International Journal of Undergraduate Research and Creative Activities

Volume 3
Issue 2 *Res Cogitans 2011*

Article 10

October 2023

Defending a Dialetheist Response to the Liar's Paradox

James Finley
info@ubiquitypress.com

Defending a Dialetheist Response to the Liar's Paradox

Abstract

The Liar's paradox stands as one of the longest standing and most discussed problems in philosophical history. In this paper, I first briefly set up the requirements on a language needed for that language to have the expressive power to construct a Liar sentence and show how the Liar sentence leads to the inconsistency and triviality of said language (if it is a logically classical language). I will then quickly set up and reject responses to the Liar that keep a classical logic as the model logic for natural language (which I will take to be English here), and argue for a dialetheist response to the Liar which endorses a three-valued, para-consistent logic. A dialetheist view claims that some pairs of sentences and the negation of that very same sentence are true, and a para-consistent logic is any logic in which ex contradictione quodlibet (The inference from P and ~P to anything) fails. Such a solution, I think, provides a viable route to defusing the Liar and keeping the expressive power of natural language.

Res Cogitans (2011) 2:148–156 2155-4838 / commons.pacificu.edu/rescogitans

Defending a Dialetheist Response to the Liar's Paradox

James Finley
Williams College

Published online: 30 July 2011

Abstract

The Liar's paradox stands as one of the longest standing and most discussed problems in philosophical history. In this paper, I first briefly set up the requirements on a language needed for that language to have the expressive power to construct a Liar sentence and show how the Liar sentence leads to the inconsistency and triviality of said language (if it is a logically classical language). I will then quickly set up and reject responses to the Liar that keep a classical logic as the model logic for natural language (which I will take to be English here), and argue for a dialetheist response to the Liar which endorses a three-valued, para-consistent logic. A dialetheist view claims that some pairs of sentences and the negation of that very same sentence are true, and a para-consistent logic is any logic in which *ex contradictione quodlibet* (The inference from *P* and ~*P* to anything) fails. Such a solution, I think, provides a viable route to defusing the Liar and keeping the expressive power of natural language.

Few philosophical problems have more literature surrounding them than the Liar's Paradox, often clouding the issues at hand and making the paradoxical argument itself seem impenetrable without a wide background knowledge. Constructing an instance of the paradox simply, we can take an example Liar sentence to be a sentence of the form:

(Q): (Q) is not true

where **Q** is the name of the very sentence that declares its own untruth. More generally, we might characterize a Liar sentence as any sentence which entails, '$P \leftrightarrow$ ~True(P),' which **Q** is a clear example of.[1] In this paper, I will first briefly set up the requirements on a language needed for that language to have the expressive power to construct a Liar sentence and show how the Liar sentence leads to the inconsistency and triviality of said language (if it is a logically classical language). I will then quickly set up and reject responses to the Liar that keep a classical logic as the model logic for natural language (which I will take to be English here), and argue for a dialetheist response to the Liar that endorses a three-valued, para-consistent logic. For our purposes now, we can take dialetheism to be the view that some pairs of sentences and the negation of that

[1] That is if our truth predicate fulfills all instances of the T-schema. Details on this requirement will come up later.

very same sentence are true, and a para-consistent logic to be any logic in which *ex contradictione quodlibet* (The inference from *P* and ~*P* to anything) fails.

I. Constructing an Instance of the Liar's Paradox

In order to construct a Liar sentence a language must meet certain basic requirements: (**A**) it must have some form of self-reference, (**B**) a truth predicate meeting certain intuitive stipulations, and (**C**) a classical negation operator on sentences. Intuitively, (**A**) seems simple enough: the sentence, 'This sentence contains five words,' seems both a well-formed sentence and true. Formally, having a self-reference requirement means that we must have names for sentences in the language and that these names can be a part of the very sentence they refer to, as is the case with **Q** constructed above.[2] As for (**B**), the sentence must contain a truth predicate, which we can naively think of as referring to a property of sentences.[3] Informally, we often say that certain things people say are true (or not), everything someone says is true, a theory is true, if what that person said was true then I agree, etc. The intuitive truth predicate then is a predicate which applies to sentences. While not necessary for the construction of the Liar sentence itself, in order for inconsistency to be derived from the Liar it is important that this truth predicate fulfills all instances of the T-schema. The T-schema claims that a truth predicate is satisfied by a sentence if, and only if, that sentence has a semantic value of one (alternatively, that sentence is the case).[4] In other words, if a sentence is true in a language then it satisfies the truth predicate, which is intuitive enough.[5] Last, in order to meet (**C**) we need a classical negation operator which we can just take to be the usual logical connective here, naively thinking of it of as a function which maps sentences onto truth values. This functions almost exactly like the phrase, 'it is not the case that,' preceding a sentence in English; If a sentence is true then its negation is false, and vice versa. In classical logics, this negation operator is both exhaustive and

[2] It is actually enough if the language can express its own syntactic properties (equivalently, that basic arithmetic can be encoded in the language). One can then prove that for any property expressible in the language there exists a materially equivalent sentence the attributes that property to itself. This result is typically called the Godel-Tarski Diagonalization (or Fixed Point) Lemma. I do not have time to go over the proof in this paper, but it should just be noticed that the requirement is actually weaker than having names for sentences that can appear in the sentence they refer to.

