THE POWER OF A PARADOX: THE ANCIENT AND CONTEMPORARY LIAR

THE POWER OF A PARADOX: THE ANCIENT AND CONTEMPORARY LIAR

By DANIEL COREN, B.A. (Honours)

A Thesis Submitted to the School of Graduate Studies in Partial Fulfillment of the Requirements for the Degree Master of Arts

McMaster University MASTER OF ARTS (2013) Hamilton, Ontario (Philosophy)

TITLE: The Power of a Paradox: The Ancient and Contemporary Liar AUTHOR: Daniel Coren, B.A. (Honours) (McMaster University) SUPERVISOR: David Hitchcock

Number of Pages: v, 97

Abstract

This sentence is whatever truth is *not*. The subject of this master's thesis is the power, influence, and solvability of the liar paradox. This paradox can be constructed through the application of a standard conception of truth and rules of inference are applied to sentences such as the first sentence of this abstract. The liar has been a powerful problem of philosophy for thousands of years, from its ancient origin (examined in Chapter One) to a particularly intensive period in the twentieth century featuring many ingenious but ultimately unsuccessful solutions from brilliant logicians, mathematicians and philosophers (examined in Chapter Two, Chapter Three, and Chapter Four). Most of these solutions were unsuccessful because of a recurring problem known as the liar's revenge; whatever truth is *not* includes, as it turns out, not *just* falsity, but also meaninglessness, ungroundedness, gappyness, and so on. The aim of this master's thesis is to prove that we should not consign ourselves to the admission that the liar is and always will just be a paradox, and thus unsolvable. Rather, I argue that the liar *is* solvable; I propose and defend a novel solution which is examined in detail in the latter half of Chapter Two, and throughout Chapter Three. The alternative solution I examine and endorse (in Chapter Four) is not my own, owing its origin and energetic support to Graham Priest. I argue, however, for a more qualified version of Priest's solution. I show that, even if we accept a very select few true contradictions, it should *not* be assumed that inconsistency inevitably spreads throughout other sets of sentences used to describe everyday phenomena such as motion, change, and vague predicates in the empirical world.

Acknowledgements

I am very grateful to my thesis supervisor David Hitchcock and to my second reader Nicholas Griffin, both of whom gave me very useful comments and corrections on multiple drafts, patiently answered my many questions, and gave me invaluable support and encouragement along the way. For comments on specific sections, I am indebted to Patrick Bondy, Mark Vorobej, Owen Pikkert, Matti Eklund, Joseph Sneep, Andrew Pineau, and Adam Woodcox. Thanks also to Richard Arthur for reading a paper on a related topic to that of this master's thesis and for introducing me to logic, and to Mark Johnstone for helpful advice.

I am indebted – and will always be – in a rather different way to my parents, Michael and Bernadette, for their love and constant support. Thanks also to my siblings, Lucy, Oliver and Elizabeth, for helping to preserve my sanity. Finally, thanks of a very special kind to Agnese.

Table of Contents

Introduction
What is the liar paradox?

Many sentences in natural language predicate a particular something about a particular something else (a subject). One might call such sentences ‘statements’. Statements may say something about anything: a tree, a dog, a person, a god, or even themselves. The domain of possibilities for statements is very broad indeed, but not as broad as that of all sentences. Suppose for now that one adopts an Aristotelian conception of truth and falsity: “To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true” (*Metaphysics*, 1011b25-27). Applying this classical conception of truth and falsity to statements, it follows that saying of what is that it is or of what is not that it is not constitutes the *truth condition* for statements. If this truth condition is satisfied, then the statement is *true*. Similarly, suppose that saying of what is that it is not or of what is not that it is constitutes the *falsity condition* for statements. If this falsity condition is satisfied, then the statement is *false*. For example, the statement ‘That professor is intoxicated’ is identified as true if the truth condition is satisfied, namely that the unfortunate professor in question is actually intoxicated when the sentence is uttered or written. It is false if the falsity condition is satisfied, namely that the professor in question is *not* intoxicated.

There is a peculiar minority of statements, though, which have themselves as their subject and are thus ‘self-referential statements’. One should not, of course, automatically discriminate against this syntactic minority simply because it appears peculiar. This sentence is the third of this paragraph. The previous sentence is a statement, has itself as its subject, and what it states about itself, namely the predicate of being the third of this paragraph, is true because it is in fact the third sentence of this paragraph. The truth value of that particular self-referential statement is fairly easily understood and articulated, because its truth conditions are relatively simple; all one has to do is perform the unremarkable feat of counting (or perhaps even reading) the first three sentences of this paragraph, while retaining some semblance of full consciousness. This is the twenty-seventh sentence of this thesis. The previous sentence is also easily identified, but this time it is a false statement, not a true one, because it satisfies the falsity condition of saying of what is not that it is: it is actually the seventeenth sentence of the thesis, but claims that it is the twenty-seventh. Equally innocuous are self-referential statements such as: 'This sentence has five words' (true); 'This sentence has seventeen letters' (false); 'This sentence has eight syllables' (a true statement), 'This sentence is fourth in a list' (true), 'The purpose of this sentence as an example is probably redundant' (true) and so on. It is clear, then, that many – perhaps even most – self-referential statements are not puzzling or problematic; rather they are true or false.

However, within the broad category of all statements, and within the sub-category of all self-referential statements, there is a further sub-category of those statements which say something about their own truth value. Many of the constituents of this further sub-category still seem relatively harmless, ex: 'Some English sentences are true'. That example statement is a self-referential one[1] because it predicates something of a subject which includes English sentences (though not necessarily *all* of them). Its subject is: some English sentences. Moreover, it is in fact an English sentence (in this case, an English sentence which is also a statement). Identifying the truth value of this statement is not particularly difficult, because its only truth condition is that certain English sentences are in fact true. Certain sentences in English *are* true, and so the sentence 'Some English sentences are true' says of what is that it is, making it true as

[1] The idea is not to get too pedantic here about the definition and interpretation of what constitutes a self-referential statement. My interpretation of self-reference may be ignoring some nuances which the literature on the subject would no doubt make clear. Rather, my aim is simply to show that there are many clearly self-referential statements in natural language which are completely innocuous and not the least bit paradoxical.

well. It is this sub-category, though, which contains the infamous liar sentence[2]: (L1) 'This sentence is false'. Identifying the truth value of this sentence, also a statement, is more difficult than in the case of (S1) 'Some English sentences are true', because of three important differences: first, the only truth condition and falsity condition for L1 is contained in *itself*, whereas the truth condition for S1 is contained in *other* sentences; second, L1 states its own falsity, whereas S1 states the truth of other sentences; and third, the satisfaction of the truth condition for S1 results in S1 being true (and the satisfaction of the falsity condition for S1 would result in it being false), whereas the satisfaction of the *truth* condition for L1 results in L1 being *false*, and the satisfaction of the *falsity* condition of L1 results in L1 being *true*. It follows, therefore, that L1 is false if it is true *and* true if it is false.

It has been shown that L1 is true if it is false and false if it is true. One might ask at this point, though: So what? Why does this make L1 problematic or puzzling? There is a problem with L1 if it is presupposed that all sentences, or even simply all statements, must be either true or false (and cannot be *neither* true nor false). This is called the principle of bivalence. But if the first option is the case, and L1 is true, then it has been established that L1 is false, and if the second option is the case, and L1 is true, then it is false. Thus by constructive dilemma L1 is both true *and* false. Saying that L1 is neither true nor false would not seem much better. This contradicts the principle of bivalence. Now the natural question is: Why not just throw out the principle of bivalence, allowing some sentences to be *neither* true nor false? Suppose, then, that the concepts truth and falsity are not contradictory values but rather merely contrary ones; that is, they are mutually *exclusive* but not *exhaustive* of the domain of possibilities. Therefore, a

[2] While the classical liar sentence is a self-referential one, there are various forms of the liar which do not require self-reference at all, such as the following pair of sentences: 'The next sentence is true. The previous sentence is false'. Assigning a consistent truth value to this pair is just as difficult as assigning a consistent truth value to the classical liar sentence. For a much more rigorous discussion of the role – or lack thereof – of self-reference in semantic paradoxes, see Yablo (1985), Yablo (1993) and Yablo (2006).

sentence can be true, false, or neither true nor false, allowing L1 to fall into the category of having a value which is neither true nor false. The object-language thesis corresponding to the principle of bivalence is a central axiom of classical logic, whose original function was one of Aristotle's three central axioms of human reason: the law of excluded middle, which states that for any sentence α, either α or not-α is the case (an inclusive disjunction, in which *at least* one of the two values must hold).

Even after this move, the difficulties seem to continue because of sentences such as L2: This sentence is not true. If L2 is true, then it is not true. If L2 is not true, then it is true. This time, asserting that these problematic sentences fit into some sort of gap between the two given values seems much less useful, since it is difficult to see how there could be anything in between truth and untruth. Truth is the negation of untruth, and untruth is the negation of truth. So if we use the same strategy as the strategy we used to deal with L1, and say that L2 is neither true nor not true, then it turns out that L2 is both not true and not not true. But this is surely a problem (whether or not we use a different principle, called the principle of double negation, to reduce 'not not true' to 'true'). It's a problem because of another central principle of classical logic, which also used to be – in the context of his metaphysics, though, rather than just his system of logic – one of the three fundamental Aristotelian axioms: the law of non-contradiction (LNC) which stipulates that for any sentence α, it cannot be the case that both α and its negation not-α. So it's not a legitimate possibility for L2 to fit into an assignment between the two values. The contradiction, then, seems to be inescapable. If a contradiction is true, then various systems of logic including classical logic and intuitionist logic hold that it must follow from this true contradiction that *every* sentence is true (known affectionately today as 'the principle of explosion', but for centuries under its less accessible Latin tag 'ex falso quodlibet'). Anything can be derived from a contradiction, and ‘anything’ is a worryingly substantial group, as it includes the truth of any sentence, including all conjunctions consisting of a sentence α and its negation *not* α, no matter how nonsensical or bizarre.

But in order to understand why the liar paradox has received so much attention over the past couple of thousand years, it is crucial to emphasize what makes the contradiction resulting from the liar paradox *especially* salient and problematic, other than the fact that it seems inescapable, is that it involves the fundamental notions of truth and non-truth not just indirectly but also directly. A puzzle is called a paradox if it uses apparently acceptable premises to arrive by apparently indisputable inferences at an apparently unacceptable conclusion – the conclusion is usually unacceptable because it involves a contradiction. The paradox of the stone, for instance, seems to show that an omnipotent god both is and is not omnipotent, by asking whether such a being could create a stone so heavy that he could not lift it. If there were a bizarre paradox which showed through apparently acceptable reasoning that the sentence ‘That graduate student is inebriated’ is inseparable from – and in fact necessarily results in – its contradictory sentence ‘It is not the case that that graduate student is inebriated’, the contradiction might still be puzzling but would not be nearly as worrisome as that of the liar paradox. This is because a lucid understanding of the concept of truth is much more central in philosophy than a lucid understanding of the concept of intoxicated graduate students, or even perhaps the concept of an omnipotent god. One might suppose that there is something inherently problematic about sentence entailing its negation, but this is not so[3]. The real problem arises when the rejection of a sentence because it entails a contradiction commits one to a sentence that also entails a contradiction. Then one is committed to a contradiction on the basis simply of whatever

[3] A brilliant example of this fact can be found in Euclid’s proof that there is no largest prime number, which works by inferring from the proposition that some number *k* is the largest prime number that *k* is not the largest prime number. See Euclid’s Elements Book IX Proposition 20.

assumptions and rules of inference were used in generating the two contradictions (other than assuming the sentence and then assuming its contradictory). This muddles the classical conception of truth mentioned in the first paragraph of this introduction (and indeed most conceptions of truth presented since then, paraconsistent systems of logic aside). Such a sentence would indirectly involve truth and non-truth, because the reason concepts such as that of a 'contradiction' and a 'contradictory sentence' exist, the reason principles such as non-contradiction exist, is that sentences are supposed to correspond to what goes on in reality (in the world). And as any realist will argue, (including Aristotle), things are as they are, and things are not as they are not. A pair of contradictory sentences *p* and *not p*, by definition, say that a thing both is and is not. But L2 is particularly important because it not only *indirectly* involves the concept of truth and its negation, but also *directly* involves truth and its negation. The trouble is that truth is a concept at the center of philosophy, and especially at the center of most systems of logic and areas of scientific inquiry. Therefore the liar paradox consists of an apparently perfectly legitimate sentence which, when combined with a standard interpretation of truth along with uncontroversial rules of inferences, seems to result in a highly problematic and inescapable contradiction.

Chapter One
What was the liar originally supposed to demonstrate?

The liar is one of seven paradoxes which are credited to a 4th century BC Megarian logician and contemporary of Aristotle named Eubulides:

> To the school of Euclides belongs Eubulides of Miletus, the author of many dialectical arguments in an interrogatory form, namely, *The Liar, The Disguised, Electra, The Veiled Figure, The Sorites, The Horned One,* and *The Bald Head.* (Diogenes Laertius, Lives II 108 Trans. Hicks 1925)

My goal in this chapter is to uncover precisely what the liar paradox was originally supposed to demonstrate, and how it was originally intended to be solved (if at all). Achieving this goal seems worthwhile for two main reasons: first, from the standpoint of the history of philosophy, especially the history of logic, many of the great philosophical minds throughout history have struggled with the ancient sophisms and issues intimately associated with these sophisms. Second, understanding what the liar paradox was originally designed to accomplish and how it was meant to be solved may shed some light on how one should – and should not – go about attempting to solve it in contemporary discussion.

But what is the ideal strategy which should be put into action in order to achieve this goal? Well, the fact that he created some fascinating paradoxes is virtually all that is known about Eubulides, except for the interesting detail that he "kept up a controversy with Aristotle and said much to discredit him" (Diogenes Laertius, Lives II 109 Trans. Hicks 1925). This controversy between Eubulides and Aristotle may help form the foundation for a relatively plausible conjecture that the Eubulidean paradoxes were used to attack important Aristotelian principles. Moline (1969) argues, for example, that the controversy between Aristotle and Eubulides, together with the fact that Aristotle does address one of the paradoxes by name[4] and

[4] For the Hooded or Veiled man (by name), see *Sophistic Refutations* 179a30 ff.

discusses the content of several others[5], makes it likely that the sorites was used to attack the Aristotelian doctrine of the mean in the *Nicomachean Ethics*. Wheeler (1983) argues that, because Eubulides was from the Megarian school, a group for which there is at least some reliable evidence to suggest that they were followers of Parmenides, and given Wheeler's fairly detailed analysis of all four paradoxes which seems to show how they have implications supporting Eleatic principles, it is a reasonable conjecture that the Eubulidean paradoxes were originally intended to defend Parmenidean conclusions. So on Wheeler's account, Eubulides and the Megarians supported the Parmenidean view that "negative statements are indeterminate and indefinite and so can't state how things are", and so "the liar was advanced to show that non-being cannot be coherently spoken of, that we cannot sensibly say what is not" (Wheeler 1983, pp 3-4). Wheeler may well have been onto something by emphasizing the pattern that in all the Eubulidean paradoxes a central role is played by the concept of negation, and it certainly seems possible that this pattern had something to do with the Megarian support for Eleatic principles.

So Wheeler's argumentation is undeniably astute and reasonable, and Moline's conjecture is certainly plausible. However, given the dearth of actual primary source material on the subject of the Megarians, even the most reasonable and astute examination which focuses primarily on historical context is doomed to be more resourceful than plausible. The strategy I will carry out will focus on the semantic context first and the historical context second, using the scanty but important historical evidence as a background. By 'semantic context' I mean the patterns in the presuppositions, structure, and implications of all the Eubulidean paradoxes; I will apply these patterns to show what the liar paradox was originally supposed to demonstrate. It is, of course, essential to bear in mind that all these paradoxes were created at the same time by the

[5] For the Electra (but not by that name) and argumentation relevant to its solution, see *Sophistic Refutations* 179a35,170bl2 ff,171al8 ff,179b29 ff, 175b28 ff, 181a7 ff, *Posterior Analytics* 71a25 ff. For the Liar (by name), see *Sophistic Refutations* 180a23-181bl3, particularly 180b2-7

same person, but the identity of that person and his school of philosophy is largely an unsolvable mystery, only because of the lack of primary source material and historical evidence. The only apparently unsolvable mystery about the Eubulidean paradoxes themselves, though, is contained in their respective solutions. The content of the paradoxes themselves is comparatively clear.

So what exactly did these paradoxes originally look like? Their original form, as constructed by Eubulides, has not been preserved. Three crucial observations and felicitous coincidences, though, make accurate reconstructions of their original form highly feasible: first, though sophisms (and those who propounded them) were often the subject of derision, they were taken quite seriously by some important ancient philosophers, and thus were fairly often involved in the discussion and testimony of several ancient writers – the sorites and the liar seem to pop up especially often[6]. Chrysippus, the leading Stoic philosopher, apparently wrote at least ten books on the subject of the liar paradox![7] Second, the Megarians, including Eubulides, were interested in the logical identity and significance of whole sentences, unlike Aristotle's interest in the logic of predicates and terms[8]. Third and perhaps most importantly, while the Megarians were very interested and involved in dialectic, and Diogenes specifically reports that Eubulides' sophisms were "arguments in dialectics"[9], what is more perhaps more telling is that the R.D. Hicks translation uses the specific word 'interrogatory' to describe the dialectical arguments propounded by Eubulides, suggesting a question-answer structure in the original form of the liar and the other sophisms.

