

Causal Theory of Truth

A Hypothesis for Treating the Truth Value of the Liar Paradox with Causal Relation

Andy Brannan

Introduction

To anyone who is not acquainted with its far reaching implications, the Liar Paradox may seem only a funny quirk of language, a kind of puzzling feature of certain types of sentences. For those who have studied the Liar, it is much more than that. The Liar Paradox is a problem in the way the truths of self-referential declarative sentences are interpreted under traditional views. Even worse, the paradox is not restricted to linguistics, but has been found to exist in first order logic, as well as mathematics. The implications of this may not at once be obvious: aberrations in natural language are one thing, fundamental flaws in the foundations of the formal systems of logic and mathematics are quite another. How can we trust proofs obtained by the use of formal systems which contain such a defect? All of the knowledge which scientific study has given us relies upon some basic assumptions about validity, yet these basic assumptions are called into question by the Liar Paradox.

The problem I have been alluding to is in the mechanics of truth. The classic form of the Liar Paradox is the declarative sentence:

This sentence is false

The above sentence cannot be true, because its claim of falsity would make it false. It cannot be false, as that would make its proposition true. Theories have been proposed which claim that that the Liar is neither true nor false, or that it is both true and false, that it is meaningless, or even that it is syntactically flawed. However, each of these options seems to present its own set of problems.

In this paper, I will discuss a number of key variations upon the Liar Paradox, and present a hypothesis which I believe may eventually lead to a new understanding of the paradox itself. I would like to present a solution proper, but I'm afraid the best I am able to achieve at present is a sort of "armed truce" with this two thousand year old enemy. I will offer a hypothetical solution, but realize that, at this time, it is only a rough sketch, and will likely require much revision in order to be viable.

My hypothesis is quite simple in essence, but requires a significant shift in thinking (at least in the Western World) to actually be applied. I view the problem with the Liar as a symptom of certain underlying principles which apply to interpreting sentences. I believe that if we choose the correct set of principles, that symptom will vanish. In order to accomplish this, one short step must be taken in our conceptualization of the idea of "truth value" in declarative sentences. First, I will argue that we must consider the truth or falsity to be an *effect* of a sentence, rather than one of its *properties*. Second, I will suggest a certain view of *causality* which is likely to yield the best results when evaluating the truth values of sentences.

Truth Defined, Part 1

In this paper, I will hold to the *transparent* conception of truth, which is to make the assertion that “it is true that *a*” and “*a*” are essentially interchangeable. More accurately, transparent truth is defined in the statement “it is true that *a*” if and only if “*a*”. Symbolically, we can say:

$$1. T([a]) \Leftrightarrow a$$

Where *T* is the global truth predicate, *a* is a proposition, and [*a*] the name of the proposition.

This rule of substitution allows for generalizations about truth values which would not be possible otherwise, and I will further generalize this conception of truth to include *deflationary* theory of truth, Tarski’s *T*-schema truth references, as well as the more recent *Catch and Release* theory. The reason for this broad definition of truth is to refer to a family of truth conceptions, any of which should be interchangeable in the context of this paper. The importance of this family of truths will become apparent in the section on Cause and Effect, below. (Beall, Spandrels of Truth, 2009) (Damjanovic & Stoljar, 2010) (Priest, 2007)

The Liar’s Club

To begin this discussion, we must consider the actual form of the Liar. The canonical Liar, it would seem, can be expressed like this:

2. Sentence (3) is true
3. Sentence (2) is false

The problem with (2) and (3) is that there are no sets of truth values for the two sentences which are consistent:

(2)	(3)	Problem with truth values
T	T	If (3) is true, (2) must be false
T	F	If (2) is true, (3) must not be false
F	T	If (3) is true, (2) must not be false
F	F	If (2) is false, (3) must not be false

This problem condition arises because, under traditional interpretation, we are presented with a proposition which attempts to affirm another proposition which in turn disaffirms the first proposition. The syntactic action involved can be expressed in the simplified form:

4. Sentence (4) is false.

We can make this simplification under the principle that any declarative sentence contains an implicit assertion of its own trueness, making the proposition in (2) unnecessary. This is the most common expression of the Liar, or at least its less formal variant: “This sentence is false.” Sentence (4) seems to be the most popular because of its simplicity – the apparent paradox is easy to detect, even if formal definition is more closely aligned with the pair of sentences (2) and (3). However, these two forms are functionally identical, and are subject to similar attempts at solution.

