The Liar Paradox:
A Consistent and Semantically Closed Solution

By
Ryan Edward Young

A thesis submitted for the degree of Doctor of Philosophy
of The Australian National University

July 2011

This thesis develops a new approach to the formal definition of a truth predicate that allows a consistent, semantically closed definition within classical logic. The approach is built on an analysis of structural properties of languages that make Liar Sentences and the paradoxical argument possible. By focusing on these conditions, standard formal definitions of semantics are shown to impose systematic limitations on the definition of formal truth predicates.

The alternative approach to the formal definition of truth is developed by analysing our intuitive procedure for evaluating the truth value of sentences like "P is true". It is observed that the standard procedure breaks down in the case of the Liar Paradox as a side effect of the patterns of naming or reference necessary to the definition of Truth as a predicate. This means there are two ways in which a sentence like "P is true" can be not true, which requires that the T-Schema be modified for such sentences.

By modifying the T-Schema, and taking seriously the effects of the patterns of naming/ reference on truth values, the new approach to the definition of truth is developed. Formal truth definitions within classical logic are constructed that provide an explicit and adequate truth definition for their own language, every sentence within the languages has a truth value, and there is no Strengthened Liar Paradox. This approach to solving the Liar Paradox can be easily applied to a very wide range of languages, including natural languages.

The work contained in this Thesis is my own and, unless otherwise attributed, is not the result of any joint effort or research.

Ryan Young

ACKNOWLEDGEMENT. Firstly, I would like to thank Peter Roeper for his time, effort, patience and attention to detail. His willingness to pause and take time to consider and think through all of my ideas, even the worst ones, has been greatly appreciated. While our discussions tended to go round in circles, he succeeded in bringing order and rigour to my work which I would have struggled to achieve otherwise.

I am very grateful to Ingolf Max, and the other academics at the Institut für Logik und Wissenschaftstheorie at the University of Leipzig, for their time and support during my stay there. I learnt a lot and my ideas matured significantly in the stimulating environment they have created.

I am also very grateful to Brian Garrett, Daniel Stoljar, Margaret Brown and many others for their ability at various times to put up with me and make sure that I successfully got to the end of this thesis.

Finally, I have to thank Rebecca, Gwyllym and Sebastian for being there, not complaining and distracting me when I needed it. Their inspiration and ability to keep me grounded has contributed immeasurably to the quality of the work.

Contents

Introduction

The Liar Paradox is one of the oldest and has been one of the most intractable paradoxes in the Western philosophical tradition. It dates back at least to the time of Aristotle, as its use is attributed to Eubilides, one of Aristotle's contemporaries and adversaries.[1] However, despite constant philosophical and logical attention since this time, and in stark contrast to other ancient paradoxes such as Zeno's Paradoxes of Motion, no consensus has been reached over how to respond to the Liar Paradox. Moreover, this lack of consensus has become more pronounced in recent times, despite the fact that more specific attention has been paid to the paradox since the middle of the twentieth century than in any previous period. The lack of consensus is not, however, due to any failure by philosophers and logicians to devise solutions since "the problem we face is far from a lack of solutions; rather, we have an overabundance of conflicting ones."[2]

These problems are on the surface surprising, since one of the defining features of the Liar Paradox is the simplicity with which it can be presented. A typical presentation of it only requires a couple of sentences, and it does not depend on any specialised or technical concepts. Moreover, an ordinary presentation is sufficient to allow one to grasp the basic mechanism by which the paradox occurs and have an idea of what is required in order to solve the paradox.

To see this, we will consider the following sentence:

1. This sentence is not true.

This sentence (1) is simple and easy to understand. However, if we make the obvious assumption that the 'This sentence' in the sentence refers to the sentence itself, we run into the following problem:

> If sentence (1) is true, then it follows that what it says must be true. What it says is that it itself is not true. Therefore, if (1) is true, it follows that (1) is not true. However, if we assume the other alternative, that sentence (1) is not true, then what sentence (1) says is in fact true, from which it follows that (1) is true. Therefore, if (1) is not true, then (1) must be true.

[1] Roy Sorenson. *A Brief History of Paradox*. Oxford: Oxford University Press, 2003, Ch 7, pp83-99
[2] Michael Glanzberg. "The Liar in Context". In: *Philosophical Studies* 103 (2001), pp. 217–251, p. 217

In the space of a few lines, we have established that if the sentence (1) is either true or not true, then it must also be the other. Thus it must be either both true and not true; or neither true nor not true. Since neither of these alternatives seems plausible, we have a paradox.

The consequences of this paradox are, if we take the conclusion established by the argument seriously, profound. For if we accept either of these conclusions, we have accepted the truth of a contradiction. In the first case it is of the form $P \wedge \neg P$ and in the second case it is of the form $\neg P \wedge \neg\neg P$. This is unacceptable on a standard understanding of reasoning. Moreover, on traditional accounts of reasoning, if there is a true contradiction, everything is provable. This means, that if the above argument is valid in English (or any fragment of English or any other language), then every sentence within English is provably true. It would therefore be impossible to use English to reason or discuss reasoning since every sentence of English can be shown to be true. If this conclusion is correct, it not only renders pointless vast amounts of writing and discussions, but on a more personal note, means that there is no point completing this thesis, since everything written in it can be proven to be both trivially true and trivially not true.

