Is the Liar Sentence Both True and False?

Hartry Field
New York University

"Will no one rid me of this accursed Priest?"–Henry II

1 Dialetheism

There are many reasons why one might be tempted to reject certain instances of the law of excluded middle. And it is initially natural to take 'reject' to mean 'deny', that is, 'assert the negation of'. But if we assert the negation of a disjunction, we certainly ought to assert the negation of each disjunct (since the disjunction is weaker[1] than the disjuncts). So asserting $\neg(A \vee \neg A)$ should lead us to assert both $\neg A$ and $\neg\neg A$. But to assert both a sentence ($\neg A$) and its negation is, in at least one sense of the phrase, to assert a contradiction.

Accepting contradictions would be intolerable if contradictions implied everything: we would be logically committed to every imaginable absurdity. But there are "paraconsistent logics" where contradictions in the above sense (pairs consisting of B and $\neg B$, for some B) don't entail everything.

It is not especially controversial that paraconsistent logics might be useful for certain purposes, e.g. analyzing certain notions of "relevant implication" and/or "what a possibly inconsistent theory should be taken as directly committed to". But I'm interested in the issue of a particular kind of use, the one motivated above: a use of paraconsistent logic to license the simultaneous literal belief in both B and $\neg B$, in full knowledge that we believe both, and where such knowledge gives no pressure to revise one of the beliefs. In short, where the beliefs, though "contradicting" each other, are not in any serious sense in conflict. I will adapt Graham Priest's term 'dialetheism' for the doctrine that we should fully accept certain sentences while also accepting their negations. This is not quite Priest's usage, as we'll see. Nonetheless, Priest is an advocate of dialetheism in this sense; in fact, its most prominent advocate.

The argument with which I began shows that if we want to disbelieve instances of excluded middle (in the sense of, believe their negations) then we should be dialetheists (not merely that we should accept paraconsistent logics for some purposes). And as Priest has often urged (e.g. [12]), the most familiar arguments against the coherence of dialetheism are seriously faulty, a result of a refusal to take the doctrine seriously.

[1] Not necessarily strictly.

I have two terminological quibbles with Priest. The more minor one concerns the use of the term ‘contradiction’. Priest revels in saying that we should accept contradictions. Here ‘contradiction’ is used either in the sense indicated above (a pair of a sentence and its negation) or in the sense of a sentence of form $B \wedge \neg B$; since Priest and I both advocate the use of logics in which any two sentences imply their conjunction and in which a conjunction implies each conjunct, there is no interesting distinction between accepting a contradiction in one of these senses and accepting it in the other, so I will henceforth not bother to make this distinction. Talk of accepting contradictions shows a flair for the dramatic, but I think it tends to put people off for bad reasons. Given the kind of logic Priest advocates, I think a better use of the term ‘contradiction’ would be: sentence that implies every other. On this alternative usage, the way to put Priest’s view is that sentences of form $B \wedge \neg B$ (or pairs $\{B, \neg B\}$) aren’t in general contradictory: they don’t imply everything. The issue of course is purely verbal; but because of the possibility of confusion, I will from now on avoid the term ‘contradiction’.

A somewhat more important terminological issue concerns the term ‘dialetheism’. Priest explains dialetheism as the doctrine

(D) Certain sentences are both true and false;

where ‘false’ means ‘has a true negation’. There is no doubt that a dialetheist should accept (D); nonetheless, (D) seems to me an unfortunate way to define the term.

To see why, let me anticipate what is to follow, by saying that one of the main prima facie benefits of dialetheism in my sense is that it allows us, despite the semantic paradoxes, to maintain the complete intersubstitutivity of $True(\langle A \rangle)$ with A (in contexts not involving quotation marks, propositional attitudes, etc.). Since $False(\langle A \rangle)$ means $True(\langle \neg A \rangle)$, this means that *False(⟨A⟩) will be completely intersubstitutable with ¬A, and hence with ¬True(⟨A⟩).*

If we agree to these properties of ‘True’, then acceptance of $A \wedge \neg A$ should be equivalent to the acceptance of $True(\langle A \rangle) \wedge False(\langle A \rangle)$. So if Priest were to have put dialetheism as the view that we should fully accept of some sentences that they are both true and false, I could have no complaint: it would be effectively the same as my own ‘true’-free formulation. But a problem with defining dialetheism as the doctrine (D) that certain sentences *are* both true and false is that while a dialetheist should certainly assert

(i) $True(\langle A \rangle \wedge False(\langle A \rangle)$

for certain A (e.g. the Liar sentence), he should deny this as well. For the dialetheist asserts both $True(\langle A \rangle)$ and $False(\langle A \rangle)$. But from $False(\langle A \rangle)$ we get $\neg True(\langle A \rangle)$, by the italicized claim in the last paragraph; so

(ii) $False(\langle A \rangle) \wedge \neg True(\langle A \rangle,$

which surely entails the negation of (i). If we assume that A entails $\neg\neg A$ (as nearly everyone would, including Priest), then from $True(\langle A \rangle)$ we get $\neg\neg True(\langle A \rangle)$, which yields $\neg False(\langle A \rangle)$ by the same italicized principle; so we also get

(iii) $True(\langle A \rangle \wedge \neg False(\langle A \rangle)$

and

(iv) $\neg True(\langle A\rangle) \wedge \neg False(\langle A\rangle)$,

each of which also entails the negation of (i). The situation for the existential generalization (D) is a bit more complicated, but here too I think Priest needs to deny it as well as assert it; hence, given the equivalence, regard it as false as well as true. Of course, it is a consequence of dialetheism that some sentences are both true and false, and there's no particular problem in the fact that the particular sentence (D) is among them. But what is odd is to take as the doctrine that *defines* dialetheism something that the dialetheist holds to be false as well as true. And it is misleading to characterize the dialetheist's attitude toward, say, the Liar sentence as the view (i) that it is both true and false, when one could equally well have characterized it as the view (iv) that it is neither true nor false, or as the view (ii) that it is false and not true, or the view (iii) that it is true and not false. On the alternative characterization of dialetheism in terms of acceptance, there are no such oddities: no dialetheist (in either Priest's sense or mine) has any reason to deny that we should accept both A and $\neg A$, for the relevant A.

Priest could resist my complaint, on the grounds that he himself does not accept the full intersubstitutivity of $True(\langle A\rangle)$ with A: he takes them not to be intersubstitutable in negation contexts ([11], secs. 4.9 and 5.4), and that destroys the equivalence between 'false' and 'not true'. Nonetheless, I don't think he ought to resist: the alternative characterization I have suggested ought to be acceptable to him, and has the advantage of not requiring the non-intersubstitutability of $True(\langle A\rangle)$ with A. Moreover, one of the advantages of dialetheism, in either sense, is that it makes possible the full intersubstitutivity of $True(\langle A\rangle)$ with A; it would be a shame to adopt a definition of dialetheism that is badly behaved when this possibility is realized.

Enough of terminological quibbles, and on to more serious issues.