One family of solutions involves attempting to prove that sentence (4) is both true and false, thereby relieving the paradox. Regardless of the merits of these theories, a simple alteration of the Liar renders them impotent:

- 5. Sentence (5) is not true

Known as the Strengthened Liar, the statement “This sentence is not true” is considered the most stringent test of any theory which attempts to diffuse the paradox. (Dowden, Liar Paradox, 2010)

As previously mentioned, the Liar Paradox is not restricted to the linguistic realm. In 1903, Bertrand Russell showed that Cantor’s *Naïve Set Theory* (an informal but important theory of mathematical sets) led to a contradiction:

- 6. $let R = \{ x \mid x \notin x \}, then R \in R \Leftrightarrow R \notin R$

This symbolic statement declares “if we let R equal the set of all sets x such that x is not a member of set x , then R is a member of R if and only if R is not a member of R .” (Irvine, 2009) The same principle is clearly at work in statement (6) as in the other renderings of the Liar Paradox above.

The significance of Russell's paradox can be seen once it is realized that, using classical logic, all sentences follow from a contradiction. For example:

1	A	Premise
2	$\sim A$	Premise
3	$A \vee B$	Disjunction introduction (1)
4	B	Disjunctive syllogism (2, 3)

Set theory yields the contradictory statement $(A \wedge \sim A)$, forming premises 1 and 2. These premises can then be used to prove any arbitrary statement through a simple process involving perfectly valid rules of inference. Because set theory underlies all branches of mathematics, this fact could be used to cast doubt upon any mathematical proof. (Irvine, 2009)

This logical paradox is easily applied to statements in natural language, as in this example:

- 7. The moon is made of cheese
- 8. Statements (7) and (8) are false

Sentence (8) implicitly asserts its own truth, while explicitly asserting its own falsity, thereby creating the required contradiction ($A \wedge \sim A$). In this case, the disjunction introduction is sentence (7), and the resulting inference can be clearly seen in the resulting truth table:

(7)	(8)	Result
T	T	(8) cannot be true due to contradiction
T	F	If (8) is false, then (7) can be true
F	T	(8) cannot be true due to contradiction
F	F	(8) cannot be false due to contradiction

The only possible combination of truth values for statements (7) and (8) seem to prove that the moon is made of cheese. (Dowden, Liar Paradox, 2010)

The Russell Paradox was discovered and understood in Naïve Set Theory, an informal system of logic. In 1936, Alfred Tarski showed that in a strong formal system of logic, truth about the system cannot be defined within the system itself. Discussion of *Tarski's Undefinability Theorem* is beyond the scope of this paper. However, it is relevant to this topic in that it, and the closely related *Incompleteness Theorem* of Kurt Gödel, mark the presence of essentially the same paradox at the most fundamental levels of logic. (Priest, 2007) (Dowden, Liar Paradox, 2010)

Possible Solutions

There are four primary families of solutions to the Paradox of the Liar.

- A. The liar sentence is meaningless, or not grammatically correct. Tarski, Quine, and Russell have all taken this tack in one form or another. In proving meaninglessness, some theories claim that language is only a crude representation of an underlying set of actual meanings and references which operate on multiple levels. The Liar sentence attempts to simultaneously reference more than one level in the hierarchy, which cannot be allowed.
- B. Another approach is to argue that the Liar is neither true nor false. Kripke adopts this view, categorizing the Liar sentence with those lacking in reference altogether, such as "The present king of France is bald" (spoken at a time in which France has no king.) Thus is introduced the concept of a "truth value gap," into which such self-referential sentences fall.
- C. A theory proposed by Prior claims that to interpret the Liar as a paradox is to make an invalid assumption about the nature of its proposition. Drawing subtle distinctions in meaning, this view holds that the Liar could be construed as forming either a *negation* of itself (making it simply false) or a *denial* of itself (making it simply true). Philosophers Barwise and Etchemendy espouse this theory.

