Preserve Self-Reference to Solve Paradoxes

David Wyde

July 20, 2023

Abstract

A self-referential sentence, when negated or joined with another sentence, should refer to the new sentence (itself) rather than the original one. Thus, the liar and the truth-teller are negations of each other. This paper uses a classical logic based on the above approach to examine the liar, Curry's paradox, and related topics.

1 Introduction

This paper describes a way of solving the liar paradox, Curry's paradox, and other problems that arise from self-referential sentences. It follows the principle that similar paradoxes should have similar solutions[1].

Section 2 sketches the approach. Section 3 examines the liar and the truth-teller as motivating examples. Section 4 attempts to justify the main idea by looking at the details of how sentence labeling works. The following sections examine a variety of paradoxes and results that the approach may apply to, then Section 15 summarizes the paper.

2 The Approach

We will use classical logic, with three important details:

1) A self-referential sentence, when negated or joined with another sentence, should refer to the new sentence (itself) rather than the original one.

2) When given a proposition p to negate,

   i) Attempt to assign p as true or false.

   ii) Find p's negation, a separate proposition, $\neg p$, and determine whether $\neg p$ is true or false.

   iii) Ensure that exactly one of p and $\neg p$ is true, by the laws of excluded middle and noncontradiction.

3) If a proposition is true, add it to a collection of axioms and theorems that can be used to prove new results. If a proposition is not true, discard it.

We will examine Detail 1 in Section 4. For Detail 2, the idea that p and $\neg p$ should be separate sentences is not new[2]. Detail 3 is in the tradition of the mechanization of mathematics, as in Hilbert's program and *Principia Mathematica*[3]. Section 6 will contest Gödel's proof that dealt a blow to those approaches[4].

3 Motivating Examples

We begin by examining the liar,

> This sentence is false,

and the truth-teller,

> This sentence is true.

It is difficult to assign either true or false to the liar. On the other hand, it seems possible to assign either true or false to the truth-teller, and our problem is that we do not know which one to pick[5].

Using the approach described in Section 2, we have the following proof:

1) The liar and the truth-teller are negations of each other (Detail 1).

2) It is not possible to assign true to the liar, but it is possible to assign true to the truth-teller (Detail 2).

3) We determine that the truth-teller is true and discard the liar (Detail 3).

4) We can thus resolve the liar paradox.

In the following sections, we will refer to the above argument as (*).

4 Sentence Labeling

Whether Detail 1 makes sense depends on the specifics of how we assign names to sentences. This section considers three approaches to sentence names: a) as equalities; b) as labels, and self-referential variables bind to the original sentence; c) as labels, and self-referential variables bind to "this sentence."

We will use p*:* $p \to q$, where q is an arbitrary proposition, as an example. We will talk more about this example in Section 8. For now, note that it is true when p and q are both true. Otherwise, the left-hand side and right-hand side disagree, and it will be a contradiction.

Suppose that we want to replace the right-hand side with $\neg p \vee q$, which has the same truth table in this paper's logic.

With a), we have

$$p = p \to q.$$

Then, the equality still holds if we do

$$p = \neg p \lor q,$$

as in a mathematical equation. The proposition remains self-referential, as in Detail 1.

For b), we have

$$p\text{: } p \to q.$$

When we change the right-hand side, we will have a new sentence. A new sentence needs a new label, and we will use r. We write

$$r\text{: } \neg p \lor q.$$

The right-hand side refers to the original sentence rather than the new one. That is how people seem to use negation when talking about the liar[6].

Such an approach has the drawback that r is not self-referential and thus has a different truth table than the original p. For example, if p and q are both false, r is true rather than contradictory. Due to the lack of self-reference, the left-hand side absorbs the truth value of the right-hand side.

For c), we have

$$p\text{: } p \to q,$$

as in b). This time, we will write

$$r\text{: } \neg r \lor q,$$

where r has the same truth table as p does in the original sentence and a).

The right-hand side refers to the new sentence rather than the old one, but the original p and c)'s r have the same structure. Both say,

$$(\text{This sentence}) \to q.$$

Sentences such as a)'s $p = \neg p \lor q$ and c)'s $r\text{: } \neg r \lor q$ are thus equivalent. Two other sentences that say the same thing with different labels are

$$s\text{: } 5 > 0$$

and

$$t\text{: } 5 > 0.$$

This paper argues that the approach taken in c) is similar to s and t, and that a) and c) work better than b). That supports Detail 1.

