The Pathological Liar: An Exclusionary Approach to Self-Referential Contradictions in Natural Language

Shomir J. Wilson, Virginia Polytechnic Institute and State University

The liar's paradox is a simple yet harrowing problem that has plagued logicians and philosophers who study natural language. It is an uncommon sentence, and perhaps a largely useless one outside of artificially-created circumstances. However, it threatens to undo any naively-constructed theory of natural language semantics, if the theory does not take the paradox into account. Indeed, many well-known theories of language semantics have been built specifically to avoid the liar's paradox or to solve it altogether. For theorists to spend so much effort on a single problem—moreover, a case with virtually no practical use—seems, intuitively, a misdirection of effort. To articulate this, first the structure and significance of the liar's paradox will be examined, as well as Richard Kirkham's five criteria for solutions that handle the paradox and preserve the integrity of natural language. Next, the liar's paradox will be classified as a *pathological problem*, which will influence the development of a theory of natural language semantics. Then, Kirkham's five criteria will be reexamined in light of the pathological nature of the liar's paradox, demonstrating how some of the criteria are unnecessary for its resolution. This will point to conclusions on what a solution to the liar's paradox can possibly mean, and what forms such a solution might have. The liar's paradox is a pathological problem, a status that dramatically affects the criteria for any solution.

In its simplest form, the liar's paradox is a single sentence that denies its own truth. For instance, the sentence

1. This sentence is false.

is a concise instance of the paradox. To assume that sentence (1) is true is to assert its meaning, that sentence (1) is false. To assume that sentence (1) is false is to assert the negation of its meaning, that sentence (1) is true. In either case, the sentence is simultaneously true and false—not one or the other, but both at the same time. In our standard conception of logic, true and false are mutually exclusive and comprehensive properties for declarative statements. That is, it should be possible to classify any declarative sentence as either true or false, but never both. Because it is effectively both at once, sentence (1) is indeed a paradox.

Additionally, the liar's paradox need not be confined to one sentence. Consider the pair of sentences

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

To assume that sentence (2) is true leads us to assert that sentence (3) is false. However, to assert that sentence (3) is false leads us to assert that sentence (2) is false, contrary to our original assumption; and if sentence (2) is false, then sentence (3) is true, contrary to our derivation. To assume that sentence (2) is false leads us to assert that sentence (3) is true. However, to assert that sentence (3) is true leads us to assert that sentence (2) is true, contrary to our original assumption; and if sentence (2) is true, then sentence (3) is false, contrary to our derivation. Beginning our analysis of this sentence pair by assuming a truth value for (3) produces similar results. No matter how we begin, the final result violates the exclusive nature of *true* and *false* that we generally take for granted, producing a paradox.

Furthermore, it is possible to expand the liar's paradox to an arbitrary number of sentences. Consider a set of sentences that take the form

> *n*. Sentence (*n*+1) is true.
>
> …
>
> N. Sentence (4) is false.

with *n* consisting of values from 4 to N-1 and some constant N at least 5. The reader may verify that the result of assuming the truth value for any sentence in the above set leads to the violation of bivalence for all sentences in the set.

These paradoxes are a reason for more than mild concern. Both consist of a sentence or sentences that are well-formed and consistent with the grammar for our natural language. The number labels may be removed from the multiple-sentence paradox by using "next" in place of (n+1); the result is a set of sentences just as linguistically clear as the one-sentence paradox. Syntactically, these paradoxes seem perfectly fine, so nothing immediately obvious prevents them from being formed. However, semantic analysis reveals that they utilize our intuitions on the meaning of truth to produce a seemingly impossible result. It is as if natural language has turned on us in rebellion by producing results that, in a sense, cannot be. Something must be wrong, but everything up to the paradoxical result seems fine. Accordingly, it has been the goal of many philosophical enterprises to justify our collection of intuitions on natural language while ousting the troublesome liar's paradox.

In response to the apparent need for a solution to the liar's paradox, Richard Kirkham has created five criteria that any solution to the liar's paradox ought to meet.[1] Two of the five shall be referred to as *positive-oriented* criteria, meaning they ensure that a solution is able to comprehensively dissect and solve the problem. The remaining three are *negative-oriented criteria*, which ensure that a solution does not interfere with the intuitive framework that it operates it within. While these criteria seem initially reasonable, it will be shown later in the paper that some are actually misleading.

The two positive-oriented criteria are *specificity* and *completeness*. Specificity requires a solution to clarify exactly what misstep in the semantic analysis of the liar's paradox leads to the impossible results. This prevents a solution from bluntly rejecting the problem that the liar's paradox poses while claiming that the derivation is coherent—a paradoxical solution in itself. Completeness requires a solution to address all instances of the liar's paradox, including all those composed of one, two, and N sentences. This criterion primarily prevents a solution from solving either one of the one-sentence paradox or the multiple-sentence paradox while leaving the other unharmed, since the multiple-sentence paradoxes are inductively congruent.