[3] If one would rather think of truth as a property of propositions or some other truth bearer, that is fine. I do not think anything I say here will be affected.

[4] Formally, 'True(<A>) ↔ A,' where the brackets serve as a substitutional device in which one substitutes in the whole sentence (in quotations) in order to mention the sentence and not use it.

[5] This is not, strictly speaking, correct, but it is enough to get us off the ground here. I will go into detail more on arguments for the truth-predicate meeting this requirement later, as well as look at responses that deny that the truth-predicate does in fact meet these requirements.

https://digitalcommons.cwu.edu/ijurca/vol3/iss2/10
DOI: -

exclusive. It is exhaustive because for any sentence, *P*, either *P* or the negation of *P* is true. It is exclusive because for any sentence, *P*, it is not the case that both *P* and the negation of *P* are true. These two properties of the negation operator can be represented by the logical laws of excluded middle and non-contradiction, which we might formally represent as schema of the form:

Law of Excluded Middle (LEM): *P* or ~*P*
Law of Non-Contradiction (LNC): ~(*P* and ~*P*)

In a classical logic these are equivalent, but it will be useful for later discussion that we separate them here. English seemingly has all of these properties, which means we can construct a Liar sentence in English. A more colloquial example would be the sentence, 'This sentence is not true.'

From the ability to create a Liar sentence within a language and some basic properties of truth, one can conclude, given classical forms of inference, that the language is question has the properties of inconsistency and triviality.[6] A language is inconsistent if it entails both some sentence and that very same sentence's negation, and trivial if it entails all sentences. Take the Liar sentence above, **Q**. Instantiating the T-schema for **Q** gives us the sentence, 'True(Q) ↔ Q,' and then substituting the Liar sentence in for its name gives us the sentence, 'True(Q) ↔ ~True(Q).' In a classical logic, this second sentence entails a contradiction of the form, '*P* and ~*P*,' and one can conclude that **Q** is both true and false.[7] From a contradiction in a classical logic, one can prove anything they like given the following steps where *R* is any arbitrary sentence:[8]

1. *P* or ~*P* (Form of the Contradiction)
2. *P* (Simplification from 1)
3. *P* or *R* (Addition from 2, where *R* is any sentence we like)
4. ~*P* (Simplification from 1)
5. *R* (Disjunctive Syllogism from 3 and 4)

The unacceptable conclusion of the Liar's paradox is that our proof procedure in English is trivial, because one could prove any sentence from the Liar, and that the truth predicate is inconsistent, because it both is satisfied and is not satisfied by some sentences.

In order to defuse the Liar's paradox and avoid the unacceptable conclusions, one must reject that natural language meets at least one of the three requirements laid out above. It also lies open to explain why one of the seemingly unacceptable conclusions (i.e.

[6] By classical forms of inference I just mean inferences that are valid within a classical logic, which for my purposes I will just take to be the predicate logic of Frege and Russell.

[7] I will not take the space to go into the derivation here, but it is a simple enough exercise using the classical inferences from the bi-conditional, material equivalences of the conditional P → Q with ~P or Q, and simplification from, 'and.'

[8] This is just a proof of *ex falso quodlibet* in classical logics.

inconsistency or triviality) is not unacceptable after all. In this case, accepting triviality seems untenable as a trivialist would accept that it is true that their own position is unacceptable, and moreover, anything you please. We then are left with the options of either rejecting that English enjoys the properties we think it does (i.e. self-reference, truth predicate, and classical negation) or somehow accepting inconsistency without accepting triviality. I will argue for a response that takes the latter route, but we must first quickly show why responses which take the first route are untenable.