2 Rejection

I've mentioned that we have a compelling motivation for dialetheism as long as we have a compelling motivation to deny instances of excluded middle. There is however an oddity here. For though accepting $\neg[A \vee \neg A]$ leads to accepting both $\neg A$ and $\neg\neg A$, that in turn leads to accepting $A \vee \neg A$. Indeed, the inference is immediate: $\neg A$ surely implies $A \vee \neg A$. ($\neg\neg A$ implies it too, in any logic where $\neg\neg A$ implies A; and the logics that Priest and I are interested in mostly have this feature. Indeed, in the future I will for simplicity assume that $\neg\neg A$ is equivalent to A.) So if to *reject* B is to have an attitude that precludes accepting it, then *denying* instances of excluded middle is not a way of rejecting them, and indeed *is incompatible with* rejecting them! If we want to *reject* a given instance of excluded middle, we had better not deny it. (I have not ruled out rejecting some instances without denying them, and denying others without rejecting them. But we can't reject and deny the same instance.)

How should we understand rejection? As a propositional attitude, on par with acceptance. To a first approximation anyway, accepting A is having a high degree of belief in it; say a degree of belief over a certain threshold T, which may depend on context but must be greater than $\frac{1}{2}$. (Degrees of belief are assumed to be real numbers in the interval $[0,1]$.) To the same degree of approximation, rejecting A is having a low degree of belief in it: one lower than the co-threshold $1-T$. This has the desired result that rejection precludes acceptance. (And it allows, as of course we should, for sentences that we are uncertain enough about to neither accept nor reject.)

Now, if degrees of belief obeyed the laws of classical probability, then rejecting A would have to be the same as accepting its negation. For in classical probability theory, $P(A) + P(\neg A)$ is always 1; so $P(A) < 1 - T$ (rejection) iff $1 - P(\neg A) < 1 - T$ iff $P(\neg A) > T$ (acceptance of negation). But if $P(A) + P(\neg A)$ could be greater than 1, then we could accept the negation of A without rejecting it; indeed if it could be sufficiently greater than 1, we could accept both A and $\neg A$ and therefore reject neither. And if $P(A) + P(\neg A)$ could be less than 1, we could reject A without accepting its negation; if it could be sufficiently less than 1, we could reject both A and $\neg A$. It's clear that a dialetheist ought to allow that $P(A) + P(\neg A)$ can be significantly greater than 1 (perhaps as high as 2): if you accept A and accept $\neg A$, you give both high probabilities (perhaps as high as 1); and you do not reject either since rejection precludes acceptance. Similarly, someone who rejects an instance $A \vee \neg A$ of excluded middle (not necessarily a dialetheist) will reject both A and $\neg A$, hence for that person $P(A) + P(\neg A)$ will be substantially less than 1 (perhaps as low as 0).

The upshot is that there is no problem distinguishing rejection from acceptance of the negation, in nonclassical logics that either don't include certain instances of excluded middle or include the negations of certain instances of it.

Where there may be a difficulty for dialetheism, though, is in conducting debates about what to reject. Suppose I reject the existence of God, and offer to my theistic friend compelling arguments against it. I expect my friend to try to rebut my arguments, or at least be worried about them (or, more optimistically, to recant his belief); but to my chagrin my friend turns out to be a dialetheist, and though he accepts my arguments and agrees with me about the nonexistence of God, he also believes in the existence of God. What I really want to do is alter his attitudes: get him to reject the existence of God, not merely disbelieve it. How can I proceed? Well, perhaps I can show him that the existence of God together with other things he accepts entails some other claim Q that I assume he'll reject; say, one for which I already know he accepts its negation. But if he's willing to carry his dialetheism far enough, he may be "dialetheist with respect to Q" as well as "dialetheist with respect to the existence of God": he may accept Q along with its negation. Perhaps there are certain sentences that he truly rejects, rather than merely accepts their negations; but these may not provide a sufficient basis to allow any argument that he should reject the existence of God, even when he can be convinced to accept the non-existence of God.

The worry, then, is that if the acceptance of $\neg A$ doesn't suffice for the rejection of A, it is unobvious how debates about rejection are to be conducted.

I take it to be an important challenge to dialetheism to answer this, but I don't mean to say that it is obvious that the dialetheist can't meet the challenge; indeed, I increasingly incline to the view that he can.[2]

3 Truth Paradoxes (I)

But is there any serious motivation for adopting a dialetheist position and therefore having to meet the challenge just mentioned? I'm skeptical. Obviously there's no way of completely ruling out in advance that there might be some problem solvable better by dialetheist means than by non-dialetheist, but I don't think there is any reason whatever to believe that this will be the case. Indeed, I think it very likely that any problem that can be solved by dialetheism can be solved without it, and when the best solutions of each sort are set side-by-side the non-dialetheic solution will always seem more attractive; in which case dialetheism is a position that we do not need.

I'm going to spend the rest of this paper illustrating this conjecture with a single example, but it is the example widely viewed as the dialetheist's best case: the semantic paradoxes.

The naive theory of truth presupposes a background syntactic theory, which can be formalized in arithmetic. In addition to this, it has at least two components. The first is the Tarski axioms:

$$(T): \qquad True(\langle A\rangle) \text{ if and only if } A.$$

The second is the principle mentioned already: $True(\langle A\rangle)$ should be everywhere intersubstitutable for A (in a language free of quotational contexts, intentional contexts, and so forth); that is, if B and C are two sentences alike except that one has $True(\langle A\rangle)$ in one or more places where the other has A, then B implies C and conversely. It may seem redundant to list these components separately, for they are equivalent in classical logic. But as we'll see, there are non-classical logics in which they are not equivalent. (Either direction of implication might fail.) Myself, I'm interested mostly in logics "classical enough" for the equivalence to hold; but we need to bear in mind the possibility of logics that aren't "classical enough".

In classical logic itself, each component of naive truth theory is inconsistent, given the background syntactic theory. For the syntactic theory allows us to construct a Liar sentence Q_0 which is interderivable with $\neg True(\langle Q_0\rangle)$; so the second component of the naive theory would make Q_0 interderivable with $\neg Q_0$. That would make Q_0 and $\neg Q_0$ each inconsistent, so their disjunction would be

[2] To be fair, the person who advocates restricting the law of excluded middle faces a somewhat analogous challenge: for on such a view we should sometimes reject A without accepting $\neg A$, and an account is needed of how to carry out debates about when this is to be done. In this case the question is how to deal with a view that rejects both 'there is a God' and its negation, perhaps on verificationist grounds, because it refuses to accept the corresponding instance of excluded middle.

For another observation on behalf of the dialetheist, see note 19.

inconsistent; but their disjunction is an instance of excluded middle and hence classically valid. So the second component of the naive theory is classically inconsistent; and since the first component is classically equivalent to it, it is classically inconsistent as well.

Kripke ([8]) shows that we can consistently retain **one component of** the naive theory of truth, by weakening classical logic to the logic K_3 obtained from the strong Kleene 3-valued truth tables by taking only the "highest" of the three values as "designated". More exactly, let the three semantic values be 1, $\frac{1}{2}$, and 0, thinking of 1 as "best" and 0 as "worst". Let an **assignment function** s be a function that assigns objects to variables, and let a **valuation** be a function that assigns semantic values to pairs of atomic formulas and assignment functions. Extend the valuation to complex formulas by the "strong Kleene rules"

$||\neg A||_s = 1 - ||A||_s$

$||A \wedge B||_s = min\{||A||_s, ||B||_s\}$

$||A \vee B||_s = max\{||A||_s, ||B||_s\}$

$||A \supset B||_s = max\{1 - ||A||_s, ||B||_s\}$

$||\forall x A||_s = min\{||A||_{s^*} \mid s^*$ is just like s except possibly in what it assigns to $x\}$

$||\exists x A||_s = max\{||A||_{s^*} \mid s^*$ is just like s except possibly in what it assigns to $x\}$.