Stated in this form, the conclusion of the reasoning in the Liar Paradox cannot be correct. Something must have gone wrong in the argument, and our task is to identify the problem. A major project of this thesis is to provide a stable and coherent diagnosis of the problem, however as the vast literature on the subject demonstrates, this is not an easy task.

The Liar Paradox is not simple

As is clear from the exposition above, the Liar Paradox is simple to present, and the mechanism by which the Paradox arises seems easy to understand. We have a situation where we have a grammatically correct sentence that claims of itself that it is not true. This, however, contradicts the standard truth conditions for declarative sentences, that a declarative sentence is true if what it claims is true. This internal contradiction, by means of our intuitive assumptions about truth, produces the type of argument which is typical of the Liar Paradox: on the assumption that the relevant sentence is true, we can directly derive its untruth; and on the assumption of its untruth, we can derive its truth. This simplicity in presenting the Liar Paradox and understanding its basic mechanism, however, does not imply that the Liar is a simple paradox. Instead it tends to hide the complexity of the phenomena that surround it.

A quick survey of the range of literature on the Paradox neatly illustrates the complexity of the relevant phenomena. For not only do few authors agree on the correct solution to the Paradox, but not many more authors agree on how exactly to identify the problem posed by the Paradox. For example, some locate the problem in traditional assumptions about logic,[3] others locate the

[3]For example, see Graham Priest. "Truth and Contradiction". In: *The Philosophical Quarterly* 50.200 (2000), pp. 305–319; Graham Priest. "The Logic of Paradox". In: *The Journal of Philosophical Logic* 8 (1979), pp. 219–241; Hartry Field. "Semantic Paradoxes and Vagueness Paradoxes". In: *Liars and Heaps: New Essays on Paradox*. Ed. by

problem in our understanding of truth bearers,[4] others locate the problem in the role of context[5] or the limited applicability of ordinary assertions[6], while yet others attribute the difficulty to a faulty understanding of the pragmatics of language[7] or even take the Liar Paradox at least in some sense as evidence of the ultimate indefinability of truth.[8] There are strong intuitions behind each alternative, and the range of different alternative diagnoses illustrates the complexity of the issues that surround the Liar Paradox.

This complexity can be most clearly seen in the range of different examples of the Liar Paradox that have been uncovered by modern research. Historically, the examples of the Liar Paradox that were advanced usually involved somewhat artificial, contrived sentences. These were examples such as "This sentence is false"; or "I am now lying". Modern research, however, has shown that the Liar Paradox arises far more commonly than these rather artificial types of examples. The simplicity of the mechanism that generates the Liar Paradox ensures that it often arises where we would least like it to arise. There are two issues that have arisen in modern research which are particularly relevant.

The first was most forcefully demonstrated by Kripke in his seminal paper "Outline of a Theory of Truth",[9] although the type of example that Kripke uses pre-dates Kripke's work in the literature.[10] Kripke considered the ordinary assertion "Most of Nixon's assertions about Watergate are false."[11] While this is normally an unproblematic assertion, Kripke described a situation in which it becomes paradoxical, namely when Nixon's assertions are evenly balanced between truth and falsity except for the one assertion that the person who uttered the above statement is telling the truth. As can be easily checked, these two assertions create a two sentence version of the Liar Paradox. The lesson that Kripke drew from this was that "many, probably most, of our ordinary assertions about truth and falsity are liable, if the empirical facts are extremely unfavorable, to exhibit paradoxical features."[12] That means that, unlike the impression generated by the types of examples that were traditionally used, there is nothing inherently problematic in the sentences themselves that are used to generate the Liar Paradox.

The second issue in relation to the ubiquitousness of the Liar Paradox that has arisen in modern research can be summarised under the label of the Revenge Problem. Modern researchers routinely

J. C. Beall. Oxford: Clarendon Press, 2003, pp. 262–310

[4]For example, see Keith Simmons. "Reference and Paradox". In: *Liars and Heaps: New Essays on Paradox*. Ed. by J. C. Beall. Oxford: Clarendon Press, 2003, pp. 230–252; Alan Weir. "Token Relativism and the Liar". In: *Analysis* 60.2 (2000), pp. 156–170

[5]See Glanzberg, "The Liar in Context", see n. 2; Michael Glanzberg. "A Contextual-Hierarchical Approach to Truth and the Liar Paradox". In: *Journal of Philosophical Logic* 33 (2004), pp. 27–88

[6]See Jon Barwise and John Etchemendy. *The Liar: An Essay on Truth and Circularity*. New York: Oxford University Press, 1987

[7]See A. P. Martinich. "A Pragmatic Solution to the Liar Paradox". In: *Philosophical Studies* 43 (1983), pp. 63–67

[8]For example, see Charles S. Chihara. "Priest, the Liar, and Gödel". In: *Journal of Philosophical Logic* 13 (1984), pp. 117–124

[9]Saul Kripke. "Outline of a Theory of Truth". In: *Journal of Philosophy* 72.6 (1975), pp. 690–716

[10]For example, see L. Jonathon Cohen. "Can the Logic of Indirect Discourse be Formalised?" In: *The Journal of Symbolic Logic* 22.3 (1957), pp. 225–232

[11]Kripke, see n. 9, p. 691

[12]Kripke, see n. 9, p. 692

use the Liar Paradox itself to critique proposed solutions to the Paradox. The standard strategy works as follows: any proposed solution must draw a line between truth and untruth somewhere; once this line has been identified, one can usually construct a version of the Liar Paradox which operates on this line (whether or not it is expressed in terms of truth). Thus one can almost always provide a counterexample to a proposed solution to the Liar Paradox by reformulating the Paradox in a way that uses the distinctions that the solution proposes. In this sense, the Liar Paradox takes revenge on the proposed solution, and we have the Revenge Problem. The ease with which this can normally be done demonstrates the power of the Liar Paradox and has lead to the suggestion that ultimately the Liar Paradox is not solvable.