---

[6] For the liar, see Cicero, Academica, II. 96; for the sorites/ the bald man see Diogenes Laertius VII. 82. Cicero, Academia, II. 49. Horace, Epistulae, II 1(45); for the hooded man/ unnoticed man/ Electra see Lucian, Vitarum Auctio, 22; for the horned man, see Diogenes Laertius, VII. 187.

[7] This fact is listed in the catalogue of his books provided by Diogenes Laertius, Lives VII. 196-97. (in both the C.D. Yonge and R.D. Hicks translation)

[8] For commentary on this important difference between Megarian-Stoic logic and Aristotelian logic, see for example Lukasiewicz (1951) p21-2.

[9] Diogenes Laertius, Lives II. 108; and Sextus Empiricus confirms the importance of sophisms to dialecticians in *Outlines of Scepticism* II. 229: "...those who extol dialectic say that it [the question of sophisms] is indispensable for their solution". (Trans. Hicks)

Using the evidence which can be gathered from the ancient sources (mostly later writers of antiquity) most scholars agree that the hooded man, the unnoticed man, and the Electra paradoxes are merely variations of the same Electra paradox[10]. Kneale and Kneale articulate the paradox as follows: “You say that you know your brother. But that man who came in just now with his head covered is your brother, and you did not know him” (Kneale and Kneale 1962 p114). What I am interested in, though, is the most accurate reconstruction of the original Eubulidean form of the paradox; it is this original form which I will analyze and apply to other Eubulidean paradoxes, especially the liar. In the chapter of the *Sophistical Refutations* immediately preceding the chapter which contains Aristotle’s discussion of the liar paradox, Aristotle quotes the questions involved in the hooded man (or in this case, the ‘masked man’) paradox, equating it with several other problems which he believes can all be solved in the same way. It is worth quoting all the relevant, disjointed passages because Aristotle was a contemporary of Eubulides and was evidently aware of most of the Eubulidean paradoxes, thus providing a more accurate depiction of the original structure of the paradox:

> All arguments such as the following depend upon accident. ‘Do you know what I am going to ask you?’ ‘Do you know the man who is approaching’, or ‘the man in the mask?’...nor in the case of a man approaching, or wearing a mask, is to be approaching the same thing as to be Coriscus [a student of Plato and friend of Aristotle], so that if I know Coriscus, but do not know the man who is approaching, it still isn’t the case that I both know and do not know the same man... Some solve these by demolishing the question; for they say that it is possible to know and not to know the same thing, only not in the same respect; accordingly, when they don’t know the man who is coming towards them, but do know Coriscus, they assert that they do know and don’t know the same object, but not in the same respect... for he both knows that Coriscus is Coriscus and that the approaching figure is approaching...But as to the approaching figure and Coriscus he knows both that it is approaching and that it is Coriscus (*Sophistical Refutations* 179a30-b33)

It is clear from these disjointed passages that Aristotle was well aware of hooded man type paradoxes, and that these sophisms were propounded in an interrogative form, involving a questioner asking questions and a respondent providing answers. It is also particularly useful that

[10] For examples of such ancient sources, see Aristotle in *Sophistical Refutations* 179a30, and Diogenes Laertius *Lives* VII; for contemporary scholars who agree on this point, see Kneale and Kneale (1962) p114, Priest (2002), and Wheeler (1983)

Aristotle elsewhere mentions that paradoxes of this sort "made two questions into one question", making use of "the fallacy that turns upon homonymy and ambiguity", in which "it is possible for it to be true to say 'Yes' or 'No' without qualification...but still one should not answer them with a single answer; for that is the death of argument. Rather, it is as though different things had actually had the same name applied to them" (*Sophistical Refutations* 175b40-1; 176a10-14). This suggests that the original form of the paradox was concise, consisting of just one or two questions. This is consistent with the form of many other ancient sophisms, and with Aristotle's commentary. It is generally much easier for a sophism to appear convincingly problematic to the respondent if it allows the respondent little time to contemplate the true nature of the questions, and the most precise way of answering these questions – or the most precise way of separating a conflated question or concept into its component parts and clearing up the ambiguity, as Aristotle notes repeatedly. It seems likely, then, that the original form of the paradox ran as follows:

Questioner: "Do you know the man who is approaching?" **Respondent**: "No".
Questioner: "Do you know your brother?" **Respondent**: "Yes".
Questioner: "But the man approaching is your brother".

This paradox has not received nearly as much attention as the heap and the liar, but the little attention it *has* received has been rigorous[11]. Though this reconstruction of the original form of the paradox might make both the parties involved appear rather simple, it seems to illustrate more clearly exactly which problems and issues are raised by the contradiction it apparently creates. Usually this paradox is represented as having the first and second sentences in reverse order in, for instance, Kneale and Kneale (1962) p114, Wheeler (1983), and Priest (2002), but this seems unlikely for a rather simple reason: in a real human dialogue, beginning by asking if a person knows her brother, and then asking if she knows the man approaching will suggest to her that the man approaching is actually her brother.

[11] See for example, Priest (2002)

In this interrogative form, the argument presupposes that there is only a simple ‘yes’ or ‘no’ answer to each question. The respondent can either give a positive answer or a negative one, one which assents to the sentence or one which negates it, but the respondent cannot – at least in the context of this conversation – give a more qualified, considered answer which neither strictly assents to the sentence nor strictly negates it. One might quibble about the fact that the respondent does have the *ability* to give an answer which is not either yes or no, but the essential observation remains: the key to the paradox producing an apparent contradiction is in the simplistic yes or no answers. These simplistic yes or no answers represent two contradictory options: the respondent either *knows*, or she does *not* know (her brother and the man approaching her); the answer ‘no’ represents negation (that the respondent does not know her brother or does not know the man approaching her), and the answer ‘yes’ represents assent or confirmation (that the respondent does know her brother or the man approaching her). If the respondent had answered the question ‘Do you know that man approaching?’ by saying ‘Right now I can’t make out the identity of the man approaching, but when I identify him I can answer your question’, the questioner would have a much more difficult time finding a contradiction between the respondent’s answers. It seems clear, then, that this paradox hinges on the dubious presupposition that ‘yes’ or ‘no’ answers are the only answers possible, and thus exhaustive of the domain of possibilities. That is, for each of the two sentences presented to the respondent, there are two possibilities for the value of the sentence p given as an answer: either p or $\sim p$, the law of the excluded middle. The possibility that the respondent neither assents to the truth of the sentence nor to the truth of its negation is essential to the solution of the ‘paradox’, that ‘yes’ and ‘no’ answers are not exhaustive, and yet this possibility implicitly dismissed by the paradox.

One might jump to the ostensibly reasonable conclusion that the hooded man paradox (and its variations) was originally intended to attack the legitimacy of LEM, which functions as an axiom in Aristotelian metaphysics, i.e. a fundamental principle that anyone who is going to learn anything must grasp (and still fundamental in the calculus of modern classical logic, and in most standard conceptions of metaphysics): something either is or is not. Along with the law of non-contradiction, which says that nothing can both be and not be at the same time and in the same respect, and the law of identity, which says that something is what it is, the law of excluded is a fundamental law of thought for Aristotle. More precisely, Aristotle acutely formulates the law of excluded middle as follows: "...there cannot be an intermediate between contradictories, but of one subject we must either affirm or deny any one predicate" (*Metaphysics* 1011b23-24). Upon closer inspection, though, this initial conjecture loses some credibility for two chief reasons: first, the fact that there *is* an error in the reasoning process of the paradox appears intuitively clear. Second, the *nature* of this error does not seem to point to a problem with the law of excluded middle itself, but rather to a problem with the *way* the law of excluded middle is being applied. The reason that there is a problem with the application of the law in this case is that the sentences to which the law is being applied are significantly ambiguous. They are significantly ambiguous because there is more than one distinct interpretation of the meaning of each sentence, and it is not entirely clear which meaning is the relevant, correct, or intended one. Take the first question put to the respondent in the paradox, 'Do you know the man who is approaching?' To declare the truth of the *negation* of this specific sentence by answering 'No' to the question, one may be stating one of two things, or both: 1) that one cannot identify enough of the details of the physical appearance – especially the face – of the man in question with any certainty (because he is too far away, is wearing a hood, or his face is obscured in the sunlight, or any number of reasons); 2) that one can in fact identify the details of the physical appearance of the man, but that the man as he is currently 'known' to the respondent (through physical appearance alone) does not seem to be anyone that the respondent knows. So in answering 'no', the respondent is unable to specify which of 1) and 2) she means. It is not the facts themselves which are individually ambiguous, though, but rather the way the facts are being described and examined.

The paradox of 'the [h]orned one', as Diogenes calls it, seems to have received the least attention out of all the Eubulidean paradoxes, both in ancient and contemporary discussion (Lives II. 108 Trans. Hicks). Here is the relevant ancient quotation: "if you have not lost a thing, you have it; but you have not lost horns; therefore, you have horns" (Lives VII. 187 Trans. Hicks). Here is a reconstruction of the paradox which fits the interrogative description of all the Eubulidean sophisms as described in the R.D. Hicks translation, as well as Aristotle's description of how these sophisms were structured:

Questioner: "Does whoever has not lost something have it? **Respondent:** "Yes."
Questioner: "Have you lost horns?" **Respondent:** "No."
Questioner: "So you have horns". **Respondent:** "Rubbish."

The first question put to the respondent is obviously the crucial one. This question presupposes that there are two possibilities which are supposedly exhaustive: for any object *x*, either one has lost *x*, or one currently possesses *x*. But these two possibilities are *not*, in fact, exhaustive, since it is impossible to lose *x* without already possessing *x*. These possibilities are *exclusive*, but they are not exhaustive. Not having lost *x* is not a sufficient condition for having *x*. Having lost *x* is a sufficient condition for having had *x* at some point in time prior to losing *x*. We might be tempted to say, then, that the second question put to the respondent shows that applying the law of excluded middle is impossible, since neither having lost *x* nor not having lost *x* seem to accurately describe what is the case in reality. The original form of this paradox, though, was probably meant to illustrate something very similar to what the hooded man paradox was meant to show: complications which can result from the application of LEM to conversations, arguments and individual sentences in natural language and even everyday discussion. That is, it was intended to demonstrate that viewing all sentences *p* as universally always either *p* or ~*p* is a generalization more problematic than it might at first appear. And, like the hooded man paradox, it seems much more likely that the horns paradox was intended to qualify the application of LEM rather than attack the legitimacy of the law itself. This is because it is very clear that there *is* an error in the reasoning process of the paradox, and the broad nature of that error is easy to identify through a little careful analysis: one cannot lose what one never had in the first place. As will be shown in the analysis of the next Eubulidean paradox, though, the distinction between what is merely an illustration of the complexities of the application of LEM and what is *not* merely an illustration of complexities (but rather an attack on LEM itself) soon becomes a murky one.

The paradox of the heap attracted considerable attention in ancient philosophy[12], and seems to be an even more popular topic of discussion in contemporary philosophy; the issue of vagueness in logic and semantics is today considered a very important one and has received a great deal of rigorous attention, such as in Williamson (1994, 2000), Sorensen (1988, 2001), and Tye (1994). The bald man paradox and the heap may have had slightly different structures, but seem to have been considered essentially the same paradox in ancient sources, known as 'little by little' arguments or simply as the 'sorites' (from the ancient Greek word for 'heaper', and 'soros' the word for 'heap'). Today, the bald man and heap paradoxes are considered the same problem: "The Bald Man, or the Heap: Would you say that a man was bald if he had only one hair? Yes. Would you say that a man was bald if he had only two hairs? Yes. Would you..., etc.

[12] See, for instance Sextus Empiricus, *Outlines of Scepticism*, II. 253; III. 80, 261

Then where do you draw the line?" (Kneale and Kneale 1962, p114) This formulation, though lacking in precision, captures the basic structure of the paradox: a line of questioning takes a familiar predicate and, little by little, shows how the borderline between that predicate and its negation can be stretched until the predicate is beyond recognition. It is perhaps clearer in the case of the heap than in that of the horned-man and the horns that this paradox cannot function without the 'yes' or 'no' answers to the stream of questions, the statement of the truth or the statement of the negation of each individual sentence. Still, since Galen provides a very precise ancient formulation of the paradox along with some lucid commentary, it is worth examining the following passage to get an accurate picture of the ancient, dialectic formulation:

> Doubt and confusion enter into many things which relate to the doings of men in spite of the fact that knowledge of these things is obvious and plain. There are some dogmatists and logicians who call the argument expressing this doubt 'Sorites' after the matter which first gave rise to this question, I mean the heap. Other people call it the Little-by-little Argument... Wherefore I say: tell me, do you think that a single grain of wheat is a heap? Thereupon you say No. Then I say: what do you say about 2 grains? For it is my purpose to ask you questions in succession, and if you do not admit that 2 grains are a heap then I shall ask you about 3 grains. Then I shall proceed to interrogate you further with respect to 4 grains, then 5 and 6 and 7 and 8; and I think you will say that none of these is a heap. Also 9 and 10 and 11 are not a heap...It is not possible for you to say with regard to any one of those numbers that it constitutes a heap...And I know of nothing worse and more absurd than that the being and not-being of a heap is determined by a grain of corn...And by reason of this denial the heap is proved to be non-existent, because of this pretty sophism (Galen, *On medical experience* 16.1-17.3 from Long and Sedley)

If at any point the respondent answers 'yes' to *any* one of the questions such as 'is 8 grains of wheat a heap?', she will be strangely stating that 7 grains of wheat is *not* a heap but 8 grains *is* a heap. It is this Eubulidean paradox which seems to blur the borderline between a relatively innocuous illustration of the *complexities* in the application of LEM, and an actual *attack* on the legitimacy of LEM itself. Like the hooded man and the horns, it is clear that there *is* an error in the reasoning process, because predicates like 'bald' and 'tall' do seem to fairly accurately pick out things in the 'real world'. The difference between these predicates and their negation – or at the very least, the existence *of* these predicates and their negation – seems intuitively evident; most people can rather easily recognize baldness, and can differentiate between a man who is bald and one who is not. There are, of course, men who one would say are ‘balding’ rather than bald or not bald, but in this case there is still a clear negation of this predicate, namely ‘not balding’.

There is, however, an extremely important difference between the heap on the one hand and the hooded man and horns on the other: the actual nature of the error in the reasoning process in the heap is nowhere near as easily identifiable and solvable as that of the hooded man and the horns. This claim is supported by the fact that in contemporary philosophy, nearly two-and-a-half millennia after the inception of the Eubulidean paradoxes, there are at least four distinct types of contemporary solutions to the problem (with the fourth one denying that there is any significant error in the reasoning process altogether): 1) there are ideal language approaches, such as in Russell (1923) and Quine (1981), which claim that the vagueness apparent in much of natural language is entirely eliminable, and essentially denies that classical logic applies to vague predicates susceptible to sorites-type problems; 2) the epistemicist approach, such as in Williamson (1994, 2000), Sorensen (1988, 2000), and Rescher (2009), attacks one or more of the premises in the paradox, claiming that the vagueness of sorites-susceptible predicates is an epistemic issue but *not* a semantic one, an indication of the limitations of human knowledge rather than an actual lack of meaning in the terms themselves; 3) anti-bivalent responses, such as in Dummett (1975) and Keefe (2000), deny the validity of the paradox by accepting the premises but denying the conclusion, positing a truth *gap* between truth and falsity, into which the real meaning of the soritical terms can be categorized; for supervaluationists in particular the sorites has semantic consequences rather than merely epistemic ones; and 4) many-valued logic responses, such as in Hallden (1949), Tye (1994), and Field (2003), accept the argumentation in the paradox as sound, stating that rather than a truth value gap between truth and falsity, there is a third truth *value* representing some form of indeterminacy which accounts for the vagueness of soritical terms.

The fact that any semblance of an agreed-upon solution to the paradox of the heap is *far* more elusive than that of the hooded man and the horns is only helpful to my argument, though, if it can be shown that the original and principal aim of the heap was to address an issue in the application of the law of the excluded middle. Otherwise, the considerable venom in the consequences and application of the heap might have been originally used to attack some other theory or support some other principle. After all, despite some superficial similarities between the paradoxes, the main issue in the heap appears to be *vagueness*, whereas the main issue in the hooded man and horns appears to be *ambiguity*. These two concepts might seem similar but there is a crucial difference between them: vagueness arises from a lack of clarity and precision in the understanding, definition, or interpretation of a sentence or term, whereas ambiguity arises when there are multiple meanings for a sentence or term and a lack of clarity as to which meaning is the intended one (Sennet 2011*)*. In the case of ambiguity, the respective meanings may be individually clear for each distinct and possible interpretation of the term or sentence; it is the lack of clarity as to which individual meaning is the intended one which creates the confusion. In the case of vagueness, though, it is clear that there may be only one interpretation which is intended for the term or sentence, but there is a lack of clarity as to what actual meaning this single interpretation involves for that term or sentence. There were one or two sentences in the case of the hooded man and horns paradoxes respectively which seemed to complicate the application of LEM. But in the case of the heap, the individual sentences – whose number could theoretically be infinite – put to the respondent seem unambiguous, e.g. Do 4,194 grains of wheat constitute a heap? If this sentence is true, then 4,194 grains of wheat constitute a heap. If the negation of this sentence is true, then 4,194 grains of wheat do not constitute a heap[13].

One might attempt to solve this problem by arguing that there is a subtle built-in ambiguity in all the individual sorites sentences arising from the different interpretations of the meaning of the term 'heap' or the term 'bald'. For example, if the predicate in question is 'bald', then there are at least two distinct interpretations of the definition: 1) the baldness perceived through an immediate, intuitive estimate *not* based on counting the number of hairs on the relevant head; and 2) the baldness deduced by first counting the number of hairs on the relevant head and then arriving at a conclusion based on whether this number is lower than the maximum number stipulated in the criteria for baldness. On the supposition that 1) and 2) are both legitimate and distinct interpretations of the predicate in question, the soundness – if not the validity – of the sorites paradox depends on exploiting an ambiguity in a manner similar to that of the hooded man and the horns. That is, even though interpretation 1) is usually the intended meaning of the predicate (and the one assumed by the respondent), interpretation 2) is the meaning of the predicate which is necessarily implied by the individual sentences in the sorites (the meaning assumed by the questioner).