Finally, regard an inference as **valid** iff in every valuation in which the premises all get value 1, so does the conclusion; and regard a sentence as valid iff in all valuations it gets value 1. Clearly the inference from $\{A, \neg A\}$ to any sentence B is valid; similarly for the inference from $\neg(A \vee \neg A)$ to anything. So this is not a logic for dialetheists, or, virtually equivalently, for deniers of excluded middle; but it is a logic that allows **rejecting** excluded middle, since excluded middle is not valid.[3] (Of course, no inference that isn't classically valid can be K_3-valid.)

It is natural to take a logic to include not only a set of validities, but to include rules for establishing some validities from others (relevant for when one expands the logic). One such rule, correct under the strong Kleene semantics, is disjunction elimination: if A implies C and B implies C, then $A \vee B$ implies C. I will henceforth understand K_3 to include this meta-rule (and the analogous rule of $\exists$-elimination).

Kripke's "fixed point argument" shows that if we weaken classical logic to K_3, then it is consistent to assume the second component of the naive truth theory: the intersubstitutivity of $True(\langle A \rangle)$ with A.[4] But we do **not** get either direction of the Tarski schema (taking 'if...then' to be represented by '$\supset$', which

[3] Indeed, no sentence (as opposed to inference) is valid in the logic as it stands. However, we will soon consider extensions of the logic in which this is not so.

[4] Indeed we get what I'll call "strong consistency": roughly speaking, it is "consistent with any starting model that has standard syntax".

seems the only remotely reasonable way to interpret it in K_3): for given the intersubstitutivity property of 'True', each direction of the Tarski schema is equivalent to $A \supset A$, but that is not valid in K_3 (it is equivalent to an instance of excluded middle).

Truth theory in K_3 is not a very satisfactory theory.[5] Not only do we not get the full naive theory of truth, the logic is in many ways simply too weak to comfortably reason in. (The absence of a conditional obeying the law 'if A then A' is a symptom of this.) And this fact could seem a motivation for dialetheism.

For there is a minor variant of the logic–Priest's LP–that can easily seem more satisfactory. As a semantics for LP we can assign semantic values in just the same way as for K_3; we simply redefine validity. (And we include the disjunction-elimination and $\exists$-elimination metarules, as before.) In particular, we take an inference to be valid just in case in all valuations where the premises all get values other than 0, so does the conclusion; and we take a sentence to be valid just in case it doesn't get value 0 in any valuation. The Kripke fixed point proof then shows that it is consistent to assume the intersubstitutivity property in this logic too; and this logic does validate $A \supset A$, so the truth schema is validated as well. We have the naive theory of truth in its entirety; the only cost is dialetheism, for the Liar sentence and its negation will both be consequences of the truth theory.

But I do not think this is really much of an improvement over the situation with K_3.[6] For the main problem with K_3 wasn't the inability to get the Tarski schema, it was the fact that the logic is too weak to reason with in a natural way, as indicated by the absence of a reasonable conditional. That is true for LP as well: '$\supset$' is in some ways even worse as a candidate for the conditional in LP than it is in K_3, because it doesn't even validate Modus Ponens.

So far, then, we don't have a satisfactory resolution of the paradoxes, either non-dialetheic or dialetheic.

4 Truth Paradoxes (II)

A natural response to the difficulties with using K_3 and LP as the logic for truth theory is to try to supplement one of them with a new conditional. Doing so requires great care: it is not easy to find a set of laws for the conditional that have reasonable strength and don't themselves lead to paradox given naive truth theory. For instance, the Curry paradox shows that (assuming naive truth theory), if our conditional satisfies Modus Ponens then it can't validate any of the following inferences:

[5] Indeed, it isn't completely clear that Kripke's own discussion of a truth theory based on the strong Kleene tables is intended to motivate a theory based on the logic K_3: it might be intended to motivate a theory, later formalized by Feferman ([1]), based entirely on classical logic. Of course, since it is based on classical logic, the "Kripke-Feferman theory" does not satisfy either component of the naive truth theory. (In particular, in the Kripke-Feferman theory we can assert that the Liar sentence is not true; hence we can assert the Liar sentence, but cannot assert that it is true.)

[6] Here I retract an ill-thought-out suggestion in [2], p. 145.

Importation:	$A \to (B \to C) \vdash A \wedge B \to C$
$\to$-Introduction:	From $A \vdash B$, infer $\vdash A \to B$
Contraction:	$A \to (A \to B) \vdash A \to B$.

(In the case of Importation, this requires the inference from $A \wedge A \to C$ to $A \to C$, but I take it that that inference is totally uncontroversial.)[7] Moreover, there seems to be no low-cost restriction of Modus Ponens that would improve the situation. For instance, it might initially seem that we could obtain a satisfactory logic by restricting the Modus Ponens rule $A,\ A \to B \vdash B$ to the case where A does not itself contain an $\to$; that would block the derivation of the Contraction form of the Curry paradox in the previous note.[8] A minor difficulty is that such a restricted Modus Ponens would be awkward to employ: we'd need to keep track of which sentences that we're representing with sentence letters have an $\to$ hidden inside them. A more serious difficulty is that since the goal is to keep the equivalence of $True(\langle A \rangle)$ to A, we'd need to also rule out applying Modus Ponens when A contains a predication of 'True' to a sentence with an $\to$. And we'd need to rule out application of the rule when the premise applies 'True' to all members of a class that may contain sentences with an '$\to$', e.g. 'All sentences on the blackboard are true'. Not only would it be a bit tricky to formulate the rule with all these added restrictions, but the resulting rule would be so restricted that our reasoning would be crippled.

So the only serious recourse, given that we want the naive theory of truth, is to adopt a logic that does not validate either Importation, $\to$-Introduction, or Contraction. Most people find the loss of Contraction surprising, but in fact

[7] First version of Curry paradox: For any sentence B, no matter how absurd, let C_B say "$True(\langle C_B \rangle) \to B$". We now "prove" B, using Modus Ponens, Contraction, and naive truth theory, as follows. (I also use $\wedge$-elimination, but this could be avoided simply by replacing each Tarski axiom in biconditional form by two conditionals.)

1. $True(\langle C_B \rangle) \leftrightarrow (True(\langle C_B \rangle) \to B)$	(Tarski axiom for C_B)
2. $True(\langle C_B \rangle) \to (True(\langle C_B \rangle) \to B)$	(1, defn of $\leftrightarrow$, and $\wedge$-elimination)
3. $True(\langle C_B \rangle) \to B$	(2, Contraction)
4. $(True(\langle C_B \rangle) \to B) \to True(\langle C_B \rangle)$	(1, defn of $\leftrightarrow$, and $\wedge$-elimination)
5. $True(\langle C_B \rangle)$	(3, 4, Modus Ponens)
6. B	(3, 5, Modus Ponens)

That establishes the claim for Contraction. As noted later in the text, this implies the claim for Importation and for a slightly stronger form of $\to$-Introduction, viz. the inference from $\Gamma, A \vdash B$ to $\Gamma \vdash A \to B$.