5 The Liar and the Truth Teller, Continued

We now return to the liar and the truth-teller. Given a truth predicate *Tr*, consider the following sentences:

(1): Tr(1)
(2): Tr(1)
(3): ¬Tr(1)
(4): ¬Tr(4)
(5): Tr(4)
(6): ¬Tr(4).

(1) is the truth-teller, and (4) is the liar. A sentence such as

(7): Tr(7)

is equivalent to (1) and is thus also an instance of the truth-teller.

We might wonder whether any of the sentences (1)–(6) are equivalent to each other, and what the negations of (1)–(6) are.

Using Detail 1 and following the approach for a) and c) in Section 4, (2) and (6) reduce to (1) and thus to the truth-teller. (3) and (5) reduce to (4) and thus to the liar.

(1) and (4) are negations of each other. Anything that reduces to (1) has (4) as its negation, and anything that reduces to (4) has (1) as its negation.

6 Gödel's Incompleteness Theorems

6.1 Background and a Conflicting Proof

Gödel's first incompleteness theorem says that a sufficiently powerful formal system contains statements that the system cannot either prove or disprove. To create such an undecidable statement, Gödel used a sentence that asserts its own unprovability[7].

We can consider six sentences analogous to (1)–(6) from Section 5, using *Pr* as a provability predicate instead of *Tr* for truth:

(8): Pr(8)
(9): Pr(8)
(10): ¬Pr(8)
(11): ¬Pr(11)

(12): $\mathrm{Pr}(11)$

(13): $\neg\mathrm{Pr}(11)$.

(11) is Gödel's undecidable sentence. Some authors say that it is true but unprovable[8], while others question whether it should be assigned true[9] or say that it is inconsistent[10].

As with (1)–(6), it is possible to ask what the negations of (8)–(13) are, and whether any of (8)–(13) are equivalent to each other.

Again following the truth-teller and the liar, (9) and (13) reduce to (8), and (10) and (12) reduce to (11).

As in (1)–(6), (8) and (11) are negations of each other. A proof analogous to (*) renders Gödel's undecidable sentence decidable, which undermines Gödel's first incompleteness theorem.

Gödel's second incompleteness theorem says that a consistent and sufficiently powerful formal system cannot prove its own consistency[11]. The proof for Gödel's second incompleteness theorem depends on Gödel's first incompleteness theorem, so (*) is also an argument against Gödel's second incompleteness theorem.

6.2 Additional Notes

Separately from (*) and unlike the truth-teller, (8) is already considered a theorem[12]. The proof is similar to Curry's paradox, which we will discuss in Section 8.

The tools that Gödel used to create self-reference encode the symbols that appear on the right-hand side of a logical expression. A Gödel numbering, or another encoding based only on the right-hand side, cannot directly represent both (11) and (13) for the same sentence.

Whether we consider (8)–(13) as more or less fundamental than sentences created by a Gödel numbering may influence how we negate the relevant sentences and say whether any of them are equivalent to each other.

7 Tarski's Undefinability Theorem

Tarski's undefinability theorem says that a language cannot define certain aspects of itself: for truth, that can only be done in a metalanguage.

The proofs of Tarski's undefinability theorem and Gödel's first incompleteness theorem are similar. First, diagonalize to create a liar-like sentence. Then, show that the liar-like sentence and its negation both lack a given property. Finally, conclude that there are certain fundamental limits on formal systems[13].

Both Gödel's incompleteness theorems and Tarski's undefinability theorem have the same issue: by (*), the negation of the given liar-like sentence should be a truth-teller-like sentence that must be a theorem. In each case, the problematic sentence is resolved and the proof no longer works.

8 Curry's Paradox

Curry's paradox describes the following proof of any proposition q, regardless of whether q is true or false, in a small number of steps[14]:

1) p: p $\rightarrow$ q (define a sentence)
2) p $\rightarrow$ p (identity)
3) p $\rightarrow$ p $\rightarrow$ q (substitute 1 into 2)
4) p $\rightarrow$ q (contraction of 3)
5) p (substitute 1 into 4)
6) q (modus ponens on 4 and 5).

Aside from substituting variables, the main assumptions in the above proof are i) $p \rightarrow p$, ii) contraction, and iii) modus ponens. There is a survey of approaches that oppose ii) and iii) in [15]. This paper questions i).

With the standard approach to negation, *2)* $p \rightarrow p$ is a tautology. Using Detail 1, p should be self-referential, and thus *2)* $p \rightarrow p$ reduces to *2): 2)* $\rightarrow$ *2)*. Informally, *p:* $p \rightarrow q$ is "If this sentence is true, then q." Then p is "This sentence is true," and $p \rightarrow p$ is "If this sentence is true, then this sentence is true."