It might be objected that there are not two interpretations of the meaning of 'bald' but two criteria for determining whether the predicate applies in a particular case. This might apply to even precisely defined predicates such as 'over 200 pounds in bodyweight'. However, whether

[13] Chrysippus had an interesting piece of advice for how one should respond to the questions in the sorites paradox: "But sorites arguments are fallacious! So crack them if you can, so they don't bother you – they certainly will, if you don't take precautions. But we do take precautions, you [dialecticians] say: Chrysippus thinks that when one is asked to specify gradually whether, e.g., three things are few or many one should come to rest (become quiescent, as they put it) a little bit before one reaches 'many'." Cicero, *Academica* II. 93. There are, though, two main difficulties with this advice: first, silence can speak. Refusing to answer questions is normally assumed to be indicative of ignorance. Second, the identification of when exactly the questions come to the moment 'a little bit before one reaches many' is rather difficult.

or not this attempted solution to the problem – the problem of the apparent incongruity between the structure and implications of the heap and those of the hooded man and horns – is satisfactory, one should note that it seems to completely miss the crucial point. This point is that the hooded man and the horns paradoxes seem to address the potentially problematic application of LEM *not solely* because they can be solved through the articulation and solution to the ambiguity in one or more of their premises after LEM has been applied to them. Rather, it is in virtue of their more substantial elements – elements which they happen to share with the sorites – that they seem to address potentially problematic complexities in the application of LEM to whole sentences. The role played by ambiguity in the Electra and the horns *helps* to support my claim but is certainly not sufficient alone; in fact, it is essentially accidental. Ambiguity necessarily involves at least two possible, distinct meanings of *p*, but it does not necessarily involve *p* and *~p* as the only two possible meanings. Instead, the crucial elements of the sequences in the pattern are as follows: one can either know one's brother or not know one's brother; one either knows the man approaching or does not know the man approaching; one has either lost horns or one has not lost horns; for any specific number of grains of wheat *n*, either *n* constitutes a heap or *n* does not constitute a heap. All three of these Eubulidean paradoxes demonstrate complexities and difficulties in stipulating that *p* and *~p* are *exhaustive* of the domain of possibilities for any sentence *p*. In the case of the heap, the demonstration attacks not only the application of LEM but it also contradicts LEM itself, as it shows that there are an enormous number of sentences which are neither true nor not true (epistemicist solutions aside).

The presuppositions, structure, and implications of the modern form of the liar paradox were explained and defined in the introduction to this paper, but it is especially important in the case of this fourth and final Eubulidean paradox to get an accurate picture of its original formulation. It is a simple but important fact that the original liar had as its constituents not truth and falsehood, or truth and non-truth, but rather between lying and saying the truth[14]:

Questioner: “A man says that he is lying. Is what he says true or a lie?”

Respondent: “What the man says is true”.

Questioner: “If what the man says is true, then what he says is a lie.”

Respondent: “OK, then what the man says is a lie”.

Questioner: “But if what the man says is a lie, then what he says is true.”
[It is likely that the respondent would now become enraged and violence would ensue.]

If the Electra and horns paradoxes illustrated complexities in the application of LEM, and the sorites significantly blurred the line between an illustration of complexities and an attack on the LEM itself, the liar paradox violently crosses this line. It is a simple, direct, and highly effective attack; whereas the sorites applies to all sorts of predicates in natural language, the liar applies to truth and negation by directly involving truth and negation as predicates. If a lie is the direct negation of truth, then the liar shows that the distinction between truth and non-truth is nowhere near as clear as LEM stipulates. One might argue, though, that whereas the sorites attacks the claim of exhaustiveness for the possibilities p and $\sim p$ for any sentence p, the liar seems to specifically attack the claim that p and $\sim p$ are mutually *exclusive*. That is, it appears that while the heap was likely designed to illustrate that in many cases the borderline between truth and negation represents a gap – if not a third truth value – between p and $\sim p$, the liar was likely designed to illustrate that in some cases there is no borderline at all between p and $\sim p$, that the truth values are not exclusive but rather sometimes overlap. This argument would lead to the ostensible conclusion that the liar was originally designed to attack LNC rather than LEM. After all, showing that sentences can be both true and not true contradicts LNC, not LEM.

[14] This is evident in the testimony concerning the paradox in the ancient sources, such as Cicero *Academica* II. 96, and Cicero *Divinatione* II. 11.

One way of responding to such an objection is to point out a pattern: the original, dialectical form of the liar paradox mirrors that of the other Eubulidean paradoxes, involving the elicitation of 'yes' or 'no' answers to deceptively problematic questions. Putting the questions in these paradoxes to virtually any rational adult normally elicits a quizzical expression and a more considered, nuanced answer, (or even a more considered, nuanced *question),* than the possibilities 'yes' or 'no' allow[15]. Allowing the respondent a one-word answer to a question such as 'Have you or have you not lost horns?' usually results in an answer such as 'neither' rather than 'both'. This seems to make it much more likely that the paradox was originally designed to show that there are sentences which cannot be presumed to be only either true or not true, rather than showing that some sentences are *both* true and not true.

This response may meet the further objection, though, that while the similarities in the presuppositions, structure and apparent implications of all the Eubulidean paradoxes may certainly shed some light on what the liar paradox was originally supposed to demonstrate, the fact remains that the liar sentence considered objectively seems to come out both true and not true, rather than neither true nor not true (thus contradicting LNC rather than LEM). It seems to point to a glut between truth and non-truth rather than a gap. One might attempt to refute this further objection by arguing that one should look not parochially only to the problem which the liar presents but rather also to the most plausible and natural solution to this problem. As with any other problem created with a purpose, including the ancient paradoxes, the most effective way of understanding what the liar was originally supposed to demonstrate – even if one insists on considering it in isolation from the other Eubulidean paradoxes – is not by examining solely the puzzle it presents, but rather by also examining the ideal solution to the puzzle. This ideal

[15] I actually used the Eubulidean paradoxes while teaching my tutorials recently, and the answers students gave to questions such as "Have you ever lost horns?" or "Does a person who says he/she is lying tell a lie?" were almost always more complex and admittedly less clear than a simple 'yes' or 'no'.

solution may well have involved viewing sentences such as liar sentences as undefined or indefinite, if not falling between truth and non-truth then at least falling outside these values[16], as will be argued later on in this thesis. For now, when determining the original intention of the liar paradox, against the stubborn objector who insists on 1) ignoring the important pattern described in this paper at the core of all the Eubulidean paradoxes; and 2) maintaining that the only reason the liar is puzzling is because it appears patently obvious that the sentence is both true and not true rather than neither true nor not true, one may at least state that this appearance of a blatant contradiction is a specious one. What underlies the appearance of a contradiction is an inherent indeterminacy between *p* and ~*p*, a gap rather than a glut between truth and non-truth. This objection may be technically irrefutable, though, until philosophers stumble upon a completely agreed upon and satisfactory solution to the liar.

There is, however, a salient question which should be asked: If the heap and the liar, taken together, were originally intended to provide convincing counterexamples to the exhaustiveness of the values *p* and ~*p* stipulated by LEM, and yet all four Eubulidean paradoxes addressed potential or actual problems with LEM, was the purpose of the Electra and the horns a redundant one? Well, they may have served as paradoxical 'warm up exercises' for dialecticians, or they may simply have not been Eubulides' best work. It seems more likely, though, that they were intended to illustrate that the goal of dialectic is the search for truth, whether this means solving ambiguities in sophisms or discovering genuine paradoxes and not shirking the task of examining what might be radically problematic implications. It is perhaps a pity that Aristotle did not consider the Megarian challenge a very serious one, particularly in the case of the sorites

[16] Besides Kripke's (1975) approach to a solution to the liar, Tappenden (1993) and McGee (1993) argue for a unified treatment of the sorites and the heap; Field (2003) and (2008) has argued that the problem presented by the liar bears crucial similarities to that of vagueness paradoxes such as the sorites, and that the ideal solution to these problems involves completely rejecting LEM.

and the liar; one can only speculate how Aristotelian logic may have been altered had he done so. One can be certain, though, that the anonymous comic poet who wrote that, "Eubulides, that most contentious sophist, asking his horned quibbles, and perplexing the natives with his false arrogant speeches, has gone with all the fluency of Demosthenes"[17], has been proved blatantly incorrect.

[17] Diogenes Laertius, *Lives* II. 109 Alternative translation from R.D. Hicks: "Of him it is said by one of the Comic poets: 'Eubulides the Eristic, who propounded his quibbles about horns and confounded the orators with falsely pretentious arguments, is gone with all the braggadocio of a Demosthenes.'" *Lives* II. 108. I think the point is rather clear in both translations: Diogenes Laertius did not like Eubulides very much at all.

Chapter Two
Tarski's hierarchy and a refutation of revenge

2.1 -- Why is the liar paradox a paradox?

The second sentence of this chapter is true (*s1*). The first sentence of this chapter is not true (*s2*). What is problematic about the first two sentences of this chapter, other than their banality? More precisely, what is *logically* problematic about *s1* and s2? A perplexed but patient reader might reason as follows: If *s1* is true, then so is *s2*, because that is exactly what *s1* states. But if *s2* is true, then s1 is not true, because that is exactly what *s2* states. The reader notes that something seems to have gone wrong here; intuitively, there is a problem. But what precisely *has* gone wrong? One might suggest that the answer to this question depends on the foundational principles of logic to which one subscribes. In classical sentential logic, there are two relevant and foundational object language principles. The first of these principles stipulates that for any sentence *p*, either *p* or not *p*. This is known as the law of the excluded middle (LEM), and it should be noted that this principle is constituted by an inclusive disjunction; what it states, then, is the *exhaustiveness* of 1) *p*; 2) not *p*[18]. The counterpart of LEM with respect to truth values is the principle of bivalence, which stipulates that for any sentence *p,* either *p* is true or it is false. LEM and bivalence, then, are concerned with eliminating the possibility of a *gap* between *p* and *not p*, and between truth and falsity respectively. The other relevant object language principle stipulates that for any sentence *p*, not both *p* and not *p*. This is known as the law of non-contradiction (LNC), and states the *exclusiveness* of *p* on the one hand and *not p* on the other. The counterpart of LNC with respect to truth values stipulates that for any sentence *p*, *p* cannot

[18] If LEM were constituted by an *exclusive* disjunction, then it would effectively make LNC redundant, by stipulating not only the exhaustiveness of *p* and *not p* but also their mutual exclusiveness. It should be noted, then, that the possibility of the conjunction of the vales *p* and *not p* for a sentence *p* does not technically contradict LEM, but rather only LNC.

be both true and false. Also, suppose for now that one adopts an Aristotelian conception of truth and falsity: "To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, and of what is not that it is not, is true" (*Metaphysics*, 1011b25-27). Applying this classical conception of truth and falsity to sentences, it follows that a sentence is true if and only if it says of what is that it is or of what is not that it is not. A sentence is false if and only if it says of what is that it is not or of what is not that it is. It turns out that applying this relatively uncontroversial definition of truth and falsity conditions to *s1* and *s2*, along with very uncontroversial rules of inference (reductio ad absurdum in a form acceptable to intuitionists, universal instantiation, modus ponens and modus tollens), results in a contradiction which does not depend on any assumptions. Here's the proof:

| 1. Suppose, for *reductio*, that s1 is true.

| 2. Then s2 is true. (from 1, the fact that s1 says that s2 is true, and the principle that, if a sentence is true, what it says is the case actually is the case)

| 3. So s1 is not true. (from 2, the fact that s2 says that s1 is not true, and the principle that, if a sentence is true, what it says is the case is the case)

| 4. But we have a contradiction. (1 and 3)

5. So s1 is not true. (from 1-4, by RAA)

6. So s2 is not true. (from 5, the fact that s1 says that s2 is true, and the principle that, if what a sentence says is the case is the case, then the sentence is true.)

7. So s1 is not not true. (from 6, the fact that s2 says that s1 is not true, and the principle that, if what a sentence says is the case is the case, then the sentence is true.)

8. But we have a contradiction. (5 and 7)[19]

There is no appeal to LEM here, or to bivalence, or even to double negation (the principle stipulating that the double negation of a sentence *p* necessarily entails *p)*[20]. So the liar paradox is

[19] Thanks to David Hitchcock for helping to tighten up this explanation and proof, so that it becomes clear that the contradiction is derived using only LNC and a standard conception of truth. The double negative in step 7 does not mean that the actual principle of double negation is being used. There is no transformation of the sentence in step 7 from a double negative sentence into an affirmative sentence.

a paradox because it involves “an unacceptable conclusion derived by apparently acceptable reasoning from apparently acceptable premises” (Sainsbury 2009 p3). That is, a paradox is a valid reasoning process which establishes – or seems to establish – a conclusion which contradicts a belief or principle(s) previously taken to be unassailable. It is difficult to find a better example of such a principle than the law of non-contradiction, and this simple proof shows how clearly and quickly the liar seems to present a counterexample to this principle. The liar is, therefore, a very important paradox.

But is there a more immediate problem involved, which makes it merely *appear* as though a contradiction results from the application of a standard interpretation of truth and rules of inference? Well, let’s examine the suspects. Unlike the classical liar sentence, ‘This sentence is not true’, the contradiction obtained does not require any direct sort of self-reference; *s1* does not refer to *s1*, and *s2* does not refer to *s2*. Since direct self-reference is not involved, one can rule out that suspect as a likely culprit. On the assumption that it is semantically legitimate for a sentence in natural language to have as its sole content a statement about another sentence, by a process of elimination the only other plausible candidate for a culprit is the *content* of the statement: the explicit predication of truth (or the negation of truth in the case of *s2*). *s1* and *s2* both make a claim about the truth value of another sentence, and are solely constituted by such a claim. This ‘descending’ version of the liar paradox is the most pertinent formulation of the antinomy for the purpose of this chapter, first because it eliminates the direct self--reference present in the classical liar, and second because it broadly reflects the structure of a solution to the liar like that of Alfred Tarski which relies on a hierarchy of languages and meta-languages. In this chapter, after defining and explaining precisely how and why Tarski’s hierarchy *is* a

[20] The fact that in this proof there is no appeal to bivalence or the law of excluded middle is important, as some significant groups of philosophers and logicians – including intuitionists – object to one or more of these principles. This seems to make the proof just about as uncontroversial as a proof can get.

solution to the liar paradox, I will examine the main criticisms of the Tarskian solution which come courtesy of Kripke. I will then discuss arguably the most devastating objection to *any* solution to the semantic paradoxes (including Kripke's), namely the liar's revenge, and construct a modification to the Tarskian hierarchy which may help to deal with revenge problems in other semantic accounts as well.

2.2 -- Tarski's hierarchical solution

Tarski[21] claims that antinomies such as the liar result from failing to recognize that a single language cannot define and use its own truth predicate. There is a crucial distinction, he argues, between the expressive capability and richness of the *object language* on the one hand, which is the language one wishes to speak *about*, and the *meta-language* on the other hand, which is the language one *uses* to speak about the object language. Natural languages, such as English, are each considered by most people who use them to be universally applicable and all-encompassing, but antinomies such as the liar point to inconsistencies within this view. These inconsistencies result from ignoring the semantic hierarchy of object language and meta-languages. Tarski's crucial distinction is that the meta-language *must* be richer in expressive capability than the object language, allowing the meta-language to form a conception of truth for the object language. One might consider the object language $English_1$ and the corresponding meta-language $English_2$. The expressive capability of $English_1$ does not include the ability to talk about semantic concepts, but rather it includes simple sentences such as 'That graduate student should get some sleep' or 'That graduate student is drinking coffee', while a sentence in $English_2$ includes not only these sorts of sentences but also sentences such as "The statement 'That graduate student should get some sleep' is false", and "The statement 'That graduate student is drinking coffee' is true". In $English_3$, the expressive capability exhausts that of $English_1$ and

[21] Tarski, Alfred (1933), (1935), (1936), (1944), (1969), (1983a), (1983b).

$English_2$, but surpasses the collective capability of $English_1$ and $English_2$ because it also includes sentences *about* $English_2$. And so on for $English_4$, $English_5$, etc. So for any language $English_n$ the corresponding meta-language is $English_{n+1.}$ The chain of reasoning applied to liar sentences such as s1 and s2 is can only entail a contradiction by illegitimately conflating the object language with the meta-language. The appearance of a genuine paradox is, then, a specious one. Since both s1 and s2 both say something *about* the truth value of another sentence, neither s1 nor s2 is actually a sentence of a single legitimate language, since no single legitimate language contains the truth predicate for the sentences of that language in Tarski's hierarchy of languages.