- D. A more radical way out of the paradox is to accept that the Liar is both true and false, and then adapt the rules of formal logic to accommodate this condition. The use of paraconsistent logic in this methodology tends to result in a weaker formal system of logic.
(Dowden, Liar Paradox, 2010)

For any of the above solutions, there are consequences. In many cases, classical logic must be revised. In others, new problems are introduced. In his book, *Revenge of the Liar*, Beall writes “[T]he Liar’s Revenge phenomena is reflected in the apparent hydra-like appearance of Liars: once you’ve dealt with one Liar, another one emerges.” (Beall, *Revenge of the Liar*, *New Essays on the Paradox*, 2007) Elsewhere, Beall chooses the term “spandrels” to describe “unintended by-products” of re-conceptualizing truth to accommodate the Liar. (Beall, *Spandrels of Truth*, 2009)

The solution I will propose is most closely akin to family (B), in that I will say that the Liar is neither true nor false. However, the grounds for my claim are very different than those of Kripke. For Kripke’s solution, the spandrels are quite severe. For example, the existence of a truth value gap presents difficulties in defining falsity, specifically in deflationary theory. The complexities of having a partially interpreted truth value predicate seem to outweigh any advantage gained by overcoming the logical paradox, and it is not clear that Kripke’s solution is tenable under all possible circumstances.
(Damjanovic & Stoljar, 2010) (Dowden, Liar Paradox, 2010)

In another sense, it seems to me that a theory of truth should be general, avoiding special interpretation for cases such as the Liar. Since the mechanics of a Liar sentence and any other self-referencing declarative sentence are essentially the same, does the truth value gap become a chasm into which we can arbitrarily throw any sentence whose semantics happen to bother us? If the self-referential sentence “this sentence is false” should be ruled to have no truth value, should not the same be said of this sentence:

9. Sentence (9) is true

Why is it that only a sentence involving assertion of its own falsity should be stripped of its truth value and tossed into the gap? Of course, sentence (9) does not pose the same problem of logic that sentence (4), did. However, it should be clear that any treatment of the liar paradox which affects the interpretation of self-referential sentences must apply equally to all self-referential sentences, paradoxical or not.

Another type of circular sentence is as this example:

10. Sentence (10) is in English

It may be tempting to call sentence (10) self-referential, but it is not – at least not in the sense applicable to the Liar. Although sentence (10) does describe one of its own properties, it does not refer to its own truth value. That being the case, it is clear that circularity itself is not at issue, the only condition being the case of self-reference to truth value.

Truth Defined, Part 2

Previously, I declared the nature of truth I would be using in this paper – namely, deflationism, broadly defined. In light of the previous section, I will now expand this definition to include the application of truth to the Liar Paradox.

The truth value of a sentence has an interesting relationship to its sentence. Truth or falsity has historically been considered a *property* of a sentence. That view has changed significantly in the last century. A contemporary textbook on the subject of symbolic logic declares that “truth value is not really part of what, in ordinary language, we think of as a statement’s meaning. This is apparent when we consider that one can often completely understand a statement without having any idea of whether or not it is true.” (Bessie & Glennan, 2000) Even more to the point, deflationist theories of truth specifically reject truth as a property of individual sentences.

Consider two true sentences:

11. The earth revolves around the sun
12. Sacramento is the capitol of California

In some sense, both sentences (11) and (12) do share a generalized property of being true. However, if we are using the global truth predicate *T*, can we say that both of these sentences are *T*? If so, it should be the case that there is a common explanation of the reasons sentences (11) and (12) are both *T*. In this sense, there is nothing shared between the truth values of sentences (11) and (12) - (11) is true by virtue of the physics which cause earth to revolve around the sun, and (12) because of California’s history leading to the establishment of Sacramento as its capitol. (Damnjanovic & Stoljar, 2010)

Causal Theory: First Formulation

If a truth value is not a property of its related sentence, then what exactly is it? In response to this question, I offer the first of my two theses: *the truth or falsity of a declarative sentence is an effect of the existence of the sentence*. In other words, sentence and truth value have a causal relation, in which the sentence causes the truth value.