When trying to assign a truth value to 2) in *2): 2)* $\rightarrow$ *2)*, true works: both sides are true. If we assign 2) to be false, then the left-hand side is false and the right-hand side is true: a contradiction. Thus, 2) asserts something that is not necessarily true.

By similar reasoning, step 4) reduces to asserting the truth of *p:* $p \rightarrow q$. The only non-contradictory assignment of *p:* $p \rightarrow q$ is when both p and q are true. In the other three cases, the left-hand side and the right-hand side disagree, causing a contradiction. Thus, something has gone wrong by 4).

An argument that can prove any statement resembles the principle of explosion[16]. In the case of Curry's paradox, 2) is a false assumption, from which anything follows.

For a different perspective on when $p \rightarrow p$ fails, see [17].

9 Validity Curry

A variant of Curry's paradox is called validity Curry or v-Curry[18]. An example version is "The argument from this sentence to absurdity is valid." Its negation, "The argument from this sentence to absurdity is not valid," is analogous to the truth-teller. An argument that resembles (*) resolves the validity Curry paradox.

10 The Knower Paradox

The knower paradox,

(14): (14) is not known,

is another paradox with similar structure to the liar but a different predicate. It is thought to raise questions that are related to knowledge in addition to self-reference[19]. (*) works here, too.

11 The Barber Paradox

(*) also seems to apply to the barber paradox[20]. The analogy gets stretched when the barber must get a haircut in either the original sentence or its negation, but, given that assumption, the argument may work.

12 Russell's Paradox

Russell's paradox is related to the barber paradox, but the former's resolution depends on issues related to set theory[21]. There may be technical reasons that "the set of all sets that contain themselves" cannot exist, in which case (*) would not work.

13 Grelling's (Heterological) Paradox

The heterological paradox, like Russell's paradox, is connected to the barber paradox[22]. As in Russell's paradox, there are specifics that may prevent (*) from working.

Sometimes natural language behaves differently from classical logic. Examples include double-negation elimination in sentences such as "I'm not unhappy,"[23] and vacuously-true implications[24].

Also, it may not be necessary that exactly one of "autological" and "heterological" describes itself.

Thus, it does not seem that (*) works in this case.

14 Systems of Sentences

A group of sentences that refer to each other can resemble the liar; for example,

The next sentence is true.

The previous sentence is false.

As in the liar, there does not seem to be a consistent way to assign truth values these two sentences[25].

Yablo's paradox considers an infinite system of sentences that refer to each other[26]. As in the finite case, it seems impossible to assign truth values to the sentences in Yablo's paradox.

For both finite and infinite systems of sentences that refer to each other, it is not clear how to apply Details 1 and 2. Can we negate an individual sentence to get the negation of the system, or do we need to consider all of the sentences at once when finding a negation?

It is possible that Detail 1 is a special case of a more general algorithm for negating systems of sentences that refer to each other. One possible way forward is to analyze the sentences as a labeled, directed graph[27, 28, 29, 30] and try to define some notion of a complement[31]. To solve finite sentence cycles and systems like Yablo's paradox, such an approach may be needed.

15 Conclusion

This paper presents a version of classical logic that resolves the truth-teller, the liar, Curry's paradox, v-Curry sentences, the knower paradox, and the barber paradox. The approach detailed in Section 2 may also extend to other paradoxes.

A possible benefit to this approach is that it only impacts propositions that involve self-reference. All other sentences and proofs remain unchanged.

One consequence of using a two-valued logic is that we group truth-teller-like and true statements together under the truth predicate, and we group false sentences with liar-like sentences. This is a result of the pigeonhole principle[32], though we could group sentences together in a different way.

There are several ways to react to an undecidable sentence such as the one in Gödel's incompleteness theorems. We might question the law of excluded middle, since a sentence and its negation both fail to be theorems. Or, we may view the result as demonstrating fundamental limits on logic. It is also possible that the undecidable sentence occurred because we negated incorrectly.

Finite and infinite systems of sentences pose a challenge for this paper's approach. It is possible that Detail 1 is a special case of a negation algorithm that can be generalized to any number of sentences that refer to each other.

Acknowledgements

Eric Hehner reviewed previous papers and pointed out that preserving self-reference could apply to operations other than negation. Julien Murzi emphasized the need to justify why the truth-teller is true in response to an earlier version of these ideas.

References

[1] Graham Priest. The Structure of the Paradoxes of Self-Reference. *Mind*, 103(409):25–34, 1994.

[2] Harrie de Swart. *Philosophical and Mathematical Logic*, pages 23–24. Springer, 2018.