It is useful here to provide a slightly more precise and formalized frame for the problem, the main concepts involved and the solution suggested. Start with a basic object language $\boldsymbol{L}_0$ which does not contain semantic concepts such as truth and falsity. In order to use semantic concepts about the constituents of $\boldsymbol{L}_0$ such as the truth or falsity of sentences in $\boldsymbol{L}_{0,}$ one must move to a meta-language $\boldsymbol{L}_1$ which contains the truth predicate T_0 corresponding to $\boldsymbol{L}_{0.}$ So, given any sentence S which comments on the semantic status of a sentence in $\boldsymbol{L}_{0,}$ the relevant truth schema for S would be $T_0 <S> \leftrightarrow S$. By '$<S>$' I mean some name in the meta-language of the sentence S in the object-language. To state anything meaningful about the semantic concepts such as the truth and falsity of any sentence S in $\boldsymbol{L}_1$, (allowing a sentence such as "The sentence 'that graduate student is asleep' is true" to be meaningful), one simply adjusts the truth schema as follows: $T_1 <S> \leftrightarrow S$. Thus one can generalize the formula to show that for any object language $\boldsymbol{L}_n$ and its corresponding meta-language $\boldsymbol{L}_{n+1}$, the relevant truth schema is $T_n <S> \leftrightarrow S$ where S is any sentence in $\boldsymbol{L}_n$. But the most salient instantiation of the formula in terms of Tarski's solution to the liar antinomy is that in the most basic object language $\boldsymbol{L}_{0,}$ there is no legitimate instance of the truth schema $T_n <S> \leftrightarrow S$. To show this one could construct a sentence S where:

$$S = \neg T_n<S> \text{ [Liar sentence]}$$

By applying LEM, $\Phi \vee \neg\Phi$, to a liar sentence S, it might *seem* a relatively simple task to show that a contradiction can be obtained from the fact that S appears to be true if it is not true, and not true if true: $S = \neg T_n<S>$ if and only if $T_n<S>$, or $\neg T_n<S> \leftrightarrow T_n<S>$. If one could obtain the latter result, then a contradiction would certainly arise. A closer inspection, though, reveals that there is *no* contradiction entailed by S because the semantic content of S can only be understood in the higher meta-language $\boldsymbol{L}_{n+1}$, *not* the object language $\boldsymbol{L}_n$. After recognizing this crucial distinction, the liar sentence S becomes S^*:

$$S^* = T_{n+1} <S> \leftrightarrow \neg T_n <S> \text{ [Liar sentence*]}$$

Thus, in the Tarski hierarchy, S^* does not entail a genuine contradiction because in whatever language S^* is expressed, the evaluation of the semantic content of S^* will always require moving to a higher language, thus not allowing S^* to create a contradiction involving two contradictory conjuncts in a single *common* level of language. Without a meaningful liar sentence, there is *a fortiori* no liar paradox – just a harmless liar sentence.

2.3 -- Problems with Tarski's hierarchical solution

Among the most prominent criticisms of Tarski's solution, and of his entire hierarchy of languages, is that it portrays a highly unrealistic and counterintuitive picture of the way language seems to actually function. This line of attack begins with the observation that there seems to be no support for Tarski's complex hierarchy in the grammatical structure of natural language, nor in the way this grammatical structure is used. Sentences often use predicates such as 'is not true' or 'is true' in a perfectly sensible and simple way, such as the sentence 'Every sentence in this chapter is true'. Besides being perfectly well formed syntactically and grammatically, the semantics of this sentence are not in the least paradoxical. In fact, its truth value is easily understood: assuming that at least one of the sentences in this chapter is not true, the sentence in question is plainly false. Yet according to Tarski's hierarchy, such a sentence must receive the same semantic treatment as the liar sentence, entailing that it is a meaningless sentence, simply because it states something about its own truth. Likewise with the predicate 'not true', as illustrated in the following sentence: "The sentence 'Elephants are smaller than mice' is not true". This sentence uses the same apparently problematic predicate as the liar sentence, namely 'is not true', and yet the semantic implications of this sentence appear no more paradoxical than the sentence 'Elephants are smaller than mice'. Elephants are, of course, *not* smaller than mice, and so the truth value of this sentence is not difficult to determine: it is plainly false. Once again, though, according to Tarski's hierarchy, this apparently innocuous sentence is semantically defective and meaningless. This criticism of Tarski's solution, then, seems a justified one, because the hierarchy does indeed seem to distort the genuine structure and role of semantic predicates used in perfectly nonparadoxical sentences. If the way semantic predicates are used by natural language appears so different from Tarski's radical hypothesis of inconsistency and inexpressibility, then it seems unlikely that languages (and semantic predicates) function in a Tarskian hierarchy either. Thus, while the primary purpose of this hierarchy was to solve the semantic paradoxes, allowing for a much more lucid understanding of – and examination in – semantics, it seems to muddy the waters even more by ruling out seemingly innocuous and meaningful sentences. It obscures even further the possibility of understanding the semantics of any language *in* that specific language, by showing that any language – or a corresponding hierarchy – satisfying Tarski's criteria must be poorer in expressive capability than that of

English in order to obtain a meaningful semantics. The more one is able to demonstrate the prevalence of liar paradox-type circular reference in language, the stronger this criticism seems to become (Kripke 1975). On this view, then, Tarski's truth schema hardly does justice to the universality of natural language.

There seem to be two promising lines of defense one might take against the attack described above: first, one might argue that the criticism relies on the essential assumption that Tarski's hierarchy does in fact apply to natural language, and then attack this assumption; and second, one might argue that the criticism, even if conceded as true, does not take away from the fact that Tarski's hierarchy *does* successfully provide a logically consistent way to get rid of the liar paradox. The first line of defense must surely begin by noting the impressive equanimity with which Tarski conceded the puzzling and perhaps problematic implication of his hierarchy, (that it is impossible to acquire a consistent understanding of the semantics of a natural language such as English without resorting to a hierarchy of languages), and examining exactly *why* and *how* he was able to do so. There seems to be a puzzling dichotomy between two of his most important claims about the connection between natural language and semantics: on the one hand, he claims that

> A characteristic feature of colloquial language (in contrast to various scientific languages) is its universality. It would not be in harmony with the spirit of this language if in some other language a word occurred which could not be translated into it; it could be claimed that 'if we can speak meaningfully about a thing at all, we can also speak about it in colloquial language'. (Tarski 1983 p 164)

So the crucial claim here is that natural language is essentially universal; if something can be properly said, then it can be properly said in a natural language such as English. On the other hand, though, Tarski also notes that in areas of specialized intellectual inquiry such as chemistry, physics or even linguistics, there is often "no need to use universal languages in all possible situations. In particular, such languages are not needed for the purposes of science (and by science here I mean the whole realm of intellectual inquiry)" (Tarski 1969 p. 68). Though it might appear that this second claim contradicts the first, Tarski's point seems to be that while the scope of natural language is necessarily universal, the fact that it *can* be applied to all areas of inquiry does not mean that it *should* be. Colloquial language may often be useful for perfunctory and casual purposes, but in areas of inquiry in which consistency is paramount (i.e. all intellectual inquiry), it is essential to recognize the inconsistency of colloquial language, and to recognize the hierarchy of object languages and meta-languages to maintain consistency. Thus any criticism which relies on the assumption that Tarski's hierarchical solution is applicable to natural language or to the ordinary notion of truth seems to miss the crux of Tarski's project: a crucial distinction between the ideally consistent, *scientific* concept of truth and language on the one hand, and the colloquial, *ordinary* (and inevitably inconsistent) notion of truth and language on the other.

The objector might retort, though, that even if one accepts that the Tarskian approach hinges on such a crucial distinction, and even if one accepts this distinction as legitimate, the ramifications of the Tarskian hierarchy for natural language can at least serve as *clues* to the soundness of his solution. Even if Tarski wanted to distinguish as carefully as possible between the ideally consistent, rigorous language of intellectual inquiry and the inevitably inconsistent colloquial language of ordinary discussion, the fact that his hierarchy rules out non-paradoxical, innocuous sentences involving circular, semantic reference should not be ignored. Just like the liar paradox itself, which originates in natural language (and so any diagnosis of the problem must begin there), the diagnosis of the potential problems in Tarski's solution must begin with natural language as well. Specifically, Kripke (1975) observed that the reason the Tarskian hierarchy eliminates all sorts of harmless sentences is that the Tarskian restrictions rely on the identification of fixed syntactic and semantic criteria for sentences, criteria which *may or may not* indicate a problem – let alone a paradox – involved in the sentence(s). That is, Tarski's hierarchy misses the fact that whether or not a sentence (or a group of sentences) is paradoxical depends *not only* on its own structure and the content of its statement or predication, but *also* on empirical facts which are relevant to what the sentence claims. Take, for example, a pair of rather strange but semantically innocuous sentences:

(1) Most of what Kripke says about CUNY is false.
(2) Everything Priest says about CUNY is true.

Kripke notes that there is nothing intrinsically problematic or paradoxical about (1) and (2) taken together. What determines this is not the independent syntactic structure or meaning of (1) and (2), but rather the verification of relevant empirical facts such as who uttered (1) and (2) respectively. If, for example, Priest uttered (1), and (1) is the only sentence ever uttered by Priest about CUNY, while Kripke uttered (2), then there would indeed be a paradox just around the corner.

It seems, then, that Tarski's hierarchy misses the fact that considering the syntactic and semantic structure of a sentence (or a group of sentences) in isolation may often fail to identify whether or not the sentence in question is a paradoxical one. A careful consideration of relevant empirical facts may often be more important than any parochial criterion focusing solely on the predicate and the sentence itself. Facts such as the identity of the person who utters the sentence in question, the nature of the subject of the sentence, and even the content of other relevant sentences uttered by that person, may be more important than the sentence itself. Kripke's point, then, is that a coherent and consistent theory of truth presented as a solution to semantic paradoxes such as the liar must take into account not only semantic content but also semantic *context*. Semantic context takes into account relevant empirical facts. Take, for example, a much simpler pair of sentences:

(3) Sentence (3) is not true.
(4) Sentence (3) is not true.

The content of (3) is identical to that of (4). There is nothing intrinsically paradoxical – or, for that matter, unequivocally nonparadoxical – about either of these sentences considered in isolation. It is not their *content* but rather their *context* which differentiates them and renders one paradoxical and the other apparently innocuous. With this context in mind, it becomes clear that there is something problematic about (3), and nothing immediately apparent as problematic about (4). (3) states its *own* untruth, and thus is true if untrue and untrue if true, whereas (4) simply states the untruth of (3). Two things, however, should be noted here. First, there is a fundamental difference between Kripke's example contained in the pair of sentences (1) and (2), and my example contained in (3) and (4): the context of the former involves empirical facts, whereas the context of the latter involves linguistic and semantic facts. Second, (4) actually directly states the untruth of (3) and thereby indirectly states the truth of (3) in virtue of the paradoxical content of (3). Nevertheless, the purpose of (3) and (4) is to illustrate that two sentences identical in content can have completely different semantic statuses in virtue of their different contexts.

While Kripke's constructive criticism makes a very convincing case, there are responses one might make in defense of the Tarskian approach. One might argue that, while it may be important to recognize the role of external but directly relevant empirical facts in the semantic evaluation of an individual sentence or a group of sentences, Kripke seems to leave unanswered the question of where exactly is the sharp boundary between relevant and irrelevant empirical facts. Is there a distinction between those empirical facts which are directly relevant to the meaning and semantic status of a sentence, and those which are not? If there is such a distinction, is it knowable? Take, for example, the pair of sentences (1) and (2), both seemingly innocuous – if admittedly bizarre – sentences. In order to determine the veracity – and the semantic status in general – of (1) and (2), there are certain empirical facts which seem directly relevant to the semantic context, such as who uttered (1), who uttered (2), how many of Kripke's other claims about CUNY are true and how many are not true, how many other claims about CUNY Priest has made, the veracity of those other claims, and so on. The trouble is that Priest's other relevant claims might include (5) 'Everything claimed about CUNY is not true'; and (6) 'Most of the things claimed about Kripke are not true'. It seems that, by arguing that "in some sense a statement should be allowed to seek its own level, high enough to say what it intends to say... [rather than] have an intrinsic level fixed in advance, as in the Tarski hierarchy" (Kripke 1975 p.8) Kripke solves a problem in the Tarski theory only to introduce and highlight a new and equally invidious question: Where exactly in the infinite possibilities for relevant facts determining the semantic context of any given sentence does the context become too broad or complex? Kripke's suggested response, roughly, is that, using a three-valued logic[22] it is possible to demonstrate (though only relative to a 'ground model') that there is an infinite *semantic, contextual* complexity (rather than a syntactic, content-based infinite complexity) in the truth predicate for all applicable sentences. The point is that, given some sentence φ which may or may not be paradoxical (depending on its referents), Kripke doesn't need to provide some sort of precise algorithm to show exactly which sentences bear the required semantic relevance to φ; Kripke's point is that φ automatically determines its own semantic status – paradoxical or otherwise – based on its given semantic context rather than its own syntactic content or structure.

[22] Modeled after Kleene's three-valued logic; see Kleene (1938).

2.4 -- The revenge of the liar

In section 2.3 we saw that Kripke argues for a replacement for Tarski's hierarchical solution, and one which leads to liar sentences being assigned a novel value fitting into a *gap* (rather than a *glut*) between truth and falsity. What I am interested in here in 2.4 is a fundamental problem with this conclusion. This problem is interesting and worth addressing because it seems to recur in virtually all purported solutions to the liar paradox, and rears its ugly head just as effortlessly in a contextual solution which takes liar sentences to be 'gappy' (such as Kripke's solution), as in a hierarchy which takes liar sentences to illegitimately conflate an object-language predicate with a meta-language predicate (such as Tarski's solution).

The problem alluded to above is the revenge of the liar. The last few decades have shown that the revenge problem appears to be a far more pervasive one than the original liar; in fact, as others have observed, it seems to become more pervasive after each new solution is presented and inevitably refuted (Scharp 2007). A revenge-sentence ψ states not just its own falsity or untruth but also it also affirms as a second disjunct (or a third or fourth or fifth and so on) the criteria or criterion of the truth schema Γ in the relevant semantic system which is/are used to consistently deal with liar sentences. That is, ψ states the negation of not just its truth in general but also affirms as alternatives those properties ($p_1, p_2, p_3....p_n$) of Γ which are used in the relevant semantic system to assign a novel truth value to liar sentences. So a generalized model for ψ can be defined as follows:

$$\boldsymbol{\psi} = \neg\Gamma\langle\psi\rangle \vee \Gamma_{p1}\langle\psi\rangle \vee \Gamma_{p2}\langle\psi\rangle... \vee \Gamma_{pn}\langle\psi\rangle$$

Here are some examples of substitution instances for $\boldsymbol{\psi}$: in a system which subscribes to bivalence, the substitution instance of ψ for the classical liar sentence can be constructed as follows: 'This sentence RS_1 is false', or, $F\langle RS_1\rangle$. In a system which rejects bivalence, the strengthened liar sentence is 'This sentence RS_2 is not true': $\neg T\langle RS_2\rangle$. In Kripke's three-valued system (Kripke 1975) the extended liar becomes 'This sentence RS_3 is either not true or gappy': $\neg T<RS_3> \vee G<RS_3>$. In a system which claims that liar sentences have an indeterminate truth value *I*, the revenge liar RS_4 states its own untruth or its own indeterminacy, $\neg T<RS_4> \vee I<RS_4>$. In response to those who reject liar sentences as somehow 'meaningless' *M*, the revenge liar sentence RS_5 will read: $\neg T<RS_5> \vee M<RS_5>$. And finally, the revenge-sentence tailored for Tarski's hierarchy reads 'This sentence RS_6 is not true at its corresponding level T_{levelx} of language', $\neg T_{level(<RS6>)}<RS_6>$. In general, then, a liar's revenge sentence applied to any given semantic account states: *"I am whatever truth is not"*.

Here is a more focused and detailed explanation of how the revenge problem applies to Tarski's hierarchical solution: first, recall the version of the liar sentence S^* where $S^* = \neg T_{n+1}<S>$ and the pseudo-contradiction entailed by S^* $T_{n+1}<S> \leftrightarrow \neg T_n <S>$ is indeed blocked by the Tarski hierarchy[23]. Second, suppose that in the hierarchy, the language corresponding to a specific sentence is the *level* for that sentence. So for any sentence *x* let *level(x)* be the corresponding language for *x*. The relevant truth predicate for the Tarski hierarchy, then, would be $T_{\text{level}(x)}x$. The revenge liar sentence *RS* for the Tarski hierarchy is now constructed relatively easily:

$$RS = \neg T_{\text{level}(<RS>)} <RS> \quad \text{[Revenge liar sentence]}$$

The crucial difference between S^* and *RS* is that S^* was a generic liar sentence which simply stated its own falsity, without specifying that it is stating its own falsity *in* its corresponding level of language. *RS*, on the other hand, is a liar sentence tailored for the Tarski hierarchy; it is by no means a new paradox, but rather simply a new *version* of a very old paradox. S^* effectively

[23] Both Priest (2006a) and also Priest (2007) give a very nice summary of how the generalized revenge problem crops up and how it can be tailored to problematize *any* semantic account (especially one designed to solve semantic paradoxes) such as Tarski's hierarchy or Kripke's three-valued system. I base my analysis on Priest's explanation of the revenge problem as applied to Tarski's hierarchical solution.

stated, in the context of the Tarskian solution, its own falsity in a higher meta-language than the object language in which it occurred. This time, though, if *RS* is not true at its corresponding level of language, then it is true at that *same* level of language, since *RS* specifically states that it is untrue at the same level of language at which it is true. So in general, *RS* will produce the following contradiction: $\neg T_{\text{level}(<\text{RS}>)} <RS> \leftrightarrow T_{\text{level}(<\text{RS}>)} <RS>$. More specifically, if *RS* is a sentence of some language $\boldsymbol{L}_{\text{n}}$, (the language being used right now) and so $level(<RS>) = n$, at the next level up, at the corresponding meta-language $n + 1$, it follows that:

$$T_{\text{n}}<RS> \leftrightarrow \neg T_{\text{level}(<\text{RS}>)} <RS> \leftrightarrow \neg T_{\text{n}} <RS> \text{ [Contradiction entailed by } RS\text{].}$$

This is a contradiction, which if accepted shows a central inconsistency in the Tarskian solution. Here is the difference, then, between the generic liar sentence *S** [$\neg T_{\text{n+1}}<S>$] and the tailored-for-Tarski revenge sentence *RS* [$\neg T_{\text{level}(<\text{RS}>)} <RS>$]: *S** reads ‘this sentence is false’, but leaves the level of language unspecified. Tarski’s solution takes advantage of the unspecified level of language by constructing an explicit hierarchy of object languages and meta languages in which the truth value of *S** cannot be understood in the same language as *S** itself. *RS* uses Tarski’s cunning against him, by incorporating this hierarchical distinction into the original liar sentence. The revenge problem, then, might be likened to a giant snowball rolling down a mountain; trying to solve the revenge problem by assigning liar sentences a novel truth value is like using more snow to obstruct the snowball. Each solution just adds more snow, and therefore more size and speed, to the problem.