The three negative-oriented criteria are *avoidance of overkill*, *preservation of intuitions*, and *avoidance of ad hoc postulations*. Avoidance of overkill requires that a solution not overzealously affect the meaning or plausibility of sentences not involved in the liar's paradox. To preserve intuitions, it is necessary to construct a solution that does not change our standard notions of semantic concepts, particularly the bivalent notion of truth. Finally, the last negative criterion requires a solution to reject the liar's paradox in an incidental fashion, so that the problem is not the focus of the semantic theory being proposed as a solution. This is perhaps the most difficult of the five criteria, since it requires one to construct a solution to the liar's paradox while not focusing upon doing so. Later analysis will prove how this is truly a daunting (and perhaps unreasonable) standard to meet.

With the background set, it is necessary to import and expand upon a concept from mathematics to support further investigation. The term *pathological* often refers to an example function, instantiation, or other case that violates an otherwise universal notion of systemic behavior. Pathological cases are often counterintuitive and obscure, but they can illustrate how careful we must be when creating rules that have far-reaching results. At the same time, they can illustrate how some generalizations are impossible within the confines of a formal system. Although the term pathological is informal, two examples from mathematics will illustrate how some of the counterintuitive claims that will follow are not unreasonable or unprecedented.

One classic example of a pathological case is the Weierstrass function, designed specifically to violate a common assumption in calculus. The Weierstrass function is continuous at all places, yet differentiable at none—a bizarre combination of properties that illustrates how continuity everywhere cannot ensure differentiation anywhere. A more accessible pathological case involves a naïve definition of the rational numbers as simply those numbers that can be represented in the form of the fraction *a/b*, with *a* and *b* as integers. While this definition does cover all rational numbers, it also includes the case where *b* is zero. The result of dividing by zero is undefined, since the calculation is impossible. Considering the issue in the form of a literal question ("How many zeroes are there in one?") illustrates the depth of the problem. Moreover, consider the case when *a* and *b* are both zero, an even more difficult quandary. These cases are pathological in the sense that they disprove the otherwise valid assumption that *a/b* is always a clear, rational number. In fact, because of these cases, the set of rational numbers is defined as those that can be represented in the form *a*/*b*, with *a* and *b* as integers, *and b* as nonzero—a clearly exclusionary tactic to avoid the problem.

It is obvious that the existence of a pathological case is a problem for any naïvely-constructed system. However, the consequences of that problem and the form of a solution are less obvious. In the case of the Weierstrass function, the only fallout is the damage to our intuitions. The common definitions of continuity and differentiability are perfectly fine; we need only abandon our misconception that the former at all places on a function's range implies the latter at some place to retain a rigorous system. In the case of division by zero, something is wrong with the naïve definition, since it directly produced the incorrigible results. The solution almost universally recognized by mathematics is to exclude, by the definition of a rational number, that case. No other rationale is available, as this exclusion does not fit into any larger project—aside from the formation of a consistent system. In fact, if somehow division by zero did not produce an ambiguous result, it seems very likely that the definition of a rational number would not include the exclusive clause. For this reason, the exclusion of division by zero is a well-accepted *ad hoc* solution to this pathological problem.

To illustrate the relationship between these pathological problems and the liar's paradox, it must be established that natural language is also naïvely-constructed and informal. Natural language clearly satisfies both of those properties. No central, respected body of rules for natural language exists. It is a system that grows, sustains, and retracts features due to the largely random influences of history and culture. There is no widely followed effort in civilization to maintain features of natural language that lend to formalization—even when they happen to arise. It is a tool of expression that we continuously design and redesign to communicate the human experience, not an enterprise that entails a focus on formality. Any demands upon natural language must be evaluated with these properties in mind.

The luck of humanity would be incredible if natural language happened to be consistent throughout. However, the liar's paradox could be taken to imply that natural language has intractable flaws. As explained, the paradox is consistent with our syntactic rules of formation, yet it is the cause of a semantic inconsistency. Furthermore, it is useless to communication—the *de facto* purpose for natural language—because the paradox is indeed so troublesome and ambiguous. The habitat, triviality, and consequences of the liar's paradox thus render it a pathological case in the system of natural language. The existence of such a case is not even surprising, considering the established haphazard nature of this system. However, its consequences remain unexplored, and require further examination.

Although the existence of the liar's paradox remains a problem in the semantics of natural language, identifying it as pathological alters Kirkham's five criteria. Some of the criteria assume that natural language contains an underlying system capable of solving the liar's paradox, and that the structure of the system should be unearthed and then presented as the solution. However, it has been shown that natural language is constructed informally for communication, and that the liar's paradox is a pathological case because, among other reasons, it is useless toward that goal. Therefore, even if natural language does contain an implicit framework, there is no reason to expect it to resolve problems beyond its purpose.