II. Defusing the Liar's Paradox

Responses which want to keep a classical logic as a model logic for natural language must either reject that English has the property of self-reference in the right type of way, or reject that English has the type of truth predicate we intuitively think it does. Rejecting that English has a classical negation operator is equivalent to rejecting classical logic. An example response which rejects that English meets the self-reference requirement is what I will call the no-proposition view. Such a view claims that sentences like Liars fail to express a proposition, despite such sentences appearing to be well-formed at first glance. They are not grounded in the right type of way due to their odd type of self-reference. After all, it seems that sentences like the Liar which refer to themselves are odd in a pathological way. When I try to pick out which sentence the Liar is talking about, I can only point out the sentence itself with the process never grounding out like it does in normal cases of reference. Take the sentence, 'What Joe said is true.' If asked what I was referring to, I could easily respond with the words Joe himself used (unless perhaps if I was Joe). With the Liar though one might claim that I could never give a satisfactory response. The claim is that the Liar cannot be constructed as a meaningful sentence, and through this the results of inconsistency and triviality are avoided.

Without getting into more details of such a position, we can reject responses that deny that English has the right type of self-reference to make the Liar sentence meaningful. Take the pair of sentences:

(N1) N2 is not true.
(N2) N1 is true.

Neither of these sentences refers to themselves, but they create a Liar type situation with all of the same unacceptable conclusions. These two sentences are much harder to claim as being ill-formed, as it seems we make statements like these all the time. 'What Joe said isn't true,' or, 'Everything Wilma said is true,' both seem like well-formed, meaningful sentences, but if Wilma said the first and Joe the second we have a similar situation to **N1** and **N2**. The no-proposition view is then unacceptable as there are some meaningful sentences one might express in English that the view must claim do not

https://digitalcommons.cwu.edu/ijurca/vol3/iss2/10
DOI: -

express meaningful sentences.[9] Such a view cannot be a model of natural language if it cannot capture all of the sentences we take to be expressible in natural language. In other words, there are extremely strong intuitions towards English having the minimum required amount of self-reference to be problematic, and the problem must be located elsewhere on pains on not modeling English any longer.

A traditional example of a classical option which rejects the truth-predicate might be taken to be Tarski's hierarchy of truth predicates.[10] For Tarski, a language cannot construct its truth predicate within itself – the proof of inconsistency from the Liar was enough for him to show this formally. A truth predicate for a language must be formulated in an expressively more powerful language, and so there is no singular truth predicate for Tarski but rather a series of stronger and stronger truth predicates that form a hierarchy. One might take the set of all sentences in English that do not contain a truth predicate (S1), and then in a stronger language with a truth predicate (S2) express the truth of sentences in S1. To express the truth of sentences in S2 though, one would need another stronger language, and so on. The Liar sentence then cannot be formed at all because there is no truth predicate that it could use that would apply to itself and be on the same level.

Tarski's solution fails as a model for natural language because the truth predicate it creates fails to accurately portray the truth predicate we use on a daily basis. We can predicate truth of sentences freely without having to worry about what level that sentence or truth predicate is on (and in fact would probably be unable to describe what level it was on if asked). Editing the Wilma and Joe example from earlier, imagine Wilma stating: 'Everything Joe says is false,' and Joe saying, 'All of Wilma's utterances are false.'[11] Each sentence seems to include the other within the range of its quantifier. Wilma is intending to include Joe's sentence that all of her statements are

[9] A proponent of the no-proposition view might of course claim that (N1) and (N2) do not express propositions either, and that the property that allows demarcation is something like non-vicious circularity or well grounded-ness. This leads to a different problem though, as then sentences like, '(Q) does not express a proposition,' which seem like fundamental claims of the no-proposition view must not express propositions either as they cannot be well-grounded. If the fundamental claims of the theory cannot be expressed in such a way that they can be accepted (or true) then the theory seems to be in trouble.

[10] Tarski is not concerned with modelling the truth predicate for natural languages but rather instead with creating a suitable truth predicate for scientific inquiry. Nevertheless we can look at his work as a potential solution, but my rejection of it is in no way finding fault with Tarski. He is unconcerned with the issues I am concerned with, and writes off natural language as inconsistent, much as I will later. Tarski, A. (1931). "The Concept of Truth in Formalized Languages." In *Logic, Semantics, Metamathematics: Papers from 1923 to 1938.* (1956) Woodger, J.H. Trans. p. 153.

[11] This is modelled on Kripke's example from, Kripke, S. (1975). "An Outline of a Theory of Truth." *The Journal of Philosophy*, 72, 19, p.691

false, and Joe is intending to do the same regarding Wilma's words. This seems perfectly intuitive and understandable. Yet on Tarski's view of truth as a hierarchal predicate, we cannot make sense of such a situation. One of the two statements would have to be of a higher level, declaring the other one to be false, and so it seems then that Tarski's truth for a language is not *our* truth. Again, the classical solution seems to be unable to express certain meaningful sentences that we can express with our intuitive truth predicate in natural language.