2.5 -- Possible responses to the revenge/‘snowball’ problem

It seems that there are three possible options now available in response to the revenge or ‘snowball’ problem applied to a semantic account such as Tarski’s hierarchy, only one of which is at all desirable: (1) Inconsistency – accept that *RS* entails an undeniable contradiction in the hierarchy; (2) Self-contradiction – deny that $RS = \neg T_{\text{level}(<RS>)} <RS>$ is a problem because in the Tarskian hierarchy there simply is no such thing as truth fixed at a single given language; (3) Incompleteness – somehow deny that the predicate $T_{\text{level}(x)}x$ can be meaningfully expressed in the Tarskian hierarchy. Unless we somehow accept that there are true contradictions, an admission which few logicians – let alone Tarski – would be willing to accept, option (1) must fail. Strike one. Option (2) fails because it is inherently inexpressible; there is not an available meta-language in which to consistently express the claim in (2), if (2) is true. Strike two. Option (3) seems plausible at first, except that it seems to defeat the central impetus for the creation – or perhaps more accurately, the identification – of the Tarskian hierarchy in the first place: providing a complete and consistent solution to the semantic paradoxes. Ignoring the revenge problem won't make it go away. Strike (3) – We're 'out', and seem to be out of options.

I will attempt to show now, though, that there is a fourth possibility which is perhaps just as promising and much less burdensome compared with possibilities such as creating a completely new semantic account from scratch or accepting the possibility of true contradictions. Since I have argued that the real kryptonite for Tarski's hierarchical solution – and indeed for any purported solution to the liar paradox – is the liar's revenge, this is exactly the problem which I will attempt to take apart here, thereby modifying the hierarchy and repairing the damage done to the Tarskian solution. To illustrate as simply and clearly as possible what precisely is going on in the liar's revenge, let's take a sentence *p* which states its own untruth:

p: *p* is not true **or** p: $\sim p$

Now let's compare *p* with another sentence *p2*:

p2: *p2* is true and *p2* is not true **or** p2: $p2 \wedge \sim p2$

Assuming that both *p* and *p2* are part of the same language $\boldsymbol{L_0}$, what precisely is the difference between *p* and *p2*, which makes *p* a paradox and *p2* simply a contradiction and thus false? The difference is that in the case of *p*, the contradiction is *entailed*, *implicit*, and thus *external* to the sentence, whereas in the case of *p2*, the contradiction is *explicitly* stated and thus *internal* to the sentence itself. As far as I can tell, this single, subtle but apparently enormously important difference is the sole reason logicians have usually dismissed sentences like *p2* as simply false (as an unproblematic contradiction) but have been inexorably drawn to puzzle about the paradox in *p*. The only reason, then, that it is possible to construct a revenge sentence in any apparently consistent system is through the acceptance of the assumption that a revenge sentence (which states, of itself, the negation of the relevant truth schema in the given semantic account) is fundamentally distinct from a sentence such as *p2*. I think that this assumption deserves far more sceptical attention than it has received in the past, and that there is a plausible argument to be made for the thesis that there really is no semantic difference between a sentence such as *p* and a sentence such as *p2*.

It might seem that, in order to support my claim, I'd need to resort to a rather extreme universal generalization remarkably similar to Prior's[24] suggestion: something to the effect that all sentences implicitly include a statement of their own truth value (Prior 1955; 1961; 1976). In this way, *p* not only *implicitly* but also *explicitly* states its own untruth *and* its own truth: *p* is true and *p* is not true. This would be a simple contradiction, rather than a complex paradox. Such an extreme syntactic and semantic view of sentences, though, carries with it some well documented problems when applied to perfectly innocuous sentences; it solves the semantic paradoxes, but

[24] Prior acknowledged that this idea of his builds on the work of medieval thinkers such as Buridan. Simmons (1993) gives a nice overview of late medieval thinking on the liar paradox; it turns out that virtually all the types of solutions put forward and developed by modern philosophers were explored between the 12th and 15th century (though of course not using the tools of modern logic and often not in as great a detail as some modern solutions).

muddles much of the rest of ordinary sentences in natural language[25]. But my proposed modification to the Tarskian hierarchy does not require such extreme measures. Rather, it builds on Kripke's criticism of the hierarchy and a more modest version of Prior's suggestion. I propose the following modification: in a sentence α which is solely constituted by a statement about the semantic status of a sentence β, there is an implicit statement in α of the truth of β. Thus this criterion applies not only to immediate relatives of the classical self-referential liar sentence, but also to instances of the semantic paradoxes which do not involve any direct self-reference, such as the 'descending liar' contained in the first two sentences of this chapter (*s1* and *s2*). The crucial difference, though, between the treatment of a classical self-referential liar sentence and a liar sentence not involving direct self-reference, is that in sentences such as *s1* and *s2*, what is substituted for the implicit statement of truth is *not* a statement of *self-referential* truth but rather a statement of *the truth of the sentence to which the sentence refers* (the truth of its referent). So not only do I want to leave innocuous, ordinary sentences alone, but I want to scrutinize and re-examine the crucial distinction between the semantic structure of *directly* self-referential liar sentences and the semantic structure of *indirectly* self-referential ones. This is all rather confusing without a concrete example, so let's start with the latter and more problematic category (those sorts of semantic paradoxes which do *not* involve direct self-reference):

S1: *S2 is true.*
S2: S1 is not true.

Right now, *S1* and *S2* taken together derive a contradiction through the simple but effective proof given on the second page of this chapter. Applying my modification to this pair, however, where each sentence implicitly states the truth of the sentence to which it refers, they would actually read:

[25] It is interesting to note, though, that Buridan and Prior's solution involving an implicit statement of truth in all sentences was quite convincingly revived and developed by Mills (1998).

$$S1^*: S2 \wedge S2$$
$$S2^*: S1 \wedge \neg S1$$

Now, while the pair may still appear problematic at first glance, they do not actually produce a paradox. This is due to the felicitous fact that *S2* is an explicit contradiction, and is thus false, making *S1* false as well (since *S1* states that *S2* is true). Applying this modified semantic and syntactic criterion to the Tarskian hierarchy of languages, the revenge of the liar seems to lose its venom, as it goes from *RS*:

$$RS = \neg T_{\text{level}(<RS>)} \text{ producing the contradiction } \neg T_{\text{level}(<RS>)} <RS> \leftrightarrow T_{\text{level}(<RS>)} <RS>$$

$$\text{To: } RS^* = \neg T_{\text{level}(<RS>)} \wedge T_{\text{level}(<RS>)}$$

*RS** is an explicit contradiction and is therefore false. Thus the liar's revenge problem is not actually a problem for Tarski's hierarchy of languages, on the assumption that my modification to the semantic account is a consistent one.

2.6 -- Objections to my solution, and my responses

There are some potential objections which should be raised and to which I should provide a cogent response if I wish to maintain my modification[26]. One of the most important objections might argue that my suggestion is a rather ad hoc one, solely motivated by the need to solve paradoxes such as the revenge of the liar. It might well be argued, after all, that my suggestion is both artificial and arbitrary, since it seems difficult to find evidence in the way language actually functions to support the claim that there is an implicit statement of truth in those sentences which are solely constituted by a claim about the semantic status of another sentence (or of itself). What particularly comes to mind here is Kripke's crucial observation that there are many sentences in natural language which are *contingently* paradoxical rather than *necessarily* paradoxical. These sentences may or may not be difficult to assign a truth value to; whether or not they are

[26] Thanks to Matti Eklund for raising this objection after reading an earlier draft of this paper.

problematic is dependent not solely on their fixed syntactic structure but also on their relative semantic context. Suppose, for example, that on the chalkboard of University Hall room 101 there is a sentence, written clumsily with a quickly-crumbling piece of chalk, which can nonetheless be deciphered as follows:

(P1) The sentence written on the chalkboard in University Hall room 103 is false[27].

There does not appear to be anything intrinsically problematic about this sentence on its own, but one cannot be sure just yet. It is *only* by finding out the truth value of the sentence written on the chalkboard in UH 103 (call this *P2*) that it becomes possible to try to assign a truth value to *P1*. Suppose that the sentence in UH 103 says something which is quite obviously and innocuously false and does not contain a reference to *P1:*

(P2) Elephants are smaller than mice.

P1 can in this case be assigned a truth value in a straightforward way: it is true (and certainly not paradoxical). Yet, according to the modification to the semantic account I propose, *P1* would have to implicitly state the truth of *P2* and thus *P1* would in fact be an explicit contradiction, and therefore patently false. My modification, the objector notes, seems to have turned a perfectly harmless sentence with a consistent truth value into an explicit contradiction. It is only the rare instance in which *P2* is actually constituted by a reference to the semantic status of *P1* that my modification seems of any use:

(P2) The sentence written on the chalkboard in University Hall 101 is true.*

Now *P1* and *P2** taken together produce a paradox in a manner exactly analogous to the example involving *s1* and *s2*. It is here, then, that my modification can come to the rescue, rendering *P1*

[27] I should note that, since the discussion here is concerning the liar's revenge rather than any particular or classical liar sentence, one could (and should) replace the predicate 'false' with the operator or schema representing the negation of the truth operator or schema in the relevant semantic account. So for example 'The sentence written on the chalkboard in UH 103 is false and ungrounded' would be the appropriate form of P1 in a semantic account which claims that liar sentences are not grounded.

an explicit contradiction and therefore false, and therefore rendering *P2** false as well (since it states the truth of *P1*). The objection, however, seems to have made its point: my modification does block the liar's revenge, but it also makes self-contradictory a great deal of perfectly normal innocuous sentences as well.

In responding to this objection, the first point I should make is that my modification leaves most sentences in natural language untouched, since most sentences are not constituted by a claim about the semantic status of another sentence. Certainly in my own experience writing on chalkboards and reading from them, most sentences (both inside rooms in University Hall and elsewhere) do not make a claim about their own truth value or the truth value of another sentence. This first point, though, is a subsidiary and largely speculative one, compared with the second: upon closer inspection, *P1* in *both* cases (when coupled with *P2* and when coupled with *P2**) had a *contingent* truth value and semantic status *before* it was combined with *P2* or *P2**. While the objector was correct in pointing out that *P1* did not produce a paradox when combined with *P2*, and so my modification turned a seemingly harmless sentence with a clear truth value into an explicit contradiction, it turns out that, even in that case, *P1* taken on its own was without any fixed or definite truth value or semantic status. When examined as an individual sentence in isolation, sentences such as *P1* have a content which is *incompletely expressed*. It is only through examining *P2* or *P2** *in conjunction* with *P1* that *P1* had a completely expressed content – and in the latter instance, taken together these two sentences managed to produce a paradox. The point is that *P1* as an individual sentence is not much more meaningful than an actual contradiction.

The objector may retort, though, that if sentences such as *P1* examined individually are no more meaningful than explicit contradictions, my modification is still missing the concrete justification for specifically claiming that *P1* contains an implicit statement of the truth of the sentence to which it refers (the truth of *P2* or *P2**). This stipulation, the objector will say, seems like an arbitrary and unjustified way of artificially turning all such sentences into explicit contradictions, and one which ignores the important difference between an actual contradiction and a sentence such as *P1*. This objection, however, misses the point: the difference between an explicit contradiction and *P1* is precisely what allows the recognition of the implicit statement of truth in *P1*. A sentence containing an explicit contradiction, (ex: *P3*: 'I have a desk and I do not have a desk'), appears different from *P1* because *P1* is meaningless and without a truth value until it is coupled with the sentence to which it refers, but *P3* is a contradiction right off the bat. It is precisely because of this crucial difference that *P1* must contain an implicit statement of the truth of *P2* or *P2**. *P1* must say something *before* it can negate what it has said; negation is parasitic, and relies on a sentence which says *something* in order to function, in order to negate (say the opposite of) whatever that sentence says.

Another objection: It might be argued that my modification requires different criteria and motivation when applied to directly self-referential liar sentences than when applied to those liar sentences which do not contain self-reference. Since a significant aspect of the motivation I described involves the fact that a sentence such as *P1* cannot be assigned a truth value when considered in isolation from the sentence to which it refers (as it is impossible to evaluate whether it says of what is that it is) the motivation for my position seems largely undermined when one considers a directly self-referential liar sentence. Take, for example, the self-referential liar's revenge sentence tailored for Tarski's hierarchy: $RS = \neg T_{\mathrm{level}(<\mathrm{RS}>).}$ In this case, there is a problem of an entirely different nature: *RS* can certainly be assigned a truth value when considered in isolation, unlike *P1*. *RS* <u>*does*</u> necessarily have an immediately apparent reference, unlike *P1*. The problem is that it appears impossible to assign a *consistent* truth value to *RS*. It is, therefore, *necessarily* paradoxical, whereas *P1* is *contingently* paradoxical. Why, then, should there also be an implicit statement of truth in *RS*, if the problem appears entirely different and the motivation entirely distinct? The most promising response to this question likely involves attacking the semantic distinction between a self-referential sentence which states its own negation on the one hand, and a non-self-referential sentence which denies a separate sentence[28]. To attack this distinction, one must ask and answer a deeper question: why exactly does *RS* necessarily deny a particular sentence, thus making it separate from sentences such as *P1*? More precisely: what exactly makes *RS* grounded unlike *P1*, and what exactly is *RS* denying? The answer to this question seems to be that *RS* is really no different from *P1*; *RS must* implicitly state the truth of the sentence to which it refers – which in this case just so happens to be itself – in order to deny that same sentence. Without this implicit statement of truth, the negation in *RS* is just as meaningless as the negation in *P1*.

A further objection has to do with the category of complex and conjunctive sentences, particularly those which contain both a liar's revenge sentence and a perfectly normal sentence containing no peculiar semantic reference. Take for example the following sentence:

P3: This sentence is neither true nor ungrounded, and bachelors are unmarried.

On my account, *P3* does not include an implicit statement of the sentence to which it refers because it is not solely constituted by a claim about the semantic status or value of a sentence. This might seem to cause the paradox to emerge again, but it doesn't. Here's why: *P3* is a single sentence composed of two separate statements. Since this is the case, *P3* refers to both bachelors <u>and</u> *P3*. If *P3* refers to itself in the first conjunct, then surely it refers to itself in its entirety; even

[28] What is particularly interesting here is that, while a sentence such as P1 is contingently paradoxical while RS is necessarily paradoxical, it turns out that P1 is also contingently self-referential. If P1 refers to a sentence such as P2 then P1 is not self-referential, but if it refers to P2* then it is indirectly self-referential.

though the self-referential statement is contained in the first conjunct, the actual referent of that first conjunct must be the *entire* sentence (both conjuncts in *P3*). So *P3* becomes:

P3: The sentence 'This sentence is neither true nor ungrounded, and bachelors are unmarried' is neither true nor ungrounded.*

There is no paradox here either, because *P3** implicitly includes a statement of the truth of the sentence to which it refers, becoming:

P3 [with implicit statement]: The sentence 'This sentence is neither true nor ungrounded, and bachelors are unmarried' is both true <u>and</u> not true.*

*P3** is an explicit contradiction and therefore false. Neither (1) nor (2) produce a paradox.

The only other major objection to my solution which I can imagine being put forward is arguably the most important one, and can be succinctly put as follows: my suggestion is false. If I choose to reject this objection, then I seem to have implicitly agreed that the sentence 'my suggestion is false' has a clear meaning which I have understood – even if I understood the meaning before I rejected it. Quite apart from a logical or scientific perspective, this objection draws instead on a very familiar but deceptively convoluted philosophical concept: our intuitions, particularly those intuitions concerning the universality of natural language. The objector will argue to the effect that sentences such as 'that suggestion is false' or 'what you've said is false' are used all the time, in both everyday discussion and rigorous philosophical discussion, without any serious confusion or problems. Moreover, it would be absurd to suggest that whenever someone says 'that's false' they are somehow actually saying 'that's true and that's false', or when someone says 'your suggestion isn't true' they're actually saying 'your suggestion is true and not true'.

This objection is a very natural one, but I will show now that it is not convincing. There are a great many phrases and terms used in natural language which get the point across without actually being used properly; in fact, it is often possible to use a term or phrase in a completely nonsensical or utterly vague way and still convey more or less the intended meaning. Take, for instance, the commonly used phrase ‘well, it is and it isn’t’. This phrase, taken on its own, seems to clearly involve a contradiction – there are actually very few examples of clearer contradictions present in phrases commonly used in the English language. But it usually gets the intended point across, not to mention the fact that it is almost always followed immediately by a qualification about the different ways in which the given thing ‘is’ compared to the other ways the given thing ‘isn’t. The phrase is not supposed to mean ‘for any specific thing x which is currently being discussed, x both exhibits the property P and does not exhibit P in exactly the same way: $Px \land \neg Px$. Virtually everyone who hears the phrase used realizes this fact. Without any context or qualification, though, the phrase does in fact constitute a straight forward and explicit contradiction. The same thing seems to apply to sentences which involve a statement about the semantic status of another sentence, such as ‘that’s false’. It’s a very commonly used phrase and it normally gets the point across through a reasonable and qualified context, but taken on its own, ‘that’s false’ is absolutely meaningless. It needs a referent. Once coupled with a referent such as ‘Some apples are purple’, ‘that’s false’ now acquires a fairly straight forward context and meaning. But if the referent of ‘that’s false’ is the sentence ‘the statement which refers to this one is true’, there’s suddenly a paradox. This first way of responding to the objection, then, appeals to a crucial distinction which was brought up earlier in the chapter, between the ideally consistent, *scientific* concept of truth and language on the one hand, and the colloquial, *ordinary* (and inevitably inconsistent) notion of truth and language on the other.