A second (and more famous) proof of the Curry paradox works even for the weaker form of $\to$-Introduction. (But unlike the proof for the stronger form of $\to$-Introduction, this one assumes that Modus Ponens can be used even in hypothetical arguments; that is, this one assumes the full rule $A, A \to B \vdash B$, rather than the weaker rule used in the first proof that from $\vdash A$ and $\vdash A \to B$ we can infer $\vdash B$). The proof:

1. $True(\langle C_B \rangle) \vdash True(\langle C_B \rangle) \to B$	[Intersubstitutivity property of 'True']
2. $True(\langle C_B \rangle),\ True(\langle C_B \rangle) \to B \vdash B$	[Modus Ponens]
3. $True(\langle C_B \rangle) \vdash B$	[1,2]
4. $\vdash True(\langle C_B \rangle) \to B$	[3, $\to$-Introduction]
5. $True(\langle C_B \rangle) \to B \vdash True(\langle C_B \rangle)$	[Intersubstitutivity property of 'True']
6. $\vdash True(\langle C_B \rangle)$	[4,5]
7. $\vdash B$	[3,6]

[8] A further restriction to block the use of Modus Ponens in hypothetical proofs would block the second derivation; alternatively, someone might be tempted by the view that we should give up $\to$-Introduction but not Contraction, making only the first derivation threatening.

the only obvious arguments for Contraction presuppose one of the other two principles. For instance, Importation would get us from $A \rightarrow (A \rightarrow C)$ to $A \wedge A \rightarrow C$; from there, the further inference to $A \rightarrow C$ is totally compelling. Alternatively, we might note that from $A \rightarrow (A \rightarrow C)$ and A, we can infer C by two applications of Modus Ponens; so with a slightly stronger version of $\rightarrow$-Introduction (allowing a side premise), we can get from $A \rightarrow (A \rightarrow C)$ to $A \rightarrow C$. I think that these reflections on the assumptions underlying the obvious arguments for Contraction make it less surprising that Contraction might be given up.[9]

There is in fact a well-known logic in which Importation, $\rightarrow$-Introduction and Contraction all fail: "fuzzy logic", aka Lukasiewicz continuum-valued logic (with 1 as sole designated value). It's an extension of K_3, and it's commonly taken as a good logic for reasoning with vague concepts, so it does seem to meet the criterion of usability. Unfortunately, it will not do for naive truth theory: although it evades many of the paradoxical arguments, naive truth theory is still inconsistent in it. (See [7].)[10]

But a few years ago I began thinking seriously about the program of extending Kleene logic with a new conditional that does allow for the naive theory of truth (including substitutivity of $True(\langle A \rangle)$ with A even within the scope of the new conditional), and have obtained some positive results. The first attempt ([3]) was actually a rather mixed success: though naive truth theory is consistent in it (indeed, "strongly consistent" in it, in the sense of note 3), still the $\rightarrow$ of the logic does not reduce to the ordinary $\supset$ when excluded middle is assumed for the antecedent and consequent; this leads to some counterintuitive results. My second attempt ([5]) solved these problems: I called the logic LCC. I'm not sure it is the best possible logic for saving naive truth theory, but it seems more than adequate.

Let me clarify something: when I speak of a logic "for naive truth theory", I don't mean to suggest that we use a different logic here than elsewhere. My view is that LCC is a good candidate for our basic, all-purpose logic. Might regarding LCC as a general logic cripple our reasoning in physics, mathematics, etc.? No: we can add all instances of excluded middle in the language of pure mathematics and pure physics as non-logical axioms. We might want to resist adding excluded middle when some of the terms are vague: that's a separate issue I don't want to get into here. (Let me just say that as far as I can see, LCC handles vague notions at least as well as "fuzzy logic" does.) And of course if one thinks that certain parts of the language of basic physics or even basic mathematics (such as quantification over all ordinals) are vague, then abandoning excluded middle for vague terms will lead to abandoning it for those parts of physics and

[9] The term 'Contraction' is sometimes used for a structural rule: giving up contraction in *that* sense would involve supposing that a sentence B might be "a consequence of A taken twice" without being a consequence of A *simpliciter*. I hope it is clear that I am *not* giving up contraction as a structural rule. (Doing so would seem to me to do serious violence to any normal notion of consequence.)

[10] Restall [13] obtained an earlier result, slightly short of this but enough to show that this logic will not do for truth theory.

mathematics. But the point is that nothing in my account **requires** that it be abandoned outside of the special case of the use of 'True' in "self-referential" contexts.

I don't want to get into the details of how LCC works, but I'll say something to indicate it's general flavor. As remarked, it is an extension of the Kleene logic K_3 (including the meta-rule of disjunction elimination). In my presentation of it in [5] I used only the three semantic values of K_3, viz. 1, $\frac{1}{2}$ and 0. In such a presentation, $\rightarrow$ cannot be value-functional: the value of $A \rightarrow B$ (relative to an assignment s) is not determined by the values of A and of B (relative to s). All we get is the following table of possible values:

$\rightarrow$	$B=1$	$B=\frac{1}{2}$	$B=0$
$A=1$	1	$\frac{1}{2}$,0	0
$A=\frac{1}{2}$	1	1,$\frac{1}{2}$	$\frac{1}{2}$,0
$A=0$	1	1	1

Which values we do get for given A and B is determined in the theory, but not just by the semantic values assigned to A and to B.

There is, though, a way of subdividing the value $\frac{1}{2}$ into infinitely many distinct semantic values, in a way that will make all the connectives value-functional; indeed, the consistency proof offered in [5] could be rewritten in these terms. To describe the space of such "fine-grained semantic values", consider any initial ordinal Π greater than ω, and let F^{Π} be the set of functions from $\{\alpha|\alpha < \Pi\}$ into $\{1, \frac{1}{2}, 0\}$. Call a member f of F^{Π} **cyclic** if there is a ρ_f such that $0 < \rho_f < \Pi$ and for all β and σ, $f(\rho_f \cdot \beta + \sigma) = f(\rho_f + \sigma)$ (when these are defined, i.e. when $\rho_f \cdot \beta + \sigma < \Pi$). Call a member f of F^{Π} **regular** if (i) it is cyclic and (ii) either it is a constant function or else $f(0) = \frac{1}{2}$. (Note that given (i), (ii) implies that if f is not constantly 1 or constantly 0 then for some ρ_f in the interval $(0, \Pi)$, f assumes the value $\frac{1}{2}$ at each right-multiple of ρ_f. It is this consequence that gives (ii) its importance.) Let V^{Π} be the set of regular members of F^{Π}. For a sufficiently large Π, V^{Π} will serve as the space of semantic values.