If one cannot reject that English has the right type of self-reference or truth predicate, it seems the problem of the Liar must then lie within the classical negation operator.[12] This in turn means that such responses must endorse a non-classical account of negation, and so a non-classical logic. Such responses either reject that negation meets **LEM** or **LNC**: solutions which deny **LEM** are called para-complete, and solutions which claim **LNC** is false in certain instances are called dialetheist.[13] I do not have space to go into para-complete solution in this paper, but those interested should look at Hartry Field's recent book, *Saving Truth from Paradox.* Any rational dialetheist solution will endorse a para-consistent logic else accept triviality, as they accept that some sentences and their negations are true, which would entail triviality if *ex contradictione quodlibet* was valid. A dialetheist is only worried about saving the system from triviality, not inconsistency. This may seem unintuitive at first, but before going into the motivations for such a position we should set up an example para-consistent logic.

A simple para-consistent logic can be set up as a three-valued logic, where instead of assigning each sentence in a language one of two semantic values (1 for true and 0 for false), we assign sentences one of three semantic values (1, 0, ½). The logic I will be setting up here is typically referred to as Logic of Paradox (LP). Sentences are assigned 1 if true, 0 if false, and ½ if they are truth gluts, which are sentences that are both true and false. We must also think of predicates as both having extensions, sets of objects they are true of, and anti-extensions, sets of objects they are false of. Predicates in LP are exhaustive as they map every object onto a truth -value, but not exclusive as they can give certain objects multiple truth-values, namely true and false. In LP, *ex falso quodlibet* fails because disjunctive syllogism is no longer a valid form of inference, given that, '*P* or *R*,' could be true if *P* was a truth glut (equivalently named as a dialetheia) and *R* was false alone, and then ~*P* would be true (because it is a dialetheia as well), but if we try to infer *R* from this we would be going down a faulty path. After

[12] I could in no way cover all classical solutions to the Liar here. Nevertheless, many responses to other classical responses will take a similar form. Self-reference and the truth predicate seem constitutive of any good model for natural language, and it seems like most classical responses give up expressive power in order to keep classical logic intact.

[13] The status of LNC in dialetheist solutions is tricky, as it will end up being both true and false for pathological sentences.

https://digitalcommons.cwu.edu/ijurca/vol3/iss2/10
DOI: -

all, it is false and false alone, and so disjunctive syllogism no longer preserves truth. Step 5 then fails in the triviality proof above.

For the dialetheist the Liar shows the inconsistency of the semantics of our natural language, but this is fine as long as inconsistency does not entail triviality. The dialetheist claims that there are no good reasons for accepting **LNC** without good argument. The main classical defense of LNC is found in Aristotle's *Metaphysics Γ*, and the arguments there have been shown to be faulty many times over.[14] We have good reason for thinking that the Liar is a truth glut, given that any argument that establishes a paradox does so from a seemingly valid and sound argument. If it did not, it would not be a paradox. Every other time one has a valid and sound argument one is supposed to believe the conclusion and has good reason to do so. In the case of the Liar's paradox, it just means I have good reason for thinking that the Liar is true, and that it is false. This is fine because I do not have good reason to think that these categories must be exclusive. One can avoid triviality but accept inconsistency, accept the intuitive conclusion of the Liar sentence (that it is both true and false), keep the T-schema's truth (although it may also be false in pathological cases), and keep the self-referential power of English.

While this may seem attractive enough to some, in order to help motivate the position a bit more I want to address a few common worries raised about the para-consistent, dialetheist position. A common worry raised for non-classical solution revolves around the possibility of revenge Liar sentences, which are specially crafted Liar sentences targeted towards the semantics of a non-classical solution. The revenge Liar for LP would look something like: Q2: 'Q2 is neither true nor both true and false.' The objector claims that the dialetheist cannot give the sentence an appropriate semantic value: if we say it is true (i.e. has a semantic value of 1) then the sentence tells us it is neither true nor a glut, and so must be false (i.e. have a semantic value of 0). If it is false, then it is either true or a glut. If it is a glut, then it is false, and perhaps the intuition is that it is false alone. LP then cannot adequately express the semantic value of the sentence, and it looks like the dialetheist is in a bind.