Paradoxical vs Non-Paradoxical

While it will be shown that such a pessimistic conclusion is not warranted, it captures the scope of the challenges presented by the Liar Paradox. Any meaningful discussion of, or solution to, the Liar Paradox must draw a clear distinction between the paradoxical cases and the non-paradoxical cases, and any proposed solution must satisfactorily deal with all paradoxical cases without incorrectly affecting any non-paradoxical cases. Both Kripke's argument and the Revenge Problem, however, demonstrate that this is not an easy distinction is to make accurately.

Kripke's argument undermines a lot of work and discussion of the Liar Paradox, since these are often characterised by discussions of Liar Sentences. While there are sentences which unquestionably trigger the Paradox, Kripke showed that "it would be fruitless to look for an *intrinsic* criterion that will enable us to sieve out ... those sentences which lead to paradox."[13] That is, there is nothing in Liar Sentences as *sentences* which can distinguish them from non-paradoxical cases. The clearest example of this are cases like Kripke's where the same sentence can be either paradoxical or non-paradoxical depending on the context. This means that any attempt to draw the distinction between the paradoxical and the non-paradoxical purely at the level of sentences cannot work. This raises the significant question of how we are to draw this distinction.

The Revenge Problem, on the other hand, raises the question of whether it is in fact possible to draw a solid distinction between the paradoxical and the non-paradoxical. As mentioned, the problem turns on the fact that once a proposed solution has distinguished between truth and untruth in any sense, the paradox can be usually formulated using this distinction. However, any proposed solution must draw the line somewhere between truth and untruth, and in doing this must develop some distinction between the paradoxical and the non-paradoxical. When the Revenge Problem bites, the line between truth and untruth is shown to be inadequate, and therefore the distinction that was made between paradoxical and non-paradoxical cannot be correct. If the Revenge Problem cannot be avoided, then it will also be impossible to make a solid distinction

[13] Kripke, see n. 9, p. 692

between the paradoxical and the non-paradoxical.

This process can be neatly illustrated by the simple solution to the Liar Paradox which proposes the introduction of a category of sentences which are neither true nor false as a solution to the Liar Paradox. This is commonly done on consideration of paradoxical sentences of the form "This sentence is false". However, as is well known, any simple version of this solution fails to account for any example which replaces "false" with "not true". That is, the solution proposes the category of 'neither true nor false', which means that the line between truth and untruth shifts to being between 'true' and 'neither true nor false'. As soon as we reformulate the Liar Paradox in terms of this distinction ("This sentence is *not true*"), the Liar Paradox arises again. The analysis of the paradox which supports this proposed solution failed to correctly draw the line between the paradoxical and the non-paradoxical, and therefore the solution failed. While this is a simple example, the frequency with which the Revenge Problem causes exactly the same problems suggests that it may in fact be impossible to definitively draw a line between the paradoxical and the non-paradoxical.

It will be demonstrated in this thesis that this suggestion can remain as a suggestion, as it is in fact possible to make a sharp distinction between the paradoxical and the non-paradoxical. However, a sharp distinction is only possible if we take Kripke's point outlined above very seriously. If we want to draw a distinction between the paradoxical and the non-paradoxical, we need to be clear about what *sort of thing* we are considering as paradoxical or not. For example, it is often assumed, either explicitly or implicitly, that in the case of the Liar Paradox it is the sentences that are paradoxical. If this is the case, then the distinction between the paradoxical and the non-paradoxical has to be a distinction between the paradoxical sentences and the non-paradoxical sentences. Kripke showed, however, that this is not possible as many perfectly satisfactory sentences can be paradoxical in unusual situations. The first challenge in drawing the required distinction is to get clear on what sort of thing we need to consider in drawing the distinction.

What *sort of thing* is paradoxical?

The problem identified by Kripke is that "it would be fruitless to look for an *intrinsic* criterion that will enable us to sieve out ... those sentences which lead to paradox."[14] That is, there is no guarantee that the sentences which are (potentially) paradoxical will share any common set of properties on which we can draw a coherent distinction between the paradoxical and the non-paradoxical and therefore on which we could base a satisfactory solution. Kripke's conclusion is based on the observation that any sentence which includes a truth ascription could be paradoxical in appropriate circumstances. There is therefore nothing different about paradoxical sentence when compared to non-paradoxical sentences, since many sentences can be both paradoxical and

[14] Kripke, see n. 9, p. 692

not paradoxical in different contexts.

Given the role that context and external situations play in Kripkean examples, some role for context has to be included in the sort of thing we want to investigate. The obvious solution to this problem would therefore be to investigate situations or states of affairs, in conjunction with sentences, as the sort of thing on which we could draw a distinction between the paradoxical and the non-paradoxical.