V^{Π} has a natural partial ordering: $f \preceq g$ if and only if $(\forall\alpha < \Pi)(f(\alpha) \leq g(\alpha))$. The partial order has a largest element $\mathsf{1}$, viz. the function that assigns everything the value 1, and a smallest element $\mathsf{0}$, viz. the function that assigns everything 0. And the partial order is symmetric around a middle point $\mathsf{\frac{1}{2}}$, the function that assigns everything value $\frac{1}{2}$; this will be the value of the Liar sentence. (If f has both the values 0 and 1 in its range it will be incomparable with $\mathsf{\frac{1}{2}}$; if its range includes $\frac{1}{2}$ and 1 but not 0 it will be strictly between $\mathsf{\frac{1}{2}}$ and $\mathsf{1}$; and if it includes $\frac{1}{2}$ and 0 but not 1 it will be strictly between $\mathsf{\frac{1}{2}}$ and $\mathsf{0}$. The earlier 3-valued semantics eliminated any distinctions among values that weren't $\mathsf{1}$ and $\mathsf{0}$, which is why the conditional could not be represented value-functionally within it.)

For each of the sentential connectives I will now describe an operation on V^{Π}. The operation f corresponding to conjunction is simply pointwise minimum: $(f \;\mathsf{f}\; g)(\alpha)$ is $min\{f(\alpha), g(\alpha)\}$. Similarly for disjunction: $(f \;\mathsf{g}\; g)(\alpha)$ is

$max\{f(\alpha), g(\alpha)\}$. It needs to be verified that these are regular if f and g are, but that's easy: for cyclicity, simply let $\rho_{f\wedge g}$ and $\rho_{f\vee g}$ be a non-zero common right-multiple of ρ_f and ρ_g that precedes Π;[11] and the satisfaction of requirement (ii) on regularity is obvious. Negation is also handled pointwise: $(f\,\mathsf{F})(\alpha)$ is $1 - f(\alpha)$. Here the preservation of regularity is even more evident, for $\rho_{f\mathsf{F}}$ can be taken to simply be ρ_f. Far more interesting is the operation $\Longrightarrow$ corresponding to the conditional; $(f \Longrightarrow g)(0)$ will be

1 if $(\exists\beta < \Pi)(\forall\gamma \in [\beta, \Pi))f(\gamma) \leq g(\gamma)$;
0 if $(\exists\beta < \Pi)(\forall\gamma \in [\beta, \Pi))f(\gamma) > g(\gamma)$;
$\frac{1}{2}$ otherwise;

and when $\alpha > 0$, $(f \Longrightarrow g)(\alpha)$ will be

1 if $(\exists\beta < \alpha)(\forall\gamma \in [\beta, \alpha))f(\gamma) \leq g(\gamma)$;
0 if $(\exists\beta < \alpha)(\forall\gamma \in [\beta, \alpha))f(\gamma) > g(\gamma)$;
$\frac{1}{2}$ otherwise.

This too preserves regularity: this time, take $\rho_{f \Longrightarrow g}$ to be of form $\gamma \cdot \omega$, where γ is a common right-multiple of ρ_f and ρ_g that precedes Π. (We could use the least such $\gamma \cdot \omega$ in place of Π in the clause for $(f \Longrightarrow g)(0)$.)

It is easily seen that the space V^Π is a deMorgan algebra with respect to the operations f, g and F; this implies that if validity is defined in terms of preserving value $\mathsf{1}$ in all valuations in this space, we obtain at least the sentential part of the logic K_3, in the narrow sense of K_3 that does not include the disjunction-elimination rule. The disjunction elimination rule requires an additional fact, that $\mathsf{1}$ be "join-irreducible", i.e. that there be no values f and g for which $f\,\mathsf{g}\,g = \mathsf{1}$ even though neither of f and g is $\mathsf{1}$. But we have that too, by the regularity requirement: for if $f\,\mathsf{g}\,g = \mathsf{1}$ and neither f nor g is $\mathsf{1}$, then neither f nor g is $\mathsf{0}$ either, and so $f(0) = g(0) = \frac{1}{2}$; so $(f\,\mathsf{g}\,g)(0) = \frac{1}{2}$, so $f\,\mathsf{g}\,g \neq \mathsf{1}$.

The other important feature of this algebra is that it is complete with respect to cardinalities smaller than Π: that is, for any subset S of V^Π with cardinality less than that of Π, the functions $\min(S)$ and $\max(S)$ that give their pointwise minimum and maximum are in V^Π. (To verify the cyclicity of these functions, simply let ρ_S be the smallest non-zero common right-multiple of all the ρ_f for f in S; it must be less than Π, by the cardinality restriction on S.) Because of this completeness property of the algebra, there will be no problem of treating quantifiers in a model whose domain has cardinality less than Π. (We'll need to use a larger space of semantic values for dealing with models of large cardinality than for dealing with models of small cardinality, but this creates no problems.) In fact, $\max(S)$ is $\mathsf{1}$ if and only if $\mathsf{1} \in S$; so we get not only the K_3-valid inferences involving $\exists$, but also the meta-rule of $\exists$-elimination.

The point of a valuation space V^Π is to form V^Π-valued models. A V^Π-**valued model** W for a language (without function symbols) consists of a domain U of cardinality smaller than Π, an assignment of an object in U to each individual constant, and an assignment, to each n-ary predicate p of the language, of a function p_W from U^n to V^Π; given such a model, we obtain a value $|||A|||$ in V^Π

[11] There is such a predecessor of Π, given that Π is an initial ordinal.

for any sentence A of the language (and for any formula of the language relative to any assignment of objects in U to the free vatiables), by using the operations described above. (I use the triple bars to indicate these fine-grained values, in contrast to the double bars used earlier for the coarse-grained values in $\{0, \frac{1}{2}, 1\}$.) A V^{Π}-valued model **treats** p **classically** if it assigns it a function whose range is a subset of $\{0, 1\}$. For each V^{Π}-valued model W of a language, there is a corresponding classical model W^{-} **on the** $\rightarrow$**-free sublanguage built from the atomic predicates that the** V^{Π}**-valued model treats classically**:[12] the classical model and the V^{Π}-valued model have the same domains, and $\langle u_1, ..., u_n \rangle$ is in the extension of p in W^{-} when $p_W(u_1, ..., u_n)$ is 1, and fails to be in the extension of p in W^{-} when $p_W(u_1, ..., u_n)$ is 0. I'll call W^{-} the **classical reduct** of W.

What do we do with these V^{Π}-valued models? Suppose we are given any decent classical model M for a base language L without 'True' or '$\rightarrow$'; where by a **decent** classical model, I simply mean one that contains within it a standard model of arithmetic. (The reason for so restricting is so that the model will contain within it a standard model for the syntax of the language L^{+} obtained by adding 'True' and '$\rightarrow$' to L; that seems a minimal prerequisite to even raising the question of the semantics of 'True'.) What we want is to be able to find a Π and a V^{Π}-valued model M^{+} of the full language L^{+} such that

(i) M^{+} has M as its classical reduct (so that M^{+} has the same domain as M and in effect gives the same classical extension to predicates of L that M gives them)

(ii) M^{+} validates the naive theory of truth.

For (ii), the main requirement is

(iia) for any sentence A in the full language, $True_{M^{+}}(u) = |||A|||$ (and therefore $|||True(\langle A \rangle)||| = |||A|||$).

In addition, we require

(iib) for any u in $dom(M)$ $(= dom(M^{+}))$ that is not a sentence of L^{+}, $True_{M^{+}}(u)$ is 0.

Then the import of the Fundamental Theorem of [5] is that there is a way of obtaining such an M^{+}, for any decent classical model M.[13]

I take this to be an adequate resolution of the paradoxes of truth, in a non-dialetheic logic.