2.7 – Concluding remarks

One of the most perplexing aspects of the study of paradoxes such as the liar is the cognitive dissonance which results from a fascination with paradox coupled with a fear of contradiction[29]. In fact, there might be more than just dissonance here – there might be a meta-paradox: deriving a clear contradiction in a proof is normally followed by a quick 'QED', yet the contradiction derived by the implications of the liar paradox contradicts the very idea that deriving a contradiction allows a QED – quick or otherwise – at the end of a proof. This is because the liar, as shown in the proof on the second page of this chapter, seems to contradict the law of non-contradiction. This paradox has remained unsolved for well over two millennia, despite the proposed solutions from some of the most brilliant minds in history. In fact, the revenge problem shows that these problematic implications are more pervasive and unsolved now than ever. The reason there might be a meta-paradox, then, is that the paradox which attracts the most fascination of all may be precisely the paradox which teaches us that contradictions are not to be feared. I have argued in this chapter that there is a way to consistently reject the liar, to demonstrate that it is not a puzzling paradox but rather an explicitly false, flat contradiction. Realistically, the chance that the solution for which I've argued will ultimately prove an exception to over two thousand years of contention is infinitesimally small. It might also be unlikely that the semantic paradoxes will be solved in the *next* two thousand years. Therefore, the question of whether and when it becomes logical to carefully examine, rather than dismiss or fear, the real nature and implications of certain apparently inescapable contradictions, seems to become more and more pressing as central problems in the philosophy of logic go longer and

[29] There have often been exceptions to this generalization: those philosophers who embraced contradictions to some extent rather than feared or dismissed them. These exceptions range from the ancient Greeks in the form of Protagoras and possibly Heraclitus through to the medieval in the form of Nicholas of Cusa through to modern philosophy in the form of Hegel through to the contemporary proponents of dialetheism such as Graham Priest.

longer unsolved. For now, though, it is essential to determine whether such contradictions are indeed inescapable. With this condition in mind, I have argued that the only inescapable contradiction is contained in the *premises* of the liar's revenge problem, *not* in the conclusion, thereby falsifying its status as a real paradox.

Chapter Three
Is the liar a paradox or just a contradiction?

It is more or less unanimously assumed in the literature now – and has been ever since the inception of the liar well over two thousand years ago – that the liar is a paradox, rather than just a contradiction. This is why we call it the liar *paradox* rather than 'liar *contradiction*'. The aim of this chapter is to show that this assumption is one which should not be treated as involving unquestionable dogma, but rather as involving an important and deceptively complex distinction which *should* be carefully examined and questioned. Naturally, the principal focus of this chapter is the question posed in its title: Is the liar a paradox or just a contradiction? The first step in answering this question is to understand as clearly as possible exactly what the question is asking. To do this, we'll need precise definitions for the three crucial concepts involved: the liar paradox (abbreviated as 'the liar' in the question), 'contradiction', and 'paradox'. The liar paradox is the central subject of this entire thesis and has been defined and commented on quite extensively, but let's review: Suppose we adopt a fairly uncontroversial *Truth Schema*: $\varphi \leftrightarrow T<\varphi>$. The angle-brackets $<\ >$ are a name-forming device for the sentence φ and T is the truth predicate. Now consider a sentence[30] L, a substitution instance of φ, which states its own untruth: $\neg T<L>$. If L is true, $T<L>$, then L is not true: $T<L> \rightarrow \neg T<L>$. Similarly, if L is not true then L is true: $\neg T<L> \rightarrow T<L>$. So we have established that L is true if and only if it is not true, $T<L> \leftrightarrow \neg T<L>$ and since truth and untruth are the only two options, L is both true and not true: $T<L> \wedge \neg T<L>$. So why is the fact that L is both true and not true a problematic conclusion? What makes $T<L> \leftrightarrow \neg T<L>$ and $T<L> \wedge \neg T<L>$ a problem? Well, this leads us to the second

[30] Substitute for the term 'sentence' any other suitable truth bearer you prefer, such as sentence, statement, assertion, and so on. The point is to remain entirely neutral on this issue (which is no doubt an extremely important and complex one). My focus is on the difference between the concept of a contradiction and the concept of a paradox, and to really satisfy this focus in a single thesis chapter rather than an entire master's thesis requires me to take a neutral stance on *some* foundational concepts.

important concept mentioned in the question posed in the title of this chapter: ‘contradiction’. It is clear that the possibility of a sentence which is both true and not true is one which runs directly contrary to our raw and basic intuitions – intuitions which seem to lie at the intersection of logic, language, epistemology and perhaps even metaphysics.[31] But is it because $T<L> \land \neg T<L>$ is a contradiction that it runs so contrary to these basic intuitions? To find out, let’s examine an eclectic assortment of various definitions of the concept of a contradiction which the great minds in the history and development of Western logic have so courteously provided:

1) “Thus it is plain that every affirmation has an opposite denial and similarly every denial an opposite affirmation…We will call such a pair of sentences a pair of contradictories”. (Aristotle, *On Interpretation* c. 350 BC 17a30)
2) “Contradictions, or sentences one of which must be true and the other false…” (Augustus DeMorgan 1846: 4)
3) “Contradictory negation, or contradiction is the relation between statements that are exact opposites, in the sense that they can be neither true together nor false together – for example, ‘Some grass is brown’ and ‘No grass is brown’”. (A.N. Prior 1967: 458)
4) “Contradictories: Two sentences are contradictories if and only if it is logically impossible for both to be true and logically impossible for both to be false”. (R.M. Sainsbury 1991: 369)
5) “A sentence is contradictory if and only if it’s impossible for it to be true”. (Daniel Bonevac 1987: 25)
6) “Contradiction: Wff* of the form ‘A&~A; statement of the form ‘A and not A’. (Susan Haack 1978: 244)
7) “A contradiction both makes a claim and denies that very claim”. (Howard Kahane 1995: 308)
8) “Contradiction: the joint assertion of a sentence and its denial”. (Baruch Brody, 1967: 61)
9) “To deny a statement is to affirm another statement, known as the negation or contradictory of the first”. (W.V.O. Quine 1959: 9)
10) “A contradictory situation is one where both B and ~B (it is not the case that B) hold for some B”. (R. Routley and V. Routley 1985: 204)
11) “To say of two statements that they are contradictories is to say that they are inconsistent with each other and that no statement is inconsistent with both of them. To say of two statements that they are contraries is to say that they are inconsistent with each other, while leaving open the possibility that there is some statement inconsistent with both. (This may be taken as a definition of ‘contradictory’ and ‘contrary’ in terms of ‘inconsistent’)”. (P.F. Strawson, 1952: 19)

[31] The intersection mentioned here provides yet another fascinating and foundational topic which I’ll have to leave without due exploration and examination, but see Resnik (2003) and Kroon (2003) for excellent examples of such discussion.

12) “We would not say that a man could, in the same breath, assert and deny the same thing without contradiction”. (Ibid. 21)
13) “Contradictory statements, then, have the character of being both logically exclusive and logically exhaustive”. (Ibid. 21)
14) “…two formulae are explicitly contradictory if and only if one is of the form q and the other of the form ‘~q’, that is, if one is the negation of the other”. (Graeme Forbes 1994: 102)

It is evident from this list that, while there is certainly some uniformity and agreement concerning the *fundamental* aspects of a contradiction, there are some differences of interpretation concerning several subtle but salient elements. For example, certain definitions use the notions of truth and falsity (and thus implicitly appeal to the principle of bivalence): DeMorgan (2), Prior (3), Sainsbury (4), and Bonevac (5). A second group of definitions appeals not to truth and falsity but instead to assertion and denial: Aristotle (1), Kahane (7), Brody (8), Quine (9), and Strawson (12). Other definitions appeal not to truth and falsity or assertion and denial but instead to the even broader notion of logical form: Haack (6), Routley (10), and Forbes (14). Strawson in (11) and (13) provides a distinction between contraries and contradictories, though this distinction is either implicit or possible in definitions (1) through (10). Definition (11) summarizes a distinction between contradictories and contraries accepted in classical logic ever since Aristotle: contradictories are exclusive (only one can be true but not both) and exhaustive (one must be true and it is not possible for both to be false), whereas contraries are only exclusive. The appeal to the values truth and falsity, while certainly reasonable, seems to narrow the scope of the definition, at least in the context of this chapter. (The respective contexts of each definition may, of course, have necessitated the appeal to bivalence). Definition (14) is the one in standard usage in contemporary philosophical logic, but it seems to be supported rather than undermined by the other two main ways of defining the concept of a contradiction evident in the others: first, a contradiction consists of two given sentences (or any other truth bearer) which cannot both be true and cannot both be false; and second, a contradiction involves an assertion and its direct and opposite denial, such that both the assertion and the corresponding denial cannot both be accepted. The fact that a standard contemporary definition such as (14) seems to build on other definitions (such as Aristotle's) in the list is unlikely to be a coincidence, if we keep in mind the general chronological point: there has been not only a history of – but also a progressive development in – Western logic.

While finding tangible consistencies in the definition of the concept of a contradiction is surprisingly difficult, there is one aspect of interpretation which lends itself far more easily to a confident generalization: all contradictions are false. Here are a few formulations of this principle[32]:

1) "…the most certain of all beliefs is that contradictory statements are not at the same time true." (Aristotle, *Metaphysics* c. 350 BC 1011b13-14)
2) "Evidently then such a principle is the most certain of all; which principle this is, let us proceed to say. It is, that the same attribute cannot at the same time belong and not belong to the same subject and in the same respect; we must presuppose, to guard against dialectical objections, any further qualifications which might be added". (Aristotle, *Metaphysics* 1005b18-22)
3) "…the law of non-contradiction, $\neg (a \wedge \neg a)$" (Priest 1987: 96)
4) "The law of contradiction asserts that a statement and its direct denial cannot be true together ('not both p and not-p') or, as applied to terms, that nothing can both be and not be the same thing at the same time ('Nothing is at once A and not-A')". (A.N. Prior 1967:461)
5) "…the law of noncontradiction: nothing is both true and false". (Priest 1998: 416)
6) "Thus there seems to be a role in dialogue for an expression whose significance is captured by the law of non-contradiction: by the principle that a sentence and its negation cannot both be accepted". (Huw Price 1990: 224)

The uniformity in the description of the law of non-contradiction itself must be distinguished from uniformity of opinion as to whether the law of non-contradiction is in fact true. It is reasonable to make the generalization, though, that most philosophers (including logicians) throughout history have assumed the truth of the law.

[32] This list was originally compiled in: Grim (2004)

The definition of a paradox seems a little less contentious and convoluted, mainly because it is not a topic as central and ubiquitous in logic (and philosophy in general). This means that there are simply fewer definitions and thus fewer variations of particular interpretive elements, allowing for relatively uncontroversial definitions such as the following: “A paradox can be defined as an unacceptable conclusion derived by apparently acceptable reasoning from apparently acceptable premises” (Sainsbury 2009). Sainsbury’s definition seems to neatly capture the essence of the concept, and it includes the important distinction between a paradox and a puzzling claim: a real paradox includes premises, connected by a reasoning process, which taken together seem to establish a conclusion. That is, a paradox is an argument, not merely a claim. A claim may constitute a contradiction in its entirety, but it cannot constitute a paradox in its entirety. The contradictory claim may function as the conclusion of a paradox, or in some cases there may be more than one contradictory claim (one functioning as a premise, one functioning as the conclusion), but a paradox is *more* than just a contradiction. Otherwise, every conceivable contradictory sentence, such as ‘I am 24-years-old and I am not 24-years-old’, would not be false but rather genuine paradoxes. It is also crucial at this point to ask: what exactly makes the apparently acceptable reasoning appear acceptable, and especially exactly what makes the apparently unacceptable conclusion appear unacceptable? The answer seems to be this: The acceptable reasoning in the premises of a paradox appears acceptable because it appears uncontroversial, and as though it does not contradict any widely held assumptions, opinions or beliefs. The apparently unacceptable conclusion of a paradox appears unacceptable because it either appears to be constituted by a contradiction or else appears to be constituted by a claim which contradicts a widely held assumption, opinion or belief. Moreover, the conclusion is especially paradoxical because it seems to follow from the apparently uncontroversial premises in the argument constituting the paradox. Of course, it is precisely because the apparently unacceptable conclusion apparently *must* be true *if* the apparently acceptable supporting premises are true that paradoxes arise. The contradiction in a genuine paradox is implicit rather than explicit. That is, the contradiction is external to the premises from which the reasoning process begins, entailed by the combination of the premises and the reasoning process, rather than by the reasoning process alone. In a paradox in which the conclusion is not an explicit contradiction but rather a rejected sentence, the contradiction is derived from the combination of all three elements – premises, reasoning process, and conclusion. One salient objection which might be raised now is that paradoxes would then appear to be simply invalid arguments; they appear to be arguments with all true premises and a false conclusion. This is not so. Rather, paradoxes should be considered reductio ad absurdum proofs. The argument takes as its premises some sentences widely supposed to be true, and attempts to use those premises to arrive at a contradictory conclusion. Paradoxes often seem to show that the problem or error is not in the reasoning process involved in the paradox itself, but rather in the sentences widely supposed to be true and which are taken as premises.

Here is a concrete example of a classic paradox to illustrate the preceding points, which were made in abstraction. In the case of the well known paradox of the stone, the reasoning process begins with the following exclusive disjunction: ‘Either an omnipotent being can create a stone so heavy that he cannot lift it, or else an omnipotent being cannot create such a stone’. It turns out, according to the paradox, that in the case of either disjunct, the being can no longer be legitimately considered all-powerful. If the being can create the stone then she cannot lift it, showing that there is something which she cannot do. If the being cannot create the stone then this also seems to show that there is something which she cannot do. The widely assumed truth taken as a premise in the paradox is that there is a legitimate possibility of the existence of an omnipotent being. The paradox takes this assumption as an implicit premise, or perhaps more accurately, it takes this assumption as a supposition. If a being is all-powerful, then she can do anything. On the assumption that the domain of 'anything' includes the task of creating a stone so heavy that she cannot lift it, then the assumption in question – that there is a legitimate possibility of the existence of an omnipotent being – is a false one. I am chiefly interested here in structure rather than solutions. It is useful to note, though, that many solutions to the paradox of the stone, such as Mavrodes (1963), appeal to the claim that the task of creating a stone so heavy that it cannot be lifted is, by definition, impossible for an all-powerful being. That is, these solutions assert that there is not *only* a contradiction *entailed* by the paradox and thus outside the reasoning process in the premises, but also a less immediately recognizable but nonetheless *explicit* contradiction *built into* the reasoning process in the premises themselves. The paradox *seems* to establish that the existence of an omnipotent being is logically impossible, by asking an all-powerful being to perform a task which at first glance appears innocuous. It could well be argued, however, that the task is itself a contradiction and a logical impossibility within the context of the range of abilities for an omnipotent being; essentially, the paradox asks the all-powerful being to be all-powerful and simultaneously *not* all-powerful. It requests an analytic impossibility. It would be a stretch to argue that the inability to draw a square triangle or to identify a married bachelor (or to create a stone so heavy that it cannot be lifted) renders an all-powerful being not in fact all-powerful.

It is essential to note, though, that the strange task requested of the omnipotent protagonist in the paradox of the stone is of a singular and especially evasive sort. After removing all the excessive distractions, superfluous accoutrements and the noise, the paradox of the stone can be stripped bare, revealing not just a regular built-in contradiction but in fact a *self*-contradiction: the paradox asks an omnipotent being to perform a task which is impossible for an omnipotent being. Or to put it another way, the paradox asks an all-powerful being, a being which should be able to perform all tasks (everything), to do something that renders the being no longer capable of doing everything, no longer all-powerful. A more extreme example of such a task could involve requesting that an omnipotent being remove her own omnipotence in such a way that her omnipotence can never be gotten back again. If she cannot perform this task, then there is a task she cannot perform and she is not omnipotent. If she can perform this task and *does*[33] perform this task, then she is not omnipotent (by definition). The paradox plays with the ambiguities associated with concepts such as 'possibility', 'impossibility', and 'anything'. If this type of solution works, then the paradox of the stone is not a paradox. It is just a contradiction, because it uses a contradiction to establish another contradiction, by including in the domain of all possible tasks those tasks which are self-contradictory. It does not use acceptable reasoning to derive an unacceptable conclusion; rather, it uses unacceptable reasoning to arrive at an unacceptable conclusion, precisely because it relies on an ambiguity in the concept of omnipotence. There is, on the one hand, omnipotence in the sense of being able to perform absolutely any task, including any contradictory or impossible task, and there is omnipotence in the sense of being able to perform any possible and non-contradictory task. Still, if this solution does not work, then there are two remaining possibilities: there is yet another possible solution which does in fact work, or there is in fact no consistent solution. In the case of the latter, the

[33] It is important to note a distinction: the omnipotent being would still be omnipotent given that she possesses merely the *possibility* of performing the task of removing her own omnipotence; before she performs the task, she is still omnipotent. But once she actually performs the task, then she is necessarily no longer omnipotent.

paradox does indeed show that the possibility of the existence of an omnipotent being is not in fact a legitimate one[34].