While it is plausible that there may be a clear distinction between paradoxical and non-paradoxical situations, any strategy of this sort runs into problems when we consider the infinitely large range of possible combinations of sentences and situations. If we consider only a sentence such as "That is not true" within English, there are infinitely many possible referents of this sentence (including itself) embedded in infinitely many possible situations. An infinite subset of these referents will in turn refer to some other sentence, or set of sentences, and in each of these there will be again infinitely possible referents of the sentences in this infinite subset. The fact that some of these referents to the sentence "That is not true" will, in the appropriate context, be paradoxical, is purely due to the combinatorics of the situation. However, identifying exactly which of these possible combinations of sentence and context are paradoxical and which are not is, at the very least, not an easy task. When we in turn consider the huge variety of sentences that we can start with, it should be clear that identifying a rule-governed method of producing all possible types of situation/sentence combinations which are paradoxical is a mammoth task, which may or may not be solvable. At the very least it is a major research project in its own right. Without such a rule-governed method, it is impossible to be sure that we are drawing the line between the paradoxical and the non-paradoxical correctly.

Moreover, even if we can draw a line between the paradoxical and the non-paradoxical on this basis, it may not be very useful in terms of grounding a solution to the Liar Paradox. The distinction between the paradoxical and the non-paradoxical will, using sentence/context sets, be largely drawn through an understanding of the combinatorics of the situation. However, we cannot change the combinatorics that lead to the Liar Paradox without a fundamental reworking of the way we use languages such as English. Requiring this as part of a solution to the Liar Paradox is neither likely to be plausible nor to gain widespread acceptance.

Given that there are problems with investigating sentences and sentence/contexts combinations, we could consider the common philosophical reflex to look at different possible truth-bearers. To take a couple of examples, we might consider looking at propositions, or sentence-tokens. Importantly, each of these requires some sort of context sensitivity in their definition. Context is often necessary to determine the proposition expressed by a sentence; and a sentence-token, as a token, necessarily belongs to a particular context. However, while each of these have been advocated as solutions to the Liar Paradox, each of these is faced with essentially similar problems.

It has been argued that taking propositions as primary truth bearers can solve the problem since Liar Sentences either do not express a proposition,[15] or express a false proposition.[16] These solutions naturally face the standard problems that all accounts of propositions face. However, independently of these problems, turning to propositions as part of an attempt to better understand the Liar Paradox produces the same problems as sentences. For whether we consider the sentence itself, or the proposition expressed by "Most of Nixon's assertions about Watergate are false"[17], nothing intrinsic can be found which separates paradoxical from non-paradoxical use of that sentence. The proposition expressed will be the same whether it is embedded within a paradoxical context or not. Considering propositions does not help us draw any line between paradoxical cases and non-paradoxical cases.

The alternative move to take sentence tokens as the basic truth bearer does not help either.[18] It would be entirely consistent on this account that one assertion of "Most of Nixon's assertions about Watergate are false" is true, and while another is false because it is paradoxical. However, this is not enough since there is nothing in the sentence tokens themselves which differentiate paradoxical tokens from non-paradoxical tokens. We still need something about the context, but that opens up the problems outlined above in the case of sentences. Thus considering sentence tokens does not in itself add any more information to an attempt to differentiate between the paradoxical and the non-paradoxical.

Thus focusing our investigation on different possible truth bears does not help us distinguish the paradoxical from the non-paradoxical since there is nothing in the truth bearers themselves which reflect the difference between the paradoxical Kripkean cases from non-paradoxical cases. Adding the relevant context directly into our considerations, however, adds enormous complexity, which brings in questions about the decidability of the problem. Moreover, a distinction on these grounds would be based on a consideration of the combinatorics of language reference, which is not something that can be plausibly altered for natural languages.

Languages

Looking at the alternatives to taking sentences as the focus of the investigation into the Liar Paradox examined here, any alternative truth-bearer to sentences runs into the same problems as Kripke showed existed for sentences, since it is not possible to distinguish between paradoxical and non-paradoxical cases without referring to external context. However, including external context in the basic unit of investigation poses problems of a different sort, the immense complexity that arises on this approach undermines attempts at progress. It is very difficult to be sure that the

[15] See Laurence Goldstein. "A Unified Solution to Some Paradoxes". In: *Proceedings of the Aristotelian Society* 100 (2000), pp. 53–74

[16] See E. Mills. "A Simple Solution to the Liar". In: *Philosophical Studies* 89 (1998), pp. 197–212

[17] Kripke, see n. 9, p. 691

[18] For one example, see Alan Weir. "Token Relativism and the Liar". In: *Analysis* 60.2 (2000), pp. 156–170

line between the paradoxical and the non-paradoxical has been drawn correctly.

Each of these alternatives focus our investigations on actual paradoxical cases. However, the structure of the paradoxical argument in the Liar Paradox suggests that this is not necessary. The Liar Paradox argument always begins with a hypothetical assumption: "Suppose this particular sentence is true, ..." of "If that particular sentence is not true, ...". This means that the Liar Paradox only requires the potential existence of a paradoxical case to arise. So long as, for example, we *can* construct an appropriate example of the Liar Paradox in English, then the consequences posed by the Paradox above need to be faced. We cannot, for example, deal with the Liar Paradox by banning the use of paradoxical sentences or statements. The Liar Paradox bites if these sentences are merely constructable.

Given that the Liar Paradox turns on the possibility of certain types of sentences or situations, this suggests that the focus our investigations into the Liar Paradox should be on the conditions which make these sentences and the Liar Paradox possible, rather than actual cases of the Paradox. This would mean that our focus should be on the linguistic structures that make paradoxical sentences possible, and the reasoning which generates the Paradox. That is, if we are to take this suggestion seriously, we should be investigating languages, rather than sentences or propositions or situations.