[3] Jeremy Avigad. The Mechanization of Mathematics. *Notices of the American Mathematical Society*, 65(6):681–90, 2018.

[4] Richard Zach. Hilbert's Program. In Edward N. Zalta and Uri Nodelman, editors, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2023 edition, 2023.

[5] Chris Mortensen and Graham Priest. The Truth Teller Paradox. *Logique et Analyse*, 24(95/96):381–388, 1981.

[6] Jc Beall, Michael Glanzberg, and David Ripley. Liar Paradox. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Fall 2020 edition, 2020.

[7] Raymond M. Smullyan. *Gödel's Incompleteness Theorems*, page 11. Oxford University Press, 1992.

[8] Ernest Nagel and James R. Newman. *Gödel's Proof*. NYU Press, 2001.

[9] Hartry Field. Which undecidable mathematical sentences have determinate truth values? In H.G. Dales and G. Oliveri, editors, *Truth in Mathematics*, pages 291–310. Oxford University Press, 1998.

[10] Francesco Berto. The Gödel Paradox and Wittgenstein's Reasons. *Philosophia Mathematica*, 17(2):208–219, 2009.

[11] Kurt Gödel. Über formal unentscheidbare Sätze der Principia Mathematica und verwandter Systeme I. *Monatshefte für Mathematik und Physik*, 38:173–198, 1931. Reprinted as: On Formally Undecidable Propositions of Principia Mathematica and Related Systems. Translated by B. Meltzer. Dover, 1992.

[12] M. H. Löb. Solution of a Problem of Leon Henkin. *The Journal of Symbolic Logic*, 20(2):115–118, 1955.

[13] Hartry Field. *Saving Truth from Paradox*, pages 23–30. Oxford University Press, 2008.

[14] Haskell B. Curry. The Inconsistency of Certain Formal Logics. *The Journal of Symbolic Logic*, 7(3):115–117, 1942.

[15] Lionel Shapiro and Jc Beall. Curry's Paradox. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Winter 2021 edition, 2021.

[16] Otávio Bueno and Scott A. Shalkowski. Modalism and Logical Pluralism. *Mind*, 118(470):295–321, 2009.

[17] Matthew Mandelkern. If p, then p! *The Journal of Philosophy*, 118(12):645–679, 2021.

[18] Jc Beall and Julien Murzi. Two flavors of Curry's paradox. *The Journal of Philosophy*, 110(3):143–165, 2013.

[19] Walter Dean and Hidenori Kurokawa. Knowledge, proof and the Knower. In *Proceedings of the 12th Conference on Theoretical Aspects of Rationality and Knowledge*, pages 81–90. Assocation for Computing Machinery, 2009.

[20] W. V. Quine. Paradox. *Scientific American*, 206(4):84–99, 1962.

[21] Andrew David Irvine and Harry Deutsch. Russell's Paradox. In Edward N. Zalta, editor, *The Stanford Encyclopedia of Philosophy*. Metaphysics Research Lab, Stanford University, Spring 2021 edition, 2021.

[22] Noson S. Yanofsky. *The Outer Limits of Reason*. MIT Press, 2013.

[23] Laurence R. Horn. Negation and opposition: Contradiction and contrariety in logic and language. In Viviane Déprez and M. Teresa Espinal, editors, *The Oxford Handbook of Negation*, pages 7–25. Oxford University Press, 2020.

[24] David Sherry. Formal Logic for Informal Logicians. *Informal Logic*, 26(2):199–220, 2006.

[25] Anil Gupta and Nuel D. Belnap. *The Revision Theory of Truth*, pages 17–18. MIT Press, 1993.

[26] Stephen Yablo. Paradox without Self-Reference. *Analysis*, 53(4):251–252, 1993.

[27] Haim Gaifman. Pointers to Truth. *The Journal of Philosophy*, 89(5):223–261, 1992.

[28] Roy T. Cook. Patterns of Paradox. *The Journal of Symbolic Logic*, 69(3):767–774, 2004.

[29] Landon Rabern, Brian Rabern, and Matthew Macauley. Dangerous Reference Graphs and Semantic Paradoxes. *Journal of Philosophical Logic*, 42(5):727–765, 2013.

[30] Timo Beringer and Thomas Schindler. A Graph-Theoretic Analysis of the Semantic Paradoxes. *Bulletin of Symbolic Logic*, 23(4):442–492, 2017.

[31] Richard J. Trudeau. *Introduction to Graph Theory*. Dover, 1993.

[32] Kenneth H. Rosen. *Elementary Number Theory*. Pearson, 2011.