From this point forward in this paper, I will omit the word “value” when I refer to truth as an effect, and simply refer to the truth effect of sentences as “ttruth,” adopting J.C. Beall’s convention for “transparent truth.” I take this measure to keep from reinforcing the misleading idea of “value,” which might indicate a “property” inherent or adherent to an object. In the sense I will be using, ttruth refers to the effect of a declarative sentence which determines whether the sentence is transparently true or not, as defined in (1), above. I will return to this topic later in this paper.

In addition to viewing ttruth as an effect, I propose that we adopt the position that *declarative sentences about the future have no present ttruth*. The idea that references to the future have no present truth value has a long history in philosophy, and there exist multiple proposed systems of logic

intended to deal with contingent truth. I am extending this concept to include ttruth, and will explore the topic, along with causality, later in this paper.

Armed with these concepts, my Causal Theory of Truth is exemplified in the following argument:

13. This sentence is false
14. The ttruth of (13) is caused by (13)
15. Any cause must temporally precede its effect

16. Therefore, at the moment of occurrence of (13), the ttruth of (13) has only a future existence
17. The law of excluded middle does not hold for future tense declarative sentences

18. Therefore, (13) is neither true nor false at the moment of occurrence

Cause and Effect

Premise (14) depends upon the causal relationship previously alluded to. At a superficial level, I find it convenient to adopt the *counterfactual* theory of causation, exemplified by the statement “ p is the cause of q just in the case that q would not have happened in the absence of p .” (Garrett, 2006) (Menzies, 2008) If this is the case, then it follows that the ttruth of (13) must have been caused by (13), because that particular instance of ttruth would not exist in the absence of (13).

Of course, there are other theories of causality which could be applied to demonstrate that ttruths are caused by sentences. In reviewing some of these theories, I find that many of them, although they might seem to support my thesis, treat causality as a relation between *events*. Counterfactual theory, as defined above, is just this way. I want to avoid thinking of ttruth as an *event*, because I suspect that such a supposition will lead our conceptualization down the wrong path by its connotation. Ttruth itself should be conceptualized not as something that happens, but as a thing that simply exists at some moment or period in time. Therefore, I appeal to a somewhat more alien theory of causation.

The Indian philosophy of Sāṅkhya gives us causation conceived of as a necessary relation between a thing and its origin. Instead of relying upon events and facts as the relata between cause and effect, Sāṅkhya relies upon objects causing the existence of other objects. Although this may sometimes seem to be the case in western speech, it is not. For example, “Cars cause thousands of deaths each year” sounds like the relation between objects “car” and “deaths.” However, the propositions contained in this example are really the *impacts* of cars and the *events* of deaths. The Indian notion of cause and effect follows more closely to the literal interpretation of the example, which could be restated as “the existence of a car impact necessitated the existence of a death” (plurals dropped for clarity.) (Garrett, 2006) (Ruzsa, 2006)

Also important to note is the close relation between cause and effect in Sāṅkhya. A cause is considered the origin of a thing; the external equivalent of the intellectual process of inference. Given the example of a potter making a pot from clay, it is the clay which is attributed the prime cause of the pot. In this way, the effect is essentially identical with its material cause. (Ruzsa, 2006) If the cause and effect are

substantially the same then, in essence, this is another way of conceptualizing the deflationist theory of truth defined in (1), the assertion that “it is true that *a*” and “*a*” are interchangeable.

References to the Future

Outside of quantum mechanics, causes must always precede their effects temporally, which is the claim of premise (15). Setting aside the possibility of simultaneity for a moment, we can visualize a cause and effect chain by the example of a billiard ball. Imagine a cue ball in motion toward a motionless eight ball, on a collision course. At some point in time, the cue ball collides with the eight, imparting energy to it. At some point in time after the moment of contact, the eight ball is in motion. We can say that the motion of the cue ball caused the eight ball to move. However, it should be apparent that the only cue ball motion relevant to *cause* is that before the instant of contact, and the only eight ball motion relevant to *effect* is that after the same instant. If we can agree that the instant in which the energy is imparted has no duration, then the cause in this example can only precede its effect – there can be no overlap between cause and effect. I posit that the analogy of the billiard balls is equivalent to the operation of truth, in that the occurrence of a declarative sentence is complete prior to the existence of the resulting truth.