This idea that the focus of investigations into the Liar Paradox should be focused on languages is supported by the brief account of the Paradox given above. As stated above, the most serious consequence of taking the Liar Paradox seriously is that our ability to use languages to express truth is called into question. This consequence is at the level of languages, and moreover any serious attempt to solve the Liar Paradox therefore either seeks to revolutionise our understanding of languages, or constructs a new formal language to demonstrate why it is not a problem. Since languages are central to solving the Liar Paradox, it is reasonable that they should become the focus of our attempts to understand and explain the paradox. That is, we should focus on understanding the conditions set up by languages which make the Liar Paradox possible.

This approach is further supported by the simple observation that in general there is a much clearer distinction between languages which are affected by the Liar Paradox and those which are not. To take slightly extreme examples, English is obviously affected in some sense, while Classical Sentential Logic is not. The fact that the Liar Paradox does not affect Classical Sentential Logic is not a deep property of that language, it simply does not possess sufficiently rich vocabulary. However this is a property that does not depend in any way on context. Similarly, the fact that English is affected is also context independent. No matter how the world is, the linguistic structures in English allow the construction of liar sentences, and once these are constructable we can run the Liar Paradox argument. This existence of a context independent distinction between the paradoxical and the non-paradoxical allows us to define the Liar Paradox much more easily

and effectively.

Each of these reasons strongly suggests that languages should be the primary focus in investigations into the Liar Paradox. They define the conditions which make the Paradox possible, they are the main focus in developing solutions and there seems to be a clear distinction between languages that are affected by the Liar Paradox and those which are not affected.

How are Languages affected by the Liar Paradox?

While we can pick a couple of examples to show that there should be a clear distinction between languages which are affected by the Paradox and those that are not, this distinction cannot be made without a clear idea of what it actually means for a language to be affected by the Liar Paradox. For there are significant differences between the ways that languages can be affected by the Liar Paradox.

The first difference is that some languages may be necessarily affected while others are only contingently affected. Any language in which it is possible to construct examples like "This sentence is not true" case above, will be necessarily affected. No matter how the world is, no matter what facts are true, the language will be affected by the Liar Paradox. However, there will be languages where the problematic sentences are ones like Kripke's examples, where they are paradoxical only if certain facts are true. Kripke's example is only paradoxical if Nixon's assertions about Watergate are perfectly balanced between true and false except for the one that refers to the speaker of the statement. Languages where examples like this are the only paradoxical cases are contingently affected by the Liar Paradox.

The focus of this thesis will be on languages which are necessarily affected for a couple of reasons. Firstly, the languages we are most interested in are necessarily affected. Secondly, constructing a language which is only contingently affected is a non-trivial task, since the linguistic mechanisms which allow the paradox in contingent cases almost always allow necessary cases. Finding clear rules which disallow necessary cases but allow contingent cases is difficult, and will not help us understand or solve the paradox.

Another difference in the way that languages are affected by the Liar Paradox can be illustrated by a couple of sketched cases. Suppose we take the language formed by adding a suitable mechanism for referring to sentences (such as a Gödel Numbering) to a Classical logic and by adding the T-Schema as an axiom schema. So long as there is at least one paradoxical sentence (the Diagonal Lemma is taken to demonstrate that this will be the case) a contradiction will be derivable in the language, and everything will be provable. This means that the language is trivialised - everything is provable and therefore one can no longer assert anything meaningful in it. If we consider English, on the other hand, it is affected by the Liar Paradox in the sense that we can construct sentences which, *under certain assumptions*, trigger the Paradox and hence from which we can derive a

contradiction. We will discuss what these assumptions might be later, but the important thing to note is that the Paradox only arises if certain non-trivial assumptions are made.

In the first case, the Liar Paradox arises purely from the definition of the language, and therefore the language is trivialised, and therefore useless. If we wish to fix the situation, it is necessary to change part of the definition of the language and therefore define a new language. Two formal languages with differing definitions and differing sets of theorems are necessarily different languages. In the second case, the paradox does not necessitate that we give up on English and define a new language to use instead. The normal response to the Liar Paradox is to question the assumptions that are necessary for the paradox, as these are not considered to be essential to the definition of English. In the first case the Liar Paradox renders the language useless, and in the second case the Liar Paradox calls into question assumptions about the language.

In order to understand this difference more clearly, it is necessary to examine a key difference between natural languages such as English, and formal languages like Classical Predicate Logic. Formal languages are normally defined so that the key principles of reasoning or logic[19] that apply to that language are a part of the explicit definition of the language. Thus, for example, the difference between the languages of Classical Predicate Logic and that of Intuitionistic Predicate Logic is not in the vocabulary or the grammatical rules on what constitutes a valid sentence (or well formed formula), but in the principles of reasoning (i.e. the axioms and /or rules of inference) defined within those languages. The particular rules of reasoning, whether defined as inference rules or axioms, are an explicit part of the definition of normal formal languages.

Natural languages, on the other hand, do not include an explicit definition of the principles of reasoning that apply to the language. If we come across a string of words such as "The dog that the ball", we can condemn it as ungrammatical and therefore not an English sentence. On the other had, if we come across an invalid argument, say "Not every dog chases balls, therefore, no cat chases balls", this is perfectly valid English. However, it is condemned on grounds of reason or logic, not on the grounds that it is not valid English. This suggests that, unlike the case of normal formal languages, principles of reasoning are not intrinsic to the definition of natural languages as one can presumably alter the principles of reasoning without altering the language.