But even in the case of the latter possibility, would the paradox of the stone actually be a real paradox? Well, suppose for the sake of argument that there is no way to solve the paradox of the stone. Suppose that the solution described above does not work for one reason or another, and that no solution will do a better job in its place. Under this supposition, it appears that there is a reasoning process which proves that the existence of an omnipotent being is logically impossible. An impartial and rational reader might ask a terse but natural question at this point: So what? Political and religious ramifications and considerations aside, there does not appear to be any sort of problem with disproving the logical possibility of an all-powerful being. There is certainly no foundational philosophical motive for the assumption that there *must* be an omnipotent being, or that without the possibility of the existence of such a being remaining intact, any hope for a consistent theory of knowledge or existence or language or logic would automatically collapse. To put the point as precisely as possible: paradoxes such as the paradox of the stone, even if unsolved, might present a surprise or a puzzle, but such puzzles do not present a genuine inconsistency or logical impossibility. An unsolved paradox of the stone is nowhere near as worrisome as an unsolved liar paradox, because the liar appears to be a counterexample to the law of non-contradiction, which has – ever since Aristotle – been the most foundational and indisputable/certain of all the principles and axioms of human reason/logic. Some philosophers might maintain here that a consistent theory of ethics or morality requires the existence of an omnipotent being. Whether or not this is the case, I am interested here in logical

[34] It is fascinating to note that Frankfurt (1964) explains that in the case of an omnipotent being of the variety proposed by Descartes' conception of God, where such a being is capable of performing contradictory tasks and is thus not limited by the bounds of logic, creating a stone so heavy that it cannot be lifted *and* still lifting that stone is a legitimate possibility.

consistency alone. The ability to define and distinguish the concepts ‘right’ and ‘wrong’ is a secondary concern compared with the ability to define and distinguish the concepts ‘true’ and ‘false’, or assertion and denial/negation, or the ability to distinguish a coherent, logically unproblematic sentence from an incoherent, logically inconsistent sentence. Without the ability to categorically state that *all*[35] contradictory claims cannot be simultaneously true, it seems to become impossible to affirm any particular theory while denying another. This fact applies to theories in ethics just as much as any other area of inquiry – chemistry, physics, mathematics, and so on. Grant the paradox of the stone its puzzling conclusion and one is left scratching one’s head. Grant the liar paradox its contradictory conclusion and one is left simultaneously scratching one’s head and not scratching one’s head.

The liar is probably the purest paradox of all, as it demonstrates in the simplest way possible the fundamental but subtle difference between the concept of a paradox and that of a contradiction. This difference is illustrated in the distinction between the liar sentence itself and the sentence it seems to necessarily entail if one attempts to assign it a consistent truth value:

1) L: L is not true. [liar sentence itself][36]
2) L is not true if it is true, and L is true if it is not true.

After applying the law of excluded middle, it follows from 2) that:

3) L is both true and not true.

Statement 3) is simply a contradiction and is thus false. If the liar paradox consisted in the sentence ‘This sentence is both true and not true’, then it is highly unlikely that you would be in the unfortunate position of reading this chapter (since the liar would just be the liar, rather than

[35] The difference between accepting all contradictory claims and accepting some contradictory claims is an absolutely crucial one. This will become apparent in the next chapter, in which I will examine the radical dialetheist solution to the liar paradox, which simply accepts the apparently unacceptable conclusion of the liar.
[36] One could satisfy any of the various definitions of a contradiction outlined earlier in this chapter very easily by altering 1) to say something like: 1)* L: ¬L.

‘the liar *paradox*’). The crucial fact which makes the liar paradox a paradox at all is that it seems to use an acceptable line of reasoning to arrive at an unacceptable conclusion from apparently acceptable premises (which in this case equates to no premises at all). It requires the implicit assumption that 1) is not an *explicit* contradiction but rather an innocuous, semantically and syntactically perfectly well formed sentence. The explicit contradiction contained in 3) appears to follow from (2), and (2) is true on the basis of no assumptions besides a standard conception of truth and standard rules of inference. The real reason, then, that there is any liar paradox at all, is because of the assumption that 1) and 3) are fundamentally different; 1) is *not* a contradiction while 3) *is* a contradiction.

It seems that a surprisingly convincing argument can be made for the hypothesis that 1) is in fact not so innocuous, and that the liar paradox may well be smuggling a disguised but nonetheless clearly implicit contradiction into its apparently acceptable reasoning process. Much like the paradox of the stone, upon closer inspection the liar seems to require the generous approval of what might well be a deceptively dubious assumption: a sentence which is solely constituted by a claim about its own falsity does *not* contain a contradiction. The assumption is that 1) can claim its own falsity (or, more precisely, its own untruth) without implicitly claiming its own truth. If 1) does implicitly claim its own truth along with its own falsity, then it is simply a contradiction. To illustrate the point as clearly as possible, here is a description of the reasoning process in the liar paradox which mirrors that of the paradox of the stone:

Premise 1: If the law of non-contradiction holds, then there is *no* sentence which can be both true and not true. *Implicit: It is implied here that if there is a sentence which proves an exception to this principle, (if the consequent is falsified), then the antecedent is necessarily falsified as well* *through application of the basic inference rule modus tollens. It is also being assumed that a sentence can state anything about any other sentence.]*

[In the paradox of the stone, the analogous premise is: If it is possible for an all-powerful being to exist, then there does not exist a task which that omnipotent being cannot perform. *Implicit: It is implied here that if there is any single task that an omnipotent being cannot perform (if the consequent is falsified through an exception to the universally quantified generalization) then the antecedent is necessarily falsified also through application of modus tollens. It is also being assumed that a task includes a task can involve doing absolutely anything, including doing anything towards anything in existence.]*

Premise 2: P2: The sentence ‘This sentence is not true’ is a legitimate sentence.

[The analogous premise in the stone paradox: The task of creating a stone so heavy that it cannot be lifted by an omnipotent being is a legitimate task.]

Conclusion 1, P3: From the universal generalization in P1 it can be instantiated that the sentence put forward in P2, ‘This sentence is not true’, must be not both true and not true.

[The analogous premise: From the universal generalization put forward in P2, it can be instantiated that an omnipotent being must be able to perform the task in question.]

P4: In the case of the first possibility, that the sentence in question is true, then it seems to follow that the sentence is not true (since it claims its own untruth).

[The analogous premise: In the case of the first possibility, that the task in question can be performed by the supposedly omnipotent being, it seems to follow that the being can no longer be considered omnipotent (since she would, by definition, not be able to lift the stone, falsifying the universal generalization in P1].

P5: In the case of the second possibility, that the sentence in question is not true, then it seems to follow that the sentence is true.

[The analogous premise: In the case of the second possibility, that the task in question cannot be performed by the supposedly omnipotent being, it seems to follow that the being can no longer be considered omnipotent (since the inability to perform any single task is enough to falsify the universal generalization in P1.]

Conclusion 2, P6: Since P3 is the case, and because of the results in either possibility found in P4 and P5, it follows that the sentence ‘This sentence is not true’ is true if not true, *and* not true if true. This result seems to falsify the consequent in P1.

[The analogous premise: Since P3 is the case, and because of the results in either possibility found in P4 and P5, it follows that the task in question, whether it is performed or not performable, renders an omnipotent being non-omnipotent.]

Final Conclusion, C3: Since the universal generalization contained in the consequent of P1 is false, it follows that the antecedent claim is falsified as well. Thus the law of non-contradiction is false.

[Analogous conclusion: Since the universal generalization contained in the consequent of P1 is false, it follows that the antecedent claim is falsified as well. Thus the existence of an omnipotent being is logically impossible.]

This informal layout of the reasoning process in both paradoxes reveals that the lynchpin of both arguments lies in the crucial assumption that the task (represented by a sentence) in question is a legitimate candidate or possibility. Both arguments require excluding the possibility of the given task/sentence involving an explicit contradiction. The explicit contradiction seems to be rather superficially disguised in the task presented in the paradox of the stone; it would be far more obvious if it asked the omnipotent being to simultaneously create a stone and not create a stone, or asked her to simultaneously lift a stone and not lift a stone. It turns out, though, that in the context of omnipotence, the task of creating a stone so heavy that it cannot be lifted does in fact consist in directly contradictory criteria. It really asks the omnipotent being to create a stone so heavy that it cannot be lifted, which will of course produce a stone which cannot possibly be lifted (since the being is omnipotent), and then do the impossible by lifting that stone. The error, therefore, is evidently in P2 (the assumption that creating a stone so heavy that it cannot be lifted by an omnipotent being is a legitimate task), once the implicit but crucial aspect of the task is made explicit.

One might argue that, since the reasoning process involved in the liar paradox seems to mirror that of the paradox of the stone, the error in the former paradox can be found in the same place as in the latter paradox. That is, one might argue that the liar paradox falsely assumes that the sentence ‘This sentence is not true’ is a legitimate sentence which can be classified – and indeed *must* be classified – either as true or as not true. This view provokes an important question: If the liar sentence is not a legitimate sentence which can be classified as true or not true, what makes it illegitimate? What makes it an exception? Well, since the implicit assumptions made in P1 include the assumption that a sentence can state anything about any sentence, it is reasonable to argue that this assumption is the most likely candidate for the specific reason which makes the liar illegitimate. Suppose, then, that this assumption is a false one. Suppose that there are some limitations on a sentence’s ability to legitimately state anything (including semantic properties) about any sentence (including itself). What, then, would these limitations look like? A natural suggestion might run along the lines of Tarski’s hierarchy, which as we saw in Chapter Two, stipulates that a sentence such as the liar is in a completely separate level of language (a meta-language) from that of more ordinary sentence such as ‘Snow is white’ (object-language sentences). But the most serious problem for the Tarskian hierarchical approach is the liar’s revenge, as was also shown in Chapter Two. Perhaps, though, the solution to the liar’s revenge problem put forward in Chapter Two becomes better supported now, in view of the parallels drawn between the reasoning process and assumptions in the paradox of the stone and that of the liar. That is, just as it has been argued that the task in the paradox of the stone involves a disguised contradiction, the sentence at the center of the liar paradox involves a disguised contradiction as well:

‘L: L is not true’ becomes ‘L*: L* is true and L* is not true’

L* is simply a contradiction. Non-self referential liar sentences, such as the following pair, can just as easily be dealt with:

‘L1: L2 is true. L2: L1 is not true’ becomes ‘L1*: L2* is true and L2* is true. L2*: L1* is true and L1* is not true’.

L2* is simply a contradiction, while L1 is simply false because it states that a contradiction is true.

There might be some immediate objections to what has just been suggested here. The first is that the justification for postulating an implicit statement of this sort in liar sentences is not nearly as secure as that of the contradictory task in the stone paradox. In the stone paradox, the task only required closer scrutiny of context and structure to uncover the implicit statement, but here the implicit statement seems at best a stretch and at worst completely unwarranted. While I discussed at length these sorts of objections to this solution in Chapter Two, let me say for now in response to this objection that, in the light of my description and definition of the concept of a paradox, if the liar is considered a reductio ad absurdum proof (rather than merely an invalid argument), then the implicit statement of the truth of the sentences to which liar sentences refer seems the logical, if not aesthetically pleasing, solution. A second and perhaps more serious objection might begin with the following question: What value is ascribed to an explicit contradiction which allows it to be dismissed? If the answer to this question is that contradictions are all false, then there still seems to be a problem. After all, if L is false, then it is true. It appears that the only way to salvage this type of solution to the liar, then, is to state something along the lines of Wittgenstein's view that contradictions are not in fact false[37] at all – nor are they true; rather contradictions have no semantic content at all. To adopt such an unorthodox and extreme view, however, would seem to involve throwing out much of classical logic and much of what remains the primary impetus for actually solving the liar paradox in the first place. While this objection appears convincing, there may well be a convincing way to respond: If *L* states both the truth and the falsity of some sentence, then what it says in that case is false, even if one half of what *L* says happens to be true. To state a conjunction involving one true conjunct and one false conjunct is to state a false conjunction, so that if *L* as an explicit contradiction is false, then it is just that – false. The problem is, of course, that if *L* is false then it is true. We might be tempted to stipulate that *L* as a contradiction is false on a different semantic level, but this type of approach falls prey to the liar's revenge (as we saw with Tarski's hierarchy of languages). The bottom line, then, is that if *L* is a contradiction then it is *still* a paradox.

Hold on, though: on my proposed solution (outlined in the latter half of Chapter Two and examined further here in Chapter Three) *L* has as its actual propositional content '*L* and not *L*', which is false. The falsehood of *L* implies that its propositional content is false, so that we get

[37] This view of Wittgenstein's can be traced originally to his early notebooks (Wittgenstein 1961: 108) and is developed in the *Tractatus.* Goldstein (2004) defends and develops this view of contradictions in the context of some prominent paradoxes.

'not both *L* and not *L*'. Now, does this make *L* true? Is there still a problem? Well, if the propositional content of *L* were simply something like '*L* is not true', then at least one of the conjuncts in the conjunction 'both *L* and not *L*' must be false. But in the case of 'both L and not *L*', if the first conjunct is the case then the first conjunct is not the case (if *L* then not *L*). If the second conjunct is the case then the second conjunct is not the case (if not not *L* then not *L*). On my account, though, *L* does not solely state its own untruth but also its own truth. Thus, the fact that we get 'not both *L* and not *L*' does not mean that *L* is true again; there doesn't actually seem to be a problem. So my solution does the job for which it was intended after all. This talk of the status of contradictions brings up a tantalizing question though: What if the liar paradox is not a reductio ad absurdum proof showing that sentences cannot state just anything about any sentence without limitations, but rather a proof showing that the principle of non-contradiction itself is in fact false? It is this seemingly radical possibility which I will examine in Chapter Four.

Chapter Four
The dialetheist solution

"As for the obstinate, he must be plunged into fire, since fire and non-fire are identical. Let him be beaten, since suffering and not suffering are the same. Let him be deprived of food and drink, since eating and drinking are identical to abstaining."[38] Avicenna, *Metaphysics* I.8, 53.13–15

The charming quotation above is from Avicenna, who was probably not a dialetheist. A dialetheist is a person who subscribes to a dialetheism[39], the position which claims that are dialetheias. A dialetheia is a sentence[40], α, such that *both* α and its negation, not-α, are true. Given the assumption that falsity just is the truth of negation, it follows that a dialetheia might also be defined as a sentence which is both true and false. If contradictions are defined as pairs of sentences in which one sentence is the negation of the other, then dialetheism amounts to the claim that there are true contradictions (Priest 2013). It is, at least *prima facie*, highly counter-intuitive that these 'dialetheias' do in fact exist. But before we dismiss dialetheism based on this intuition alone, let's try to carefully and systematically answer three central questions: First, what are the arguments provided in support of dialetheism? Second, do these arguments really prove the existence of true contradictions? Third, if they do, then what consequences follow from these true contradictions?

Let's begin with a bit of recap: As we have seen in past chapters, the liar paradox is an argument – not just a claim – which contradicts a principle known as the law of non-

[38] It's interesting to note that, ethical considerations aside, Avicenna's argument for beating, burning, and depriving the obstinate is, from a strictly logical point of view, a very poor one. It ignores an important distinction between the perspective of the obstinate on the one hand and the perspective of the people doing the beating, burning and depriving on the other. If fire and not-fire are the same for the obstinate, then it does not necessarily follow that fire and not-fire are the same for everyone else – particularly that particular portion of everyone else who will be persecuting the obstinate. What Avicenna should say, of course, is that the obstinate should burn himself, since burning and not burning are the same. Unfortunately, even this would still be rather poor reasoning, since, if burning and not burning are considered the same from the obstinate person's perspective, then there is no reason for the obstinate to choose *either* course of action.

[39] JC Beall and Graham Priest are good examples of proponents of dialetheism.

[40] The term 'sentence' is used here more or less arbitrarily; replace it your favourite title for truth bearers, whether that title is 'proposition' or 'statement' or whatever you prefer.

contradiction[41]. The law of non-contradiction states that, for any sentence *φ,* it cannot be the case that both *φ* and *not-φ:* $\neg (\varphi \wedge \neg\varphi)$*.* The law of non-contradiction has, ever since, Aristotle, been taken as one of the most – if not the most – fundamental and unassailable of all the principles of not just logic but, crucially, metaphysics as well. That is, the law of non-contradiction has, in the work of most of the major figures in the entire history of Western philosophy, been taken as a foundational assumption applicable *not just* to how language *describes* the world and how we can form an idealized version of that language – and the rules which govern truth-preserving reasoning processes in the language – into a calculus we call *logic*, but *also* applicable to how that world actually *exists* in reality (what philosophers call *metaphysics*). The counterpart of the law of non-contradiction with respect to truth values stipulates that a sentence cannot be *both* true and false. The problem is that there are sentences such as *L*: '*L* is false', which come out true if false and false if true. A natural suggestion is that *L* is merely a counterexample to the principle of bivalence (all sentences are true or false but cannot be neither true nor false) because *L* fits into a gap between truth and falsity. The suggestion, more precisely, is that *L* is neither true nor false rather than *both* true and false, as Martin (1967), van Fraassen (1968), and Kripke (1975) propose. This solution doesn't work. Here's why: Suppose we adopt a more general sentential version of Aristotle's definition of truth: for all sentences *φ*, *φ* is true if and only if *φ* says of what is that it is or of what is not that it is not. We'll use the *T-Schema*: $\varphi \leftrightarrow T<\varphi>$. The angle-brackets $<.>$ are a name-forming device for the sentence *φ* and *T* is the truth predicate.