There are views of causality which allow for *simultaneous causation*. I find the analogy of the billiard ball convincing – that there is no overlap between cause and effect. Replies to theories of simultaneous causation generally hinge on the idea that apparent examples of it are misdescribed. (Schaffer, 2007)

The law of excluded middle (I will abbreviate as LEM), referenced in (17), holds that a declarative sentence is either *true* or it is *not true*, with no other choices allowed. This is one of the defining properties of classical systems of logic. However, many philosophers going back as far as Aristotle have held that propositions about the future cannot have a truth value. For example, the statement “It will rain tomorrow” cannot be said to be true or false until tomorrow arrives, at which time we can assign truth or falsity *post hoc*. This represents a kind of loop-hole for logic which would otherwise be flawed. (Dowden & Swartz, Truth, 2004)

The future reference loop-hole is not without its problems. It has been shown that certain deductively valid arguments become unprovable in the face of it. For example:

19. We’ve learned there will be a run on the bank tomorrow.
20. If there will be a run on the bank tomorrow, then the CEO should be awakened.

21. So, the CEO should be awakened.

With classical logic, there is a strong case that this otherwise valid argument fails once deprived of its truth values by the future reference loop-hole. (Dowden & Swartz, Truth, 2004)

It remains to be seen whether the conception of ttruth as I have defined it will have a positive or negative effect on other arguments surrounding LEM or its exceptions. I suspect that it could be used as a basis for support of the future reference loop-hole, in that ttruth exists only in the future for *all* declarative sentences. However, I will leave that argument for another time.

With premise (17) I assert that the ttruth of (13) is in the future relative to the existence of (13), and conclude in (18) that it therefore cannot be applied to (13) at the time of occurrence. This conclusion shows that the sentence is not, in fact, self-referential in regard to ttruth.

Or is it? Once the ttruth of a Liar sentence obtains, could the observer somehow take that ttruth and apply it to the sentence? In addressing this question, I will start with a simple example: changing the past.

Consider this sentence:

22. I will take a drink from my coffee cup one minute from now

Because it references the future, (22) has no present ttruth. However, once a minute has past, and I do (or do not) have the drink described, has it happened that ttruth becomes assigned to (22), thereby changing something that exists in the past? Of course, this is no different than any other type of future tense assertion, and should be treated as such. (Faye, 2010)

Effects Upon the Liar

The basic form of the Liar, "this sentence is false," is used in my formulation of Causal Theory, above. It should be easily seen that the same principle applies to other linguistic varieties of the Liar Paradox, such as the Strengthened Liar shown in (5), and the "split" Liar shown collectively in (2) and (3). It doesn't matter how the Liar is redistributed, linguistic Liars are defused by Causal Theory.

More difficult is the Russell Paradox, from the field of mathematics. Copied from above for reference:

6. *let $R = \{x \mid x \notin x\}$, then $R \in R \Leftrightarrow R \notin R$*

In set theory, we can apply the same principle. To say that a set is or is not a member of itself requires the set to exist prior to the delineation thereof. In other words, at the moment of the occurrence of the set definition, the set contains no elements. After the set has been defined, it becomes populated with all elements that met the defining criteria at the moment of set definition. Set definition is the cause and the set population its effect. Since the set itself did not exist until after the moment of set definition, it cannot be considered part of the set. In this realm, the set definition is analogous to a declarative sentence, and the contents of the set to ttruth.

It would please me greatly to have the capacity to discuss Tarski's Undefinability Theorem at this point. However, I am afraid that topic will have to wait until additional research can be made, by myself or others.