However, it might be commented that natural languages do not have an explicit definitions of anything, so it means little to say that there is no explicit definition of the relevant principles of reasoning in a natural language. Formal languages are explicitly defined and therefore reasoning is explicitly defined; natural languages are not explicitly defined and therefore reasoning is not explicitly defined. While this is strictly true, it does not mean that we cannot draw a relevant distinction between the grammatical rules of English and the principles of reasoning that apply to English, which does not exist for formal languages. It is very rare to have a substantial argument

[19] This is primarily meant to include purely logical reasoning. That is, the reasoning which governs logical connectives and quantifiers.

about whether a particular sentence is grammatically sound, but it is far more common to have a discussion about whether a particular piece of reasoning is actually valid. If a grammatical rule is changed, it is a change to the English language. We accept that a language can change over time due to new grammatical constructions. However, if an accepted principle of reasoning changes, we take it as a change to our understanding of the language. It seems absurd to argue that a philosopher changes English when they present a new philosophy of language. There are however no rules of reasoning that we can change in the case of formal languages without changing the language.

One final challenge to the validity of this distinction is that it is only a surface distinction, and that when one considers the semantics of the relevant languages the distinction disappears. The difference between Classical Logic and Intuitionistic Logic does not only lie in the difference in the principles of reasoning, but also in the different semantic interpretation that the symbols in the different languages have. The different semantics for the different languages, it might be argued, determines the different principles of reasoning. Furthermore, in languages like English, the semantics of logical words like "If" determine the correct principles of reasoning for English, and therefore there is no difference between formal languages and natural languages like English.

While there may be a correct or true semantics for English, the lack of consensus on this matter means that if there is, we do not know what it is. The meaning of logical words such as "If" certainly fixes some rules of inference. It seems highly unlikely that we could accept that someone understands the meaning of "If" if they deny Modus Ponens. However, unlike the case with formal languages, the meaning of "If" does not fully determine the correct rules of reasoning about it. To borrow the terminology of Natural Deduction, while Modus Ponens defines the appropriate Elimination rule of "If", the correct Introduction rule or rules are not always clear. The debates over the nature of conditionals are a good illustration of this. A similar story can be told for many logical words and important concepts in English. The standard meanings of these words partially define the relevant reasoning (in a strict logical sense), but they are not completely defined. This is not the case for formal languages.

Importantly, this difference between standard formal languages and natural languages corresponds to the difference in the way that the Liar Paradox affects languages as identified above. If the principles of reasoning are an essential part of the definition of the language, then if a contradiction arises then the language typically becomes useless. If the principles of reasoning are, at least partly, a matter of interpretation or assumption, then the Liar Paradox called these into question, rather than making the language useless.

Grammar-Only vs Logical Languages

This distinction that has been identified between formal and natural languages is so useful in understanding the Liar Paradox, that we will adopt explicit terminology for the different types of language. We will label languages which explicitly include the logical principles of reasoning for that language, as "Logical Languages" and languages which do not explicitly include principles of reasoning as "Grammar-Only Languages". It should be noted that the sense of Grammar that is being invoked here is that there are rules governing what are and what are not valid sentences. These rules may rely on semantic information such as meanings of words, as they often do in natural languages. Thus a Grammar-Only language is one which has a set of rules about what are valid sentences, but does not explicitly include in its definition complete information about correct logical reasoning about or from sentences. As argued above, natural languages, such as English, count as Grammar-Only languages on this definition.

Although the distinction has been drawn between formal and natural languages, not all formal languages are necessarily logical languages, and not all natural languages are necessarily Grammar-Only Languages. Furthermore, it is often useful to treat a Logical language as having a Grammar-Only part, and a Logical part. Thus, for a formal language, the Grammar-Only part is the set of syntactic rules that determine what are allowable sentences, or well formed formulas. The Logical part are the axioms or rules of inference which determine which sentences are provable and which are not. This means, for example, that Classical and Intuitionistic logics have identical Grammar-Only parts, and only differ in the Logical part.

Importantly, this distinction between Logical and Grammar-Only Languages corresponds to different ways that a language can be affected by the Liar Paradox. A Logical Language is affected in the sense that it may or may not be trivialised by the Liar Paradox. That is, the Liar Paradox may render a Logical Language trivial (and useless) since everything is provable in that language. Since the principles of reasoning for a (fully) Logical Language are completely defined, it makes sense to ask whether or not the Liar Paradox trivialises the language, and almost always there will be a definitive answer. For Grammar-Only Languages, however, it does not make any sense to ask whether such a language is trivialised by the Liar Paradox. Such languages are only defined by a set of grammatical rules, and grammatical rules do not generate sets of provable sentences. Thus it does not make sense to say that the Liar Paradox means that everything is provable in English. Thus Grammar-Only Languages are not affected by the Liar Paradox in the sense that the Liar Paradox can trivialise them.

However, there is a real sense in which Grammar-Only Languages are affected by the Liar Paradox. Languages such as English are affected by the Liar Paradox since sentences can be constructed in such languages that, under the assumption of certain principles of reasoning about the language and terms within the language, lead to the argument typical of the Liar Paradox,

and hence a contradiction. Thus Grammar-Only Languages are affected by the Liar Paradox in the sense that sentences (or something equivalent) are constructable in the language which, under the relevant assumptions, trigger the Liar Paradox.

