and if “>P” is true then P is weakly true because it is verifalse. In my opinion any *reductio* of P worthy of the name ought to rule out even the weak truth of P, and this makes “~P” is a better choice than “¬P”. That is, “~P” rules out the weak truth of P if “~P” is true only, while “¬P” does not rule out the weak truth of P if “¬P” is true only. “~P”, of course, does not rule out the weak truth of P if “~P” is verifalse.

[14] Thanks to J. C. Beall and Bas van Fraassen for bringing this argument to my attention.

[15] What about “P & ¬P”? It is equivalent to “P & ~P”, so it too is not equivalent to “>P”.

[16] Normally there would be two possibilities here: If it is false that some sentence P is aletheically equivalent to a sentence which is false only, P could be either true only or both true and false. But (2) has the form of an aletheic equivalence, and as we saw above an aletheic equivalence could only be both true and false on an interpretation of a language which was itself inconsistent. In any case, if (2) were both true and false it would be true, in which case TT would be true as well, and we would still get the result that every sentence is true.

[17] As I will use the term, every truth status is a truth value, but not conversely. Truth values which are not truth statuses—such as 1 and 0—could be called *atomic truth values*.

[18] ‘< >’ being a name-forming device.

[19] See Priest (2006b), 106.

[20] Several people helped me out with this paper, but I'd like to thank Bas van Fraassen especially for his time, his encouragement, and his many helpful comments. Thanks also to Kris Kemtrup, Ben Burgis, Justin Tiwald, James Blackmon, Alex Hyun, Miguel Balboa, Tyrus Fisher, Asta Sveinsdottir, Graham Priest, J.C. Beall, and anyone who I may have forgotten.