---

[12] Indeed, there's really no need to go to the $\rightarrow$-free part of the language; we could instead interpret the $\rightarrow$ as $\supset$ in the classical valuation.

[13] To immunize against a confusion, I should point out that while the sequence of values $|A|_{\alpha}$ discussed in that paper is related to the semantic value f_A of A discussed here, they are not the same. For typically, the function assigning $|A|_{\alpha}$ to α will not have the regularity requirements needed for inclusion in V^{Π}. What we do have (by the Fundamental Theorem of the other paper) is that there is a unique member f of V^{Π} and an ordinal β less than Π such that $(\forall\alpha \geq \beta)(|A|_{\alpha} = f(\alpha))$, and that f is what I'm here taking to be the semantic value of A.

5 Other Paradoxes

How about semantic paradoxes involving notions other than truth? Not a problem: the construction generalizes straightforwardly to satisfaction; and other semantic notions (denotation, definability) are explainable from that.[14]

We also get a consistent theory of properties and (non-extensional) relations with naive comprehension in biconditional form. (Instead of 'properties and relations' I'll just say relations, since properties can be conceived as 1-place relations.) This is most naturally shown in analogy with the semantic case, by starting with a given domain for a given language L (adequate to talking about natural numbers and finite sequences), and going to a larger language with the binary predicate '$Rel(z,n)$' (meaning "z is an n-place relation") and the binary predicate '$\Delta(s,z)$' (meaning "for some n, z is an n-place relation and s is an n-place sequence that instantiates z"). Then we inductively expand the domain by adding new entities: for each formula and choice of a finite set of distinguished variables (say $\Theta(x_1,...,x_n,u_1,...,u_k)$, with the x_is distinguished), and any entities $o_1,...,o_k$ that are either in the ground model or have previously been added, we add a new entity $\lambda x_1,...,x_n\Theta(x_1,...,x_n,o_1,...,o_k)$. Given this background, we can construct a valuation for the instantiation predicate Δ which validates the naive comprehension schema

$$\forall u_1...\forall u_k\exists z[Rel(n,z)\wedge\forall x_1...\forall x_n[\langle x_1...\forall x_n\rangle\Delta z\leftrightarrow\Theta(x_1,...,x_n,u_1,...,u_k)]],$$

by proceeding in complete analogy with the treatment of satisfaction. (Call this the autonomous approach; it is set out in more detail in [4].) Alternatively, one can reduce the property case to the semantic case, by modelling n-place relations

[14] How does this fit with the Appendix to Chapter 1 of [11], which appears to show that the Berry paradox (for the naive theory of denotation) arises without assuming excluded middle? The answer is that the proof there uses a principle about the least number operator μ that is tantamount to excluded middle or to a restricted version of it: viz.,

(*) $\exists xA(x)\rightarrow\exists y[y=\mu xA(x)]$

(where the quantification is over numbers, and where $A(x)$ is allowed to contain vocabulary that leads to breakdowns of excluded middle). To see that (*) is really a (possibly restricted) form of excluded middle, let B be any sentence, and let $A(x)$ be

$[x=1]\vee[x=\mathbf{0}\wedge B]$.

Then $\vdash A(1)$; so $\vdash\exists xA(x)$ and $\vdash\forall y[y=\mu xA(x)\rightarrow(y=\mathbf{0}\vee y=1)]$, and so (*) implies

(**) $[\mathbf{0}=\mu xA(x)]\vee[1=\mu xA(x)]$.

At this point the discussion divides. (I) On the most natural reading of the least number operator, $\mathbf{0}=\mu xA(x)$ is equivalent to $A(\mathbf{0})$ and hence to B, and $1=\mu xA(x)$ is equivalent to $A(1)\wedge\neg A(\mathbf{0})$ and hence to $\neg B$, so (**) is in effect $B\vee\neg B$, and so (*) implies a general form of excluded middle. (II) There is also an alternative construal of the least number operator, on which $\mathbf{0}=\mu xA(x)$ is equivalent to $Det[A(\mathbf{0})]$ and hence to $Det[B]$, where Det is an operator meaning "it is determinately the case that", and $1=\mu xA(x)$ is equivalent to $Det[A(1)]\wedge\neg Det[A(\mathbf{0})]$ and hence to $\neg Det[B]$; in that case, (**) is in effect $Det[B]\vee\neg Det[B]$, so (*) implies only excluded middle restricted to determinateness claims. But that restricted excluded middle is enough to breed paradox all by itself, without the least number operator, via sentences that assert of themselves that they are not determinately true. Any treatment of the paradoxes of truth that accords with the naive truth schema will thus have to avoid excluded middle for determinateness claims, and so again cannot accept (*). (The semantics outlined in this paper does allow for determinateness operators in the language, but these operators do not obey excluded middle: see sections 5 and 6 of [5], and sections 6-8 of [6].)

as "objectified formulas", that is, as pairs $\langle A, s\rangle$ where A is a formula that may contain the satisfaction predicate and s is a function that assigns objects to all variables in A other than the particular variables $v_1, ..., v_n$. Then $\langle o_1, ..., o_k\rangle$ instantiates $\langle A, s\rangle$ iff "the combination of $\langle o_1, ..., o_k\rangle$ with s" satisfies A.[15] In either case, the model is one where all properties and relations are definable from the ground model, but that is just for getting a consistency proof; not all models of the naive theory of relations will have this form (as is completely evident on the autonomous approach).

Of course, excluded middle cannot be assumed for instantiation claims generally: for instance, if R is the property of not instantiating itself (on the reductive approach this would be $\langle$'$\neg Sat(x_1, x_1)$', $\varnothing\rangle$, where $\varnothing$ is the null assignment function), then the assumption that R either instantiates itself or doesn't leads to contradiction. All we can say in general is that '$\langle o_1, ..., o_k\rangle$ instantiates $\langle A, s\rangle$' will get a semantic value in the space V^{Π} (and that it gets precisely the same value as A gets with respect to the combination of $\langle o_1, ..., o_k\rangle$ with s). In contrast, the modelling just given shows that we can consistently assume of anything that it either is a relation or isn't, and assume of any two relations that they either are the same or they aren't: we can give a value in $\{0, 1\}$ to any claim not involving the instantiation relation, even if it is about properties and relations.

Can we also get a consistent theory of extensions (relations-in-extension, including classes as the $n = 1$ case), with naive comprehension in biconditional form? I'm not sure: there do seem to be complications in modifying the construction so as to ensure extensionality, while also getting the other desired laws

[15] This reductive approach will not work as it stands if the naive theory of relations explicitly asserts that relations are not linguistic–or rather, that they are not pairs whose first member is linguistic. We could handle this by adding "duplicates" of the objectified formulas to the ground model, and use these duplicates as the properties and relations in the model (still allowing them to apply to themselves, of course); though it is probably easier to take the autonomous approach.