This distinction between Grammar-Only and Logical Languages is useful for understanding the Liar Paradox as it helps clarify the different senses in which one can say that a language can be affected by the Liar Paradox. Languages can affected by the Liar Paradox in the sense that the language is trivialised by the Paradox, or in the sense that sentences (or similar) are constructable within the language which trigger the Paradoxical reasoning under appropriate assumptions. Languages which are trivialised by the Liar Paradox must obviously be also affected in the second sense, and the appropriate assumptions must be reflected in the principles of reasoning defined within the language. Grammar-Only Languages can only be affected in the second way; while Logical Languages can be affected in both ways.

The Problem

This distinction between Grammar-Only and Logical Languages, and the corresponding distinction between the different ways that a language can be affected by the Liar Paradox helps clarify the problem posed by the Liar Paradox, and hence helps identify strategies for solving the problem. As phrased before, the problem posed by the Liar Paradox is that if the paradoxical conclusion is correct, any statement in English can be proven and therefore is true. That is, to use the term adopted in the previous section, English would be trivialised - everything would be provable in it and therefore it would be useless.

However, it was argued in the previous section that the Liar Paradox does not trivialise a Grammar-Only Language such as English, since English does not include explicit rules of inference as part of its definition. Rather, the Liar Paradox affects Grammar-Only Languages in the sense that *under certain assumptions about logical reasoning and truth*, the Paradoxical reasoning arises. The Liar Paradox only trivialises Logical Languages.

We can however take English plus the intuitively correct principles of reasoning and truth accepted in the argument presented above to be one linguistic system, say $English^*$. In the terminology adopted here, $English^*$ is a Logical Language. The paradoxical argument given above is therefore a proof that $English^*$ is trivialised by the Liar Paradox. We cannot accept the assumptions inherent in the argument without modification, since that leads to an inconsistent logico-linguistic system in which everything is provable.[20] That is, our intuitive understanding of English, Truth and the way that languages represent Truth is inconsistent.

Put in this way, the challenge of identifying the incorrect assumptions in the paradoxical argument becomes the challenge of finding a set of assumptions about truth and reasoning which

[20] Assuming a sufficiently classical logic.

satisfy two conditions. Firstly, they capture the correct concepts and correspond to our intuitions. Secondly, that the Logical Language corresponding to the linguistic system formed by adding these to English (say $English^{+}$) is not trivialised (i.e. affected in the more substantial sense) by the Liar Paradox.

The appropriateness of this way of framing the problem can be seen in the numerous attempts in the literature to define a formal solution to the Liar Paradox. The main components of a formal solution are a formal language and a truth definition for that language, and the aim is to develop such a solution which is not affected by the Liar Paradox. The aim is that a correct formal solution would demonstrate the logical structure for $English^{+}$, the correct linguistic system which includes English and the correct assumptions about reasoning and truth. That is, these formal solutions are designed to embody the correct assumptions, and to demonstrate the mechanism by which the Paradox is avoided is we adopt the correct understanding.

Unfortunately, as commented before, there exist too many solutions rather than no good ones. There exist many formal solutions to the Liar Paradox in which the formal language is not trivialised. That means philosophically that there are many candidates for $English^{+}$ which are not affected by the Liar Paradox in the sense that $English^{+}$ is not trivialised. This should mean that the challenge that we face is choosing between these competing candidates, and developing some criteria by which we can choose between them. However, while all of these formal solutions provide a Logical Language which is not affected by the Liar Paradox, it does not mean that they can all function as candidates for $English^{+}$.

In a philosophical follow-up to his formal definition of truth, Alfred Tarski offered a powerful argument that, in the terminology adopted here, any Logical Language which satisfies certain plausible assumptions is necessarily inconsistent.[21] The key assumptions in his argument were articulated by Tarski as follows:

> (I)-We have implicitly assumed that the language in which the antinomy is constructed contains, in addition to its expressions, also the names of these expressions, as well as semantic terms such as the term "*true*" referring to sentences of this language; me have also assumed that all sentences which determine the adequate usage of this term can be asserted in the language. A language with these properties will be called "*semantically closed.*"
>
> (II) We have assumed that in this language the ordinary laws of logic hold.[22]

There is a third principle that is crucial to his argument, which Tarski considered that he had established as correct, what is now known as the T-Schema. For Tarski this is expressed as: *"X is*

[21] For the clearest account see Alfred Tarski. "The Semantic Conception of Truth: and the Foundations of Semantics". In: *Philosophy and Phenomenological Research* 4.3 (1944), pp. 341–376

[22] Tarski, "The Semantic Conception of Truth", see n. 21, p. 348

true if, and only if, p."[23] Tarski uses X as a name for the sentence p. What is important to note in this context is that the first is an assumption about the Grammar-Only part of a Logical Language, the second is an assumption about correct reasoning and the third is an explicit assumption about Truth. To translate his point into our terminology, Tarski argued that any Logical Language that includes a plausible truth definition (i.e. one that satisfies the T-Schema), that satisfies the ordinary laws of logic and whose Grammar-Only part is semantically closed, is necessarily trivialised by the Liar Paradox.

$English^+$, whatever the correct logical and alethic assumptions are, is a Logical Language whose Grammar-Only part (English) is semantically closed. For this reason, Tarski thought that English was necessarily inconsistent and needed to be replaced by a scientific, logical language for correct discussion of truth.[24] This conclusion of Tarki's has justifiably been rejected by modern philosophers and logicians, but it reveals an important aspect to the challenge of providing a formal solution which is a plausible candidate for $English^+$. The Grammar-Only part of $English^+$ is semantically closed, and therefore the Grammar-Only part of any formal language which provides a formal solution must be semantically closed. Otherwise, the formal solution cannot correctly capture the logical structure of $English^+$. That is, if we want a formal solution which correctly deals with the Liar Paradox as it poses a philosophical challenge for natural languages as we use them, that solution must be semantically closed.