In particular, the two invalidated criteria are specificity and avoidance of ad hoc postulations. Specificity assumes that it is possible to point out what step in the derivation of the liar's paradox is a mistake. On the contrary, it is possible that the derivation actually contains no mistake, and instead the result itself is the only inconsistency that we can identify. Because of that outcome, avoiding ad hoc postulations also may not be essential to formulating a reasonable solution. By its own nature, the uniqueness of a pathological case invites special handling, even possibly exclusion. As explained, such special handling should not be considered prohibitively radical, and has precedent in other formal systems. Under these restrictive circumstances, we can safely say that ad hoc is no longer necessarily a fallacy.

These problems with the five criteria suggest possible solutions to the liar's paradox that are very different from those Kirkham had in mind for examination. One solution is perhaps all too obvious at this point: as a pathological case, we may simply exclude the liar's paradox from having any truth value. At once, it is a straightforward, elegant solution and a simple tactic that avoids the complexity of many others that have been proposed. The simplicity stems from the fact that the exclusionary solution need not explain anything else about the system of natural language. Again, this is no different from the aforementioned examples from mathematics; we need not fully develop all of the properties of rational numbers—a topic in group theory, beyond the scope of this paper—to have sufficient grounds for excluding division by zero. After all, we are seeking a solution to *one problem*, the liar's paradox. Developing an orderly system for all natural language, although necessary in some solutions, should not be seen as a requirement for all of them.

Exclusion fares well when examined in light of Kirkham's remaining three criteria. It satisfies completeness, since all forms of the liar's paradox are pathological and are thus reasonable to exempt from having truth value. The solution clearly avoids overkill, since it has virtually no impact on the meaning of any other sentences in natural language. Perhaps most interestingly, exclusion preserves our intuitions on the meaning of truth. We instinctively feel that, if a sentence has truth value, then it should be exclusively either true or false. The liar's paradox defies that intuition by claiming to be equally true and false at once. Rejecting the ability of the paradox to have either truth value seems more reasonable than arbitrarily choosing one of them while rejecting the other.

The remaining two criteria are specificity and avoidance of ad hoc postulations. The inability of exclusion to meet them is apparent when compared to other solutions to the paradox. However, these two criteria have been eliminated with significant reason and thus are no longer our concern. Moreover, even if we *must* consider all five of the criteria, it is worth noting that most of the solutions critiqued by Kirkham fail to meet at least one of them. While that alone cannot justify exclusion as a solution—such justification was presented earlier in this paper—it does show that exclusion has at least the same magnitude of effectiveness as the others.

To continue the analysis of possible forms of solutions to the paradox, we turn our attention to the division between exclusion and other solutions that are considered more conventional. Many of those in Kirkham's survey are larger, sophisticated enterprises that explain much more about natural language than simply how the liar's paradox can be solved. Such extrapolation is necessary to satisfy the ad hoc criterion, although not entirely sufficient, since is possible to design a framework for natural language that clearly focuses on solving the paradox. That illustrates an additional, more sublime problem with the ad hoc criterion: essentially, it requires a solution not to appear to be a solution. While enforcement of this criterion can produce extensively useful results, the results are not solutions so much as they are systems that *facilitate* solutions by containing within themselves sufficient means to somehow remedy the liar's paradox. This disparity between apparent purpose and intended purpose is not a reason to reject a theory of the semantics of natural language. However, it raises the possibility that such a theory does not cover the array of other concerns that must be addressed in that context.

Finally, there exists another form of a solution that bears little resemblance to either exclusion or system construction. It is possible, in some sense of the term, to adopt an inclusive attitude toward the liar's paradox by granting it both truth and falsity at the same time. Unlike exclusion, this solution is very destructive for our fundamental intuitions on truth, particularly exclusive bivalence and non-contradiction. In fact, inclusion effectively denies that the liar's paradox is a problem at all—a hypothesis that exceeds the hypothetical leeway that we ought to grant to any solution. For that reason, it is the form of a solution that most clearly invites rejection. However, by contrast it illustrates how all other solutions share at least the common view that the liar's paradox is undeniably a problem, and must be dealt with in some way.

The structure, environment, and sheer nature of the liar's paradox all introduce difficulties toward crafting a solution. For that reason, we have considered classifying the paradox as a pathological case, similar to those that are recognized in mathematics. Such classification is the basis for a re-evaluation of Kirkham's criteria for any solution to the liar's paradox. It also leads us to consider exclusion from truth value as a viable solution to the paradox. While unconventional, this solution still bears significant merit when compared to others, for its directness, simplicity, and effectiveness toward resolving the semantic inconsistency in natural language.

[1] Richard L. Kirkham, *Theories of Truth* (Cambridge: MIT Press, 1997) 273.