I haven't thought about whether there is any problem getting the naive semantics and the naive theory of relations together, when relations are explicitly declared non-linguistic. I doubt that there is a problem, but for here I restrict myself to the claim that each of these theories is individually obtainable.

such as substitutivity rules for identity.[16] But even if we can't, I'm not sure that this is particularly worrisome. In the case of sets we have a perfectly good non-naive theory, the theory of iterative sets (e.g. Zermelo-Fraenkel set theory), and I do not see any obvious reason to demand more. This theory has no analog that is satisfactory for the semantic paradoxes (since the analog of the fact that the Russell set doesn't exist would have to be that the predicate 'is not true of itself' doesn't exist, which is absurd). Nor does it have a satisfactory analog for properties in the sense of the term that has application in semantics, for the point of having properties in that sense demands that there be a property corresponding to every predicate. (Similarly for relations more generally.) It is because iterative set theory has no satisfactory analog in these cases that the naive theories of truth and satisfaction and of properties and relations are so important. If there is an analogous need of a naive theory of sets (or extensional relations more generally), it is quite unobvious what it is.

I should add that the naive theory of non-extensional relations could be used to make iterative set-theory more attractive: we can use naive properties and relations for most of the purposes that "proper classes" have traditionally been put, for the extensionality of proper classes plays little role. Thus we avoid well-known puzzles about "how the proper classes differ from another level of sets". This idea of proper classes as entities of a very different nature than iterative sets was well articulated by Parsons in [10]; the view being suggested in this paragraph is a slight variant, in which we allow proper class-surrogates to "belong to" proper class-surrogates, and indeed accept naive comprehension for them.[17] Moreover, though the law of excluded middle can't be assumed

[16] It's worth mentioning that to have any hope of getting

(EXT) $\forall s(s \in y \leftrightarrow s \in z) \leftrightarrow y = z$,

we must abandon the assumption of excluded middle for identity claims. To see why, let K be the "Curry set" $\{w | w \in w \rightarrow \bot\}$; let o_z be $\{w | w \neq w\}$, and o_y be $\{w | w = w \wedge K \in K\}$. By (EXT), $o_y = o_z \leftrightarrow K \in K \rightarrow \bot$; so by the definition of K and naive comprehension, $o_y = o_z \leftrightarrow K \in K$. But on the semantics in the text, excluded middle cannot be assumed for $K \in K$. (Note that if the Russell set were used in place of the Curry set, we would not have a counterexample to excluded middle: the claim corresponding to $o_y = o_z$ would then have value 0.) The situation is in marked contrast to the non-extensional case, where excluded middle for identity claims was unproblematic.

Moreover, while we can hope to get transitivity and substitutivity in the form of pairs of rules (e.g. for transitivity, the rules $x = y \models y = z \rightarrow x = z$ and $\neg(y = z \rightarrow x = z) \models x \neq y$, the conditional forms $x = y \wedge y = z \rightarrow x = z$ and $x = y \rightarrow (y = z \rightarrow x = z)$ inevitably fail if we adhere to (EXT). (Let o_y and o_z be as above, and let o_x be $\{w | w = w\}$.) But this really isn't surprising: a failure of those forms of transitivity seems pretty much inevitable when "indeterminate identity" (failure of excluded middle for identity claims) is allowed.

The real problem with adhering to any form of extentionality is securing the validity of the rule $\forall w(w \in x \leftrightarrow w \in y) \models x \in z \leftrightarrow y \in z$, a rule which is independent of how identity is treated. (If this can be secured while retaining comprehension, then to retain substitutivity rules for identity we might need to weaken (EXT) a bit, e.g. by replacing the left to right conditional by the two rules $\forall s(s \in y \leftrightarrow s \in z) \models y = z$ and $\neg(y = z) \models \neg\forall s(s \in y \leftrightarrow s \in z)$.)

[17] The idea of a theory that abandons excluded middle to allow proper classes to belong to themselves has also been suggested previously, e.g. in [9]. But that was in the context of a logic that does not contain a reasonable conditional and does not allow naive comprehension in biconditional form, which severely limits the utility of the theory. For instance, without

generally, it can be assumed for sentences in which quantifiers are suitably restricted, e.g. to the iterative sets, or to those and the proper class-surrogates that apply only to iterative sets, or to various larger subuniverses. Excluded middle is only abandoned in connection with certain things (such as class of all classes that don't belong to themselves) which don't exist in any of the classical theories. I don't claim that this view of proper classes has huge advantages over Parsons', though I do think it has some;[18] my main point here is that with any view that postulates class-like entities with a fundamentally different character from iterative sets, an awkwardness in the iterative theory is removed, thus undermining one argument one might have had for a naive account of extensional entities like sets.

My claim then is that we have a unified account of all the paradoxes that are really in the same ballpark as the paradoxes of truth, in a non-paraconsistent logic.

6 Dialetheic variants

Can the dialetheist do as well? I'm not sure. An obvious thought is that just as the dialetheist can "dualize" K_3 to obtain LP, so too he can dualize LCC to obtain "Dual LCC". Dual LCC has just the same semantics as LCC, but in the 3-valued formulation it's designated values are 1 and $\frac{1}{2}$ rather than just 1, and in the infinite-valued (fine-grained) formulation its designated values are all values other than $\mathbf{0}$.

However, Dual LCC has the same difficulty as LP: it does not validate Modus Ponens. The exceptions to Modus Ponens will be somewhat fewer than in the case of LP, since in the semantics of Dual LCC $A \rightarrow B$ sometimes takes the undesignated value 0 when A has the designated value $\frac{1}{2}$ and B has the undesignated value 0; but sometimes $A \rightarrow B$ has value $\frac{1}{2}$ when A has $\frac{1}{2}$ and B has 0, and that's enough to ensure a violation of Modus Ponens in Dual LCC. An example is the standard Curry sentence $C_\perp$, provably equivalent to $True(\langle C_\perp \rangle) \rightarrow \perp$ and hence to $C_\perp \rightarrow \perp$ in naive truth theory; where $\perp$ is any sentence with semantic value $\mathbf{0}$. So $|||C_\perp|||$ is the same as $|||C_\perp \rightarrow \perp|||$. This value is not $\mathbf{0}$ (in fact, its value is the function that assigns $\frac{1}{2}$ to 0 and to limits,

that, one can't assert that each class x is a subclass of itself (in the sense that for all z, if z belongs to x then z belongs to x); nor can one assert that if x belongs to the class of those z such that $\Phi(z)$, then $\Phi(x)$.

[18] Parsons' own view didn't allow impredicatively defined classes of sets; the alternative suggested here does, and I do take that to be a fairly clear advantage. Parsons could of course have allowed impredicatively defined classes without allowing proper classes as members. I believe he didn't because it would have made classes look too much like "just another level of sets". But on my view that worry doesn't arise: the "classes" are just properties, and they obey completely different laws: for instance, self-instantiation is allowed.

Parsons view might be thought to have an advantage over the one suggested here, in that he can define identity for classes so that the axiom of extensionality holds. But in fact I could adopt the same definition of "identity": it's just that outside of Parson's restriction to classes of sets, the defined notion doesn't behave in the way one would expect a definition of identity to behave.