However the vast majority of (and arguably all) formal solutions to the Liar Paradox are not semantically closed. This means that these solutions cannot capture the correct assumptions about reasoning and truth which allow us to reason satisfactorily in English in the ways we habitually do. They can only be accepted as a valid solution if we change the way we use English, if they can be accepted at all. The aim of this thesis is to provide a new formal solution which allows a semantically closed Grammar-Only part to a Logical Language and therefore is a plausible candidate for $English^+$.

Summary of rest of the Thesis

The development of a new formal solution must be grounded in an accurate analysis of the Paradox. It was argued above that investigating languages offers a clear method for differentiating between the paradoxical and the non-paradoxical, and this differentiation will be the focus of investigations in the first chapter. Given that two different ways that languages can be affected by the Liar Paradox, it follows that there are two different distinctions that need to be drawn and we can divide this investigation into two projects. The first distinction is the line between Grammar-Only languages that are affected and those that are not. That is, we need to identify which properties a

[23] Tarski, "The Semantic Conception of Truth", see n. 21, p. 344

[24] Alfred Tarski. "The Concept of Truth in Formalized Languages". In: *Logic, Semantics and Metamathematics*. Oxford University Press, 1956, pp. 152–278, p. 165

Grammar-Only Language must possess for Liar Sentences to be constructable within that language. The first project therefore is to develop an account of the grammatical conditions necessary for Liar Sentences to be constructable in a (Grammar-Only) Language.

The second distinction is between those Logical Languages which are trivialised by the Liar Paradox and those which are not. We need to uncover which principles a Logical Language must satisfy for the Liar Paradox to trivialise that language. As noted before, every Logical Language includes a Grammar-Only part, some rules about which sentences are grammatically valid. In formal languages, these are normally purely syntactic rules. The Grammar-Only part of a Logical Language must therefore possess the properties necessary for Liar Sentences to be constructable within that language, which will be identified in the first project. This means that the second project is to provide an account of the principles of reasoning necessary for the Paradoxical reasoning to produce a contradiction and hence trivialise the relevant (Logical) language. These two projects will be tackled in the first chapter, with the aim of providing a comprehensive account of which languages are affected by the Liar Paradox (in both senses) and therefore a deeper understanding of what the Liar Paradox is.

The second chapter will examine various formal solutions that have been advanced. Tarski's work will be a particular focus, since Tarski more than anyone else is responsible for defining the parameters of the modern debates about the Liar Paradox. Moreover, the various formal solutions which have been advanced since Tarski can be understood as attempts to construct non-trivial Logical Languages which are as close to being semantically closed as possible. It will be argued that many existing formal solutions suffer from a systematic limitation. They cannot, within the logical language, assign certain sentences about Liar Sentences the correct truth status, according to the internal properties of the solutions themselves.

The aim of the third chapter is to track down the cause of this systematic limitation. The fact that there is a common limitation within existing solutions suggests either that it an unsolvable problem, or that there is a systematic flaw in the strategies that are currently being used. It is important here to note the difference between what is here defined as a logical language and a formal language. A logical language is a language which includes an explicit definition of its principles of reasoning. A formal language is a language in which a mathematical or symbolic logic is defined. Given that the formal solutions that are advanced are constructed within modern formal languages, it is vital to question whether there are assumptions within modern formal logic which are incompatible with a satisfactory solution to the Liar Paradox. It will be argued that two key semantic assumptions about formal languages, namely that truth values are completely defined with respect to a semantic model and that sentence types are the primary truth bearers, are incompatible with a satisfactory philosophical solution to the Liar Paradox.

The second half of the thesis will turn to developing a more satisfactory solution to the Liar

Paradox on the basis of the analysis in the third chapter. The fourth chapter will develop a blueprint for a more satisfactory formal truth definition by focusing on our understanding of how truth is defined in a language, and particularly the way that we evaluate the truth value of sentences which include the truth predicate. With this focus, the interaction between predicates like truth and the reference structures necessarily for the definition of a truth predicate become very significant, and it is shown that care is needed in the definition of a truth predicate. This blueprint will be developed entirely independently of any particular formal logic, and will therefore be applicable to any potential formal logic. The blueprint is shown to be successful in solving the Liar Paradox in a very appealing way in the fifth chapter, where it is implemented in a Classical Sentential Logic. This definition is provably consistent, which is a remarkable result.

The sixth chapter builds on and improves the blueprint from the fourth and fifth chapters by exploring the questions about the appropriate truth bearers that arose in the analysis in the third chapter. It is argued treating name/sentence pairs as truth bearers in formal languages allows us to improve the definition in the fifth chapter to a consistent, semantically closed formal truth definition. Once again, this is defined in a Classical Sentential Logic, and arguably provides a philosophically satisfactory formal truth definition.

The seventh chapter concludes the thesis by summarising the key features of the approach adopted in chapters four to six, and discusses its applicability to understanding truth in natural languages. It is argued that the key problem the Liar Paradox produces, that we cannot be sure our use of natural languages is consistent or coherent, is solved as this approach provides an understanding of truth which is consistent, applicable in natural languages and semantically closed.