This kind of objection should not worry the dialetheist though, and, in fact, is exactly what she should expect. If the sentence is true (i.e. has a value 1) then it is false (i.e. has a value of 0). It then has both values and is a glut, which just means having the semantic value of 1 and 0. If it is false then it is true or a glut, and in either case it again has both the semantic value 1 and 0. If it is a glut, it already has the semantic values of 1 and 0. No matter how we approach the sentence, it seems to come out as a truth glut and so both true and false. The sentence has truth values it says it does not, as it asserts it is not a truth glut, but the dialetheist allows contradiction as long as it does not entail

[14] I turn the interested reader to Łukasiewicz, J. (1910). "On the Principle of Contradiction in Aristotle." *The Review of Metaphysics,* Vol. 24, No. 3, 1971. Wedin, Vernon, Trans. pg.489.

triviality. While this seem somewhat worrying, this situation is no different from the original Liar because it tells us that it is not true but turns out to be true, and, of course, false.

A second worry claims that if belief must be belief that *P* is true and not false, and dialetheia are both true and false, then we can never believe a dialetheia given that it is false. If this is true, then we can never believe in dialetheism as a view, given that it itself is both true and false. The dialetheist here must just deny the intuition that belief is belief in things that are true and true alone. Instead, the truth of *P* is sufficient for belief that *P*. We can believe inconsistently, and in fact people do so all the time. It even seems fine for me to say the sentence, 'I have inconsistent beliefs,' although if I were to figure out which beliefs of mine were in conflict, I would most likely revise my set of beliefs to make it consistent. Empirically, it seems that we do believe things that are false all the time, even if we do eventually revise. It is a bit harder to show that it can be rational to have inconsistent beliefs, but following Graham Priest, one can offer the paradox of the preface:[15] A philosopher writes a book and as she writes every sentence she thinks that sentence is true, else she would not write it. Yet, in the preface she writes that there is bound to be some error in her book. She believes each sentence individually but also the negation of the conjunction of all of the sentences, and so has inconsistent beliefs. In this case though, it seems rational to hold inconsistent beliefs.[16]

A final worry focuses on how one can resolve disagreement in dialetheism. Let's assume that I accept *P* (i.e. have reason to believe that *P*) and that someone disagrees with me and argues for ~*P*. Due to their persuasive rhetoric and convincing arguments, I come to have a reason to accept ~*P*. On a classical account, I cannot accept both *P* and ~*P* and so I have to weigh the arguments and decide which of the two I think is right. On a dialetheist account though, I could just accept the argument for ~*P* and believe both *P* and ~*P*, since I no longer have the consistency requirement on my beliefs. I would just be accepting *P* as a truth glut. It then looks like it might be impossible to convince someone to reject *P*, given that an argument for ~*P* does not entail that I ought not to believe *P*.

While such an objection seems strong at first, one only has to analyze how we make decisions. Classically, I am given a choice between three options: keeping my belief that *P,* rejecting it in order to accept ~*P*, or withholding assent to either. Here I must weigh my evidence and the strength of the arguments. On a dialetheist account the number of options just moves from three to four: I can keep *P* and reject ~*P*, reject *P*

[15] Priest, Graham. (2006). *In Contradiction*, Second Edition. p.100

[16] Perhaps one might also offer an argument that one should believe inconsistent things because dialetheism best resolves paradox, and paradox must be resolved, but this kind of response has a distinctive whiff of circularity.

https://digitalcommons.cwu.edu/ijurca/vol3/iss2/10
DOI: -

and accept ~*P*, or accept *P* and ~*P*, or withhold assent. Again, here I must weigh my evidence for each possibility and accept the one I believe to be the strongest supported. There is seemingly always some empirical evidence against accepting a dialetheia given that there seem to be a very small amount of them.[17] Very rarely do we run into potential dialetheia, which is supported by the fact that we take forms of inference such as disjunctive syllogism to be valid, an inference which fails in LP. If there were more dialetheia, it seems that disjunctive syllogism would obviously fail in more cases.[18] Evidence for ~*P* may not be evidence against *P*, but there is often good reason to think that it will be. While we may have more options for making a decision, we still do so based on the strength of evidence, and we must be practical about the evidence towards the existence of common dialetheia.

While I cannot cover every possible worry here, I hope that it is at least evident that dialetheism is a viable contender as a solution to the Liar's paradox. A para-consistent dialetheist solution models both the ability to express self-referential sentences and the truth predicate of natural language, as well as keeping a non-trivial proof procedure. One may have to endorse that some sentences are both true and false and so that the language is inconsistent, but this is not as high of a price as it seems if triviality does not follow. One might even think such a result is expected when it comes to natural language – after all, it is a product of human invention, and we are far from infallible, perfectly consistent creatures.

[17] Very rarely have I ever been convinced that I have good reason to believe both P and ~P. Perhaps my reason for belief that LNC holds in this case outweighs such a possibility (via inductive reasoning, or non-conceivability, or some other standard).

[18] One might extend this to LNC, but again one might think this is circular.