0 to odd ordinals, and 1 to even successors.) But then the inference from $C_\perp$ and $C_\perp \to \perp$ to $\perp$ is a counterexample to modus ponens.[19]

We can avoid this problem by shifting to "Almost-Dual LCC", where the designated values are those f such that $\frac{1}{2} \preceq f$; in other words, those functions that after a certain point never contain the value 0. Almost-Dual LCC does validate Modus Ponens (Modus Ponens for $\to$, not for $\supset$). But it has another defect: the disjunction $C_\perp \vee \neg C_\perp$ is designated even though neither $C_\perp$ nor $\neg C_\perp$ is; as a result, this logic would lead to a solution of the Curry paradox with a supervaluationist flavor, in that you could assert that the Curry sentence is either true or false but wouldn't be allowed to say which. Relatedly, we would have a failure of disjunction elimination: $C_\perp$ implies $\neg C_\perp$ (since it implies $\perp$, given modus ponens), and $\neg C_\perp$ implies $\neg C_\perp$, but $C_\perp \vee \neg C_\perp$ doesn't imply $\neg C_\perp$.

To avoid these problems, one would need to find an acceptable set of designated values which either excludes $C_\perp \vee \neg C_\perp$ or includes $\neg C_\perp$. (It can't include $C_\perp$ if it is closed under modus ponens.) But it can't exclude $C_\perp \vee \neg C_\perp$ if it is to include the Liar sentence, since the semantic value of $C_\perp \vee \neg C_\perp$ is strictly bigger than that of the Liar. Indeed, given the reasonable additional demand that the theory validate $\wedge$-Introduction, then the set of designated values can't exclude the value of $C_\perp \vee \neg C_\perp$ as long as it contains **any** pair of form $\{f, f\,\mathsf{F}\}$: for if f and $f\,\mathsf{F}$ are designated, then $\wedge$-Introduction requires that $f \;\mathsf{f}\; f\,\mathsf{F}$ be designated too; but $f \;\mathsf{f}\; f\,\mathsf{F} \preceq \frac{1}{2} \preceq [C_\perp \vee \neg C_\perp]$, so the Liar sentence and $C_\perp \vee \neg C_\perp$ must receive designated values too.

In short, the only alternative, for a dialetheist who wants a theory based on the algebra V and that validates reasonable rules, is to include the value of $\neg C_\perp$ but not that of $C_\perp$ among the designated values. But the semantic value of $\neg C_\perp$ is so similar in structure to that of $C_\perp$ (in both, 0 and 1 alternate at successors, with $\frac{1}{2}$ at all limits) that it is hard to see how there can be a natural

[19] Despite the inadequacy of Dual LCC, it may be interesting to contemplate: in particular, I think it helps undermine the general worry about dialetheism contemplated at the end of Section 2. The worry was that if the acceptance of $\neg A$ doesn't suffice for the rejection of A, how are arguments for the rejection of A to be conducted? Dual LCC is relevant to this because it contains a sequence of stronger and stronger "negation-like" operators. Let DA abbreviate $A \wedge (\top \to A)$, where $\top$ is any sentence of form $B \to B$; and let N_0A be $\neg A$, and for each k let $N_{k+1}A$ be DN_kA. (We could extend this a long way into the transfinite, by using the truth predicate to get the effect of infinite conjunctions at limits.) For each k–or each ordinal α, on the transfinite extension– $|||N_\alpha A \wedge A|||$ will have a value other than $\mathsf{0}$ when A has as its value a function in which $\frac{1}{2}$ and α-length sequences of 0's both appear arbitrarily late. So there is no α for which accepting $N_\alpha A$ quite suffices for rejecting A. Nonetheless, it requires very special assumptions about A to accept even $N_1A \wedge A$; accepting N_1A thus makes it **harder** not to reject A than does the mere acceptance of $\neg A$. And as α increases, it becomes harder and harder not to reject A while accepting $N_\alpha A$. Indeed for nearly any choice of semantic value that A might have, it is possible to find a sufficiently big $\sigma[A]$ such that $|||N_{\sigma[A]}A \wedge A|||$ is 0. My thought is, then, that this battery of stronger and stronger negations might serve the dialetheist's needs in arguing for rejecting undesirable claims.

If it were possible to "conjoin all the N_αs" into a "supernegation operator" N for which accepting NA compelled rejecting A, this would lead to the objection that we have made dialetheism uninteresting: perhaps N is what we should have called negation (and the truth of NA is what we should have called the falsity of A), and it's only because we've made an alternative choice that the logic is "dialetheic". But whatever the merits of this objection, it doesn't arise if it is impossible to conjoin all the N_αs; and that **is** impossible in Dual LCC.

choice for the set of designated values that includes $\neg C_\perp$ but not $C_\perp$.

This is less than an impossibility proof: technically, the question is whether the algebra contains any prime filters that are both closed under "modus ponens for $\Longrightarrow$ " and contain some sentence and its negation, and I haven't investigated this beyond the observations just made.

But even if one could do better than Dual LCC and Almost-Dual LCC, I'm not sure I see the point. In the case of LP and K_3, there was at least a *prima facie* advantage of the dialetheic logic over the non-dialetheic (inadequate though both were): despite its inadequacies from lack of a decent conditional, LP gave the full naive theory of truth, whereas K_3 didn't. But in the present case, we already get the naive theory of truth with the non-dialetheic logic, so there seems to be no special motivation for going dialetheic. The main case for dialetheism has disappeared.[20]

References

[1] Solomon Feferman. Toward useful type-free theories, I. *Journal of Symbolic Logic*, 49:75–111, 1984.

[2] Hartry Field. *Truth and the Absence of Fact*. Oxford University Press, Oxford, 2001.

[3] Hartry Field. Saving the truth schema from paradox. *Journal of Philosophical Logic*, 31:1–27, 2002.

[4] Hartry Field. The consistency of the naive(?) theory of properties. In Godehard Link, editor, *One Hundred Years of Russell's Paradox - Logical Philosophy Then and Now*. Walter de Gruyter, 2003.

[5] Hartry Field. A revenge-immune solution to the semantic paradoxes. *Journal of Philosophical Logic*, 32, 2003.

[6] Hartry Field. The semantic paradoxes and the paradoxes of vagueness. In JC Beall, editor, *Liars and Heaps*. Oxford University Press, 2003.

[7] Petr Hajek, Jeff Paris, and John Shepherdson. The liar paradox and fuzzy logic. *The Journal of Symbolic Logic*, 65:339–346, 2000.

[8] Saul Kripke. Outline of a theory of truth. *Journal of Philosophy*, 72:690–716, 1975.

[20] Thanks to JC Beall and Graham Priest for useful comments on earlier drafts.

And lest the epigram of this paper mislead, I'd like to say that Graham Priest was extremely encouraging about my earliest attempt at a non-dialetheic solution to the semantic paradoxes; I might not have pursued the matter were it not for this encouragement, and I have found his open-minded attitude in discussing these topics admirable. So I'm not calling for the remedy that Henry II was calling for!

[9] Penelope Maddy. Proper classes. *Journal of Symbolic Logic*, 48:113–39, 1983.

[10] Charles Parsons. Sets and classes. In Charles Parsons, editor, *Mathematics in Philosophy*. Cornell University, 1983.

[11] Graham Priest. *In Contradiction*. Martinus Nijhoff, Dordrecht, 1987.

[12] Graham Priest. What is so bad about contradictions? *Journal of Philosophy*, 95:410–426, 1998.

[13] Greg Restall. Arithmetic and truth in Łukasiewicz's infinitely valued logic. *Logique et Analyse*, 139–140:303–312, 1992.