Conclusion

For any declarative sentence, its truth is in its future and cannot, therefore, be meaningfully referenced from within the sentence. Without the Russell / Liar Paradox to generate contradictory premises, the paradoxical proofs described above become impossible.

This is only a hypothesis. The theory I propose will only be realized with additional research, and much deliberation, if at all. In future papers, I hope to cover the topic in more detail, with more authority in the areas of formal logic, mathematics, and semantics. I hope to explore side effects, the “spandrels” of Causal Theory, as well as objections and relationships to other theories. Can this theory be applied to classical logic in a consistent manner? Does this theory require sweeping changes to classical logic in order to be useful? Is the application of this theory limited, or can it be generally applied? These questions and more remain.

One final thought: When viewed in the light of Causal Theory, the Liar sentence should no longer seem to be truth self-referential. Of course, the natural language version of the Liar remains unchanged, and will continue to “sound odd” to the ear. Or will it? Could it be that the only reason the Liar sentence strikes us as odd is that we have spent our lives immersed in a peculiar conceptualization of cause and effect? Is it possible that future generations would read the Liar, and see it something like the potter's clay, with its truth like a pot emerging as an effect of the clay's existence?

References

- Allen, R. E. (1991). *Greek Philosophy, Thales to Aristotle*. New York: The Free Press.
- Beall, J. (2007). *Revenge of the Liar, New Essays on the Paradox*. New York: Oxford University Press.
- Beall, J. (2009). *Spandrels of Truth*. Oxford: Clarendon Press.
- Bessie, J., & Glennan, S. (2000). *Elements of Deductive Inference*. Belmont, CA: Wadsworth.
- Damnjanovic, N., & Stoljar, D. (2010, October 4). *The Deflationary Theory of Truth*. Retrieved 01 15, 2011, from Stanford Encyclopedia of Philosophy: <http://plato.stanford.edu/entries/truth-deflationary/>
- Dowden, B. (2010, 4 6). *Liar Paradox*. Retrieved 1 12, 2011, from Internet Encyclopedia of Philosophy: <http://www.iep.utm.edu/par-liar/>
- Dowden, B., & Swartz, N. (2004, September 17). *Truth*. Retrieved January 16, 2011, from Internet Encyclopedia of Philosophy: <http://www.iep.utm.edu/truth/>
- Faye, J. (2010, February 16). *Backward Causation*. Retrieved January 18, 2011, from Stanford Encyclopedia of Philosophy: <http://plato.stanford.edu/entries/causation-backwards/>

- Garrett, B. (2006). *What is this thing called Metaphysics?* New York: Routledge.
- Irvine, A. D. (2009, May 27). *Russell's Paradox*. Retrieved January 15, 2011, from Stanford Encyclopedia of Philosophy: <http://plato.stanford.edu/entries/russell-paradox/>
- Lorkowski, C. M. (2010, November 7). *David Hume: Causation*. Retrieved January 14, 2011, from Internet Encyclopedia of Philosophy: <http://www.iep.utm.edu/hume-cau/>
- Menzies, P. (2008, March 30). *Counterfactual Theories of Causation*. Retrieved January 16, 2011, from Stanford Encyclopedia of Philosophy: <http://plato.stanford.edu/entries/causation-counterfactual/>
- Priest, G. (2007). Revenge, Field and ZF. In J. Beall, & J. Beall (Ed.), *Revenge of the Liar, New Essays on the Paradox*. New York: Oxford Press.
- Russell, B., & Whitehead, A. N. (1913). *Principia Mathematica*. Cambridge: Cambridge University Press.
- Ruzsa, F. (2006, March 6). *Sāṅkhya*. Retrieved January 16, 2011, from Internet Encyclopedia of Philosophy: <http://www.iep.utm.edu/sankhya/>
- Schaffer, J. (2007, August 13). *The Metaphysics of Causation*. Retrieved 01 18, 2011, from Stanford Encyclopedia of Philosophy: <http://plato.stanford.edu/entries/causation-metaphysics/>
- Various. (1978). *Paradox of the Liar*. (R. L. Martin, Ed.) Atascadero, CA: Ridgeview Publishing Co.