Now consider a substitution instance of *φ* which is different from that of *L*, one in which a sentence *S* states its own untruth, the direct negation of the truth predicate *T* (rather than a separate predicate such as falsity): $S = \neg T<S>$. This is known as the 'strengthened liar'. Suppose first for *reductio* that S is true, $T<S>$. If *S* is true then, according to the *T-Schema, S* must by

[41] See Chapter Three for a much more comprehensive analysis of different interpretations of LNC.

definition be not true (since that is exactly what it states). So we have derived the conjunction $T<S> \wedge \neg T<S>$. But this is a contradiction. So by reductio, we have proved that S is not true, $\neg T<S>$. If S is not true, though, then it is the case that S is true (as this is exactly what S states). So we have derived $T<S>$ from $\neg T<S>$. We have derived the familiar contradictory conjunction $T<S> \wedge \neg T<S>$, only this time it's even more troubling because have derived this contradiction without relying on any assumptions whatsoever. This means that we cannot reject an assumption used in deriving the contradiction, because we are no longer within the scope of any such assumption.

The preceding proof is much more than a party trick. In fact, ever since the mysterious Megarian logician Eubulides we met in Chapter One, the past two and a half millennia have shown that the liar paradox is one of the most enduring problems in the whole of philosophy and logic, and many of its cousins[42] remain some of the most pervasive problems in set theory and in systems of mathematics. (I will discuss the latter briefly later on in this section). A problem which is now coined the 'revenge' of the liar has mirrored a dramatic increase in the number of innovative purported solutions to paradoxical sentences such as S where $S = \neg T<S>$.[43] The trouble is that most of these proposals, no matter how inventive and original, involve assigning some kind of novel truth status to S. The classic example (discussed in Chapter Two) is that of Tarski's proposal, in which S does not entail a genuine contradiction because of a crucial distinction between the expressive capability and richness of the *object language* on the one hand, which is the language one wishes to speak *about*, and the *meta-language* on the other hand, which is the language one *uses* to speak about the object language. Thus, in the Tarski

[42] The term 'cousin' is used here in a rather loose way, but the bottom line is that however similar or dissimilar the various semantic and set theoretic paradoxes are, they all function collectively as the bulk of the motivation for the dialetheist position – see Priest (2006a).

[43] Two of the most prominent of these can be found in Tarski (1933), (1935), (1936), (1944), (1969), (1983a), (1983b); and Kripke (1975).

hierarchy, without the ability of the object language to use its own truth predicate, S does not lead to a real contradiction because S is not a well formed sentence of a single legitimate language. This is because no single legitimate language contains the truth predicate for the sentences of that single language. The meaning of the semantic concepts used by S (namely the truth predicate T and its negation $\neg T$) will always require moving to a higher meta-language than the language of S, thus not allowing S to create a contradiction involving two contradictory conjuncts in a single ***common*** language. With this distinction in mind, S goes from $\neg T_n<S>$, which seems to produce the contradiction $\neg T_n<S> \leftrightarrow T_n<S>$ through a proof very similar to the proof in the preceding paragraph, to S^* where $S^* = \neg T_{n+1} <S^*>$, producing the pseudo-contradiction $\neg T_{n+1} <S^*> \leftrightarrow T_n<S^*>$. Problem solved – or so it seemed for a few decades.

It eventually became apparent that Tarski's hierarchical solution to the semantic paradoxes runs into the singularly pervasive problem alluded to above: the revenge of the liar. In general a liar's revenge sentence applied to any given semantic account states: *"I am whatever truth is not"*. More precisely, a revenge-sentence ψ does not blithely state its own falsity or untruth full stop, but rather it also states as a second disjunct (or a third or fourth or fifth and so on) the criteria or criterion of the truth schema Γ in the relevant semantic system which is/are used to attempt to consistently dismiss liar sentences. That is, ψ states the negation of not just its truth in general but also affirms as alternatives those properties ($p_{1,} p_{2,} p_{3....}p_{n}$) of Γ which are used in the relevant semantic system to assign a novel truth value to liar sentences. So a generalized model for ψ can be ostensibly defined as follows:

$$\boldsymbol{\psi} = \neg\Gamma<\psi>$$

Or:

$$\boldsymbol{\psi} = \neg\Gamma<\psi> \vee \Gamma_{p1}<\psi>$$

Or:

$$\boldsymbol{\psi} = \neg\Gamma\langle\psi\rangle\rangle \vee \Gamma_{p1}\langle\psi\rangle \vee \Gamma_{p2}\langle\psi\rangle$$

And so on: $\boldsymbol{\psi} = \neg\Gamma\langle\psi\rangle\rangle \vee \Gamma_{p1}\langle\psi\rangle \vee \Gamma_{p2}\langle\psi\rangle \ldots \vee \Gamma_{pn}\langle\psi\rangle$

There might in some cases only be one property p_1 of Γ used to assign a value to liar sentences, or even just the negation of Γ full stop; it depends on the context of the given semantic system and how liar sentences are semantically classified in that system. For examples of substitution instances of $\boldsymbol{\psi}$, see Chapter Two, section 2.4.

It is perhaps not an enormous coincidence that a more modern version of one of the other really important Eubulidean paradox discussed in Chapter One, the paradox of the heap, is another of the main arguments used as motivation for the dialetheist position. The more modern version of this ancient paradox takes various forms. In the following form, it is valid by mathematical induction (Sorensen 2001 p1):

> Base step: a collection of ten billion grains of wheat is a heap.
>
> Induction Step*:* if a collection of n grains of wheat is a heap, then so is a collection of n – 1 grains.
>
> Conclusion: a collection of one grain of wheat is a heap.

If there are some groundbreaking objections to the use of mathematical induction as an inference rule, then using the inference rule *modus ponens* 9,999,999,999 times is just as efficacious (though a deplorable waste of time, as I found out one long night). At some point in the process, of course, through the subtraction of a very large number of individual grains of wheat, one at a time, a collection of grains of wheat which *does* constitute a heap becomes a collection of grains of wheat which does *not* constitute a heap. It is immediately clear that ten billion grains of wheat makes a heap. It seems equally clear that one less than this amount also makes a heap. But the induction step implies not just conditionals such as ‘If 999,999,999 grains of wheat is a heap, then so is 999,999,998 grains of wheat’, but also conditionals such as ‘If 999,999,999 grains of wheat is a heap, then so is 2 grains of wheat' (Boolos 1991). Correspondingly, the induction step implies the conditional 'If 2 grains of wheat is not a heap, then neither is 999,999,999 grains of wheat'. Paradoxically, then, it seems that the predicate *always* applies, or that it *never* applies. The problem of vagueness, a topic which has received a great deal of attention in the literature recently, involves the application of this same reasoning process to many different predicates and concepts in natural language. This reasoning process does not solely apply to heaps of wheat – it applies to much less obscure things such as redness, roundness, tallness, baldness, blondeness, and a practically endless and impressively eclectic assortment of various ordinary concepts and predicates. In fact, Peter Unger famously used this wide range of applicability of sorites reasoning as his main premise to argue "for the denial of the existence of ordinary things, and for all that that entails" to be interpreted just about as literally, radically and provocatively as can be imagined. (Unger 1979 p 29)

But Priest does not use the paradox of the heap to deny the existence of ordinary things (unless disjunctive syllogism, which he rejects, is considered an 'ordinary thing'). Rather, he notes that most approaches to solving the vagueness problem, with the obvious exception of the epistemicist approach[44], involve some form of under-determinacy and/or rejection of bivalence assigned to the all the borderline cases which fall between the given predicate and the negation of that given predicate. The borderline cases represent a gap between the predicate and its negation. Priest argues, though, that it is just as reasonable – and perhaps even more reasonable – to assign borderline cases to a *glut* (both the predicate *and* its negation) rather than a *gap* (Priest 2013). This glutty approach to solving the vagueness problem is adopted by Colyvan (2009); Weber (2010a); Priest (2010); and Ripley (2012). One of the main reasons that any approach which relies on assigning some under-determinacy to borderline cases involved in a vague

[44] The four main types of approaches to solutions to the vagueness problem are outlined in Chapter One.

predicate tends to fail is the higher order vagueness phenomenon. The issue is that there must be some definite and very specific boundary between the range of under-determinacy and the range of what is definitely the given predicate. This boundary must be a sharply defined limit, because without such a limit, the under-determinacy spreads throughout the entire range of the predicate and the predicate's negation.

This pattern can be outlined and quantified as follows. Here is first order vagueness: it is impossible to pinpoint a sharp boundary or threshold between the exact number of hairs on a man's head which make him bald and the number of hairs which make him *not* bald. The transition between the predicate of baldness and the predicate's negation (not-baldness) is imprecise and apparently impossible to quantify in any accurate way. All we can tell is that the transition does occur at *some* point, because there certainly are men walking around who we describe as unequivocally bald, and men walking around who we describe as unequivocally *not* bald. So baldness is a vague predicate because it contains borderline cases; there are a number of points (points represented by a specific given number of hairs on a man's head, progressing incrementally) in the transition from the predicate to the predicate's negation at which it is entirely unclear whether the point is included in the range of the predicate or in the range of the predicate's negation. Therefore, there is no sharp boundary between the range of the predicate and the range of the predicate's negation, making the transition not easily quantifiable. Ascribe to this collection of points the value 'indeterminate' (though 'borderline' might work just as well; it makes no difference). Second-order vagueness occurs in response to the following question: where is the sharp boundary between the points constituting range of the predicate ('definitely bald') and those points constituting borderline cases? Correspondingly: where is the sharp boundary between the points constituting the range of borderline cases ('maybe bald;

maybe not bald') and those points constituting the range of the predicate's negation ('definitely not bald')? It turns out that the answer to either of these questions is just as elusive as a satisfactory solution to the first order vagueness problem. After ascribing some appropriate value assignment to the collection of points in between those points constituting the range of the predicate and those which are borderline indeterminate, third-order vagueness follows in the same pattern which can be represented clearly and geometrically as follows:

First order vagueness -- problem frame:

0 $\longrightarrow$

Predicate *P* ('definitely bald')} <u>Elusive</u> transition point/sharp boundary between *P* and not-*P* { Not-*P* ('definitely not')

First order vagueness -- natural and temporary solution:

0 $\longrightarrow$

Def. P} { For all borderline points, <u>*Indeterminate*</u> whether *P* or not-*P*} { *Def*. not-*P*

Second order vagueness -- problem frame:

0 $\longrightarrow$

Def. P} Elusive transition between *P* and *Ind* { *Ind* } Elusive transition between *Ind* and not-*P* { *Def.* Not-*P*

Second order vagueness -- natural and temporary solution:

0 $\longrightarrow$

Def. P} {<u>*Ind-2*</u> whether *P* or *Ind*} { *Ind* } { <u>*Ind-2*</u> whether *Ind or Def* Not-*P*} { *Def.* Not-*P*

Third order vagueness -- problem frame:

0 $\longrightarrow$

Def. P} Elusive transition {*Ind-2* }Elusive transition{ *Ind*} Elusive transition {*Ind-2*}Elusive transition {*Def* Not-*P*

Third order vagueness -- natural and temporary solution:

0 $\longrightarrow$

Def. P} {*Ind-3*} {*Ind-2*} { *Ind-3* } { *Ind* } { *Ind-3* } { *Ind-2* } {*Ind-3*} { *Def* Not-*P*

The fourth order vagueness problem frame will ensue in the same pattern from the natural and temporary solution to the third order vagueness problem. There is no apparent sharp boundary between the range of those points which are *Def-P* and those which constitute a group of 3rd level indeterminacy, between that same group of 3rd level indeterminate points and a group of 2nd level indeterminate points, and so on. In the natural and temporary solution to fourth order vagueness, there will be a 4th level indeterminacy range between the *Def-P* range and the 3rd level indeterminacy range, between that same 3rd level indeterminacy range and a group of 2nd level indeterminacy points, and so on. What a geometrical understanding of the higher order vagueness phenomenon makes particularly clear is that vagueness is a problem peculiar to gaps – what the gap represents is a transition between what a given predicate definitely applies to and what that given predicate's negation definitely applies to. More specifically, it is a problem peculiar to the task of filling gaps –what 'filling gaps' represents is the task of identifying and quantifying, in some precise way, the transition between *Def-P* and *Def not-P,* defining their limits precisely. It also shows very clearly that the strategy of assigning a third value outside of the two values 'Predicate' (ex: Bald) and 'Not-predicate' (ex: Not bald) to the borderline cases involved in the transition between the predicate and its negation is an effectively useless one. It solves nothing. It simply creates more gaps, more unclear transitions between sets of given collections of supposedly clear values or groups of points. This hopeless indeterminacy (which, as we will soon see, effectively amounts to inconsistency[45]) spreads by induction throughout the entire range of the given predicate, even if there seem to be cases to which the predicate

[45] My exposition of higher order vagueness doesn't yet show why and how Priest is able to make the jump in reasoning from gaps to the necessity of gluts; that is, the argument hasn't yet shown why we get continuous inconsistency involved in vague predicates rather than continuous indeterminacy. I will explain precisely how he makes this jump in reasoning when I compare the pattern common to the topological structure of each of these arguments (and, more specifically, the structure which ensues when the typical strategy of assuming indeterminacy is used).

definitely applies and clear cases to which the predicate definitely does not apply. Then, as a result of the extremely wide applicability of the vagueness problem, this indeterminacy generalizes all throughout our descriptions of all sorts of phenomena in the empirical world.

It might be objected that this indeterminacy does not spread by induction throughout the entire range of the given predicate for the following reason: if we are dealing with a discrete measure like the number of hairs on a person's head and we have clear cases to which the predicate definitely applies and clear cases to which it definitely does not apply, then there must be an upper limit to the orders at which there are vague boundaries. This objection does not hold. Here's why: if we are indeed dealing with a discrete measure, then there would be a limit at both ends – both in the range of the predicate and the range of the predicate's negation. But what my exposition of the higher order vagueness problem shows is that assigning 'borderline' cases some under-determined value besides the given predicate and that predicate's negation causes the indeterminacy to spread even to the clear cases which the predicate seemed originally to definitely apply to, and to those clear cases which the predicate originally seemed to definitely *not* apply to. The point is that there are clear cases at either end (predicate and predicate's negation), where cases to definitely apply, but assigning indeterminacy to borderline cases 'in the middle', as it were, results in vague boundaries continuously expanding from the middle and outwards in both directions (both toward the range of the predicate *and* toward the range of the predicate's negation).

A third and even more concrete argument for true contradictions used quite heavily by Priest argues that through a priori, purely descriptive analysis of transition states in the empirical world (the real/observable world) it becomes apparent that there *must* exist certain moments of symmetry involved in a transition from a given state to its negation at which the thing is both in the original state and *not* in that same state. The argument runs as follows (Priest 2006a, pp. 160-71): consider any given state of any given thing in the empirical world. This is an extremely broad domain, and the examples seem infinite, but let's start with a particularly parochial one: consider the strange chapter you're currently reading. At some point you will stop reading this chapter (and hopefully at the end rather than somewhere in the middle). So you will go from reading this chapter to not reading this chapter. That is a transition from a specific state to the negation of that specific state. Priest's argument posits that there must be a point of perfect symmetry between the given state and its negation, at the middle point of transition, at the very moment at which you go from reading this chapter to not reading this chapter. Now at this point of perfect symmetry, sticking with the idea that we are trying to be purely descriptive in our analysis of what is going on in the empirical world, there are four possibilities for what you are doing at that point of perfect symmetry; or more generally, Priest (2006a p161) claims that there are four possibilities for the descriptive analysis of a system s which is in a state s_0, described by a sentence α, before a time t_0 and then in a different state $s_{1,}$ described by not-α, after t_0:

(A) You **are** reading this chapter; more generally, s is solely in s_0 at t_0

(B) You are **not** reading this chapter; s is in s_1 at t_0

(C) You are **neither** reading **nor** not reading this chapter; s is in neither s_0 nor s_1 at t_0

(D) You are **both** reading **and** not reading this chapter; s is in both s_0 and s_1 at t_0

To say that this point of perfect symmetry between the original state of the given system and the negation of that state can be described as either A or B would seem, at the very least, utterly arbitrary. To describe this point as A would be to describe a perfectly symmetrical point as one which is asymmetrical, as one which is in the given state (reading this chapter) rather than the state's negation (not reading). Thus it seems not only arbitrary but also patently false to describe the state as A. The very definition of this point is precisely that there is absolute equipoise between reading this chapter and *not* reading this chapter; it is the point of perfect symmetry which *must* occur at some point if you are to undergo the transition from reading to not reading. Similarly, to describe this point as B would be to ascribe asymmetry to symmetry (a symmetrical transition point), to describe it as being in the state's negation (not reading this chapter) rather than in the given state (reading). So to describe the point as B would seem just as inaccurate as describing it as A.

Yet, according to classical logic, this point of perfect symmetry can only be described as A or as B (since classical logic stipulates both exclusiveness and exhaustiveness between a sentence α and its negation not-α). But if it is inaccurate to describe this symmetrical transition point as type A and also inaccurate to describe it as type B, then it follows that the point is not type B and *also* not type A: neither A nor B. This seems not only a logically appropriate move, involving nothing more controversial than the classical inference rule known as conjunction introduction, but it also appears a very natural and intuitive move. It is natural and intuitive because we seem faced with a dilemma; neither fork in the road is at all appealing, so we wait and search for a more sensible option, rejecting both of the unappealing options. The problem is that on the received understanding of negation (and more specifically, the principle of double negation) in classical logic, a point described as type C simplifies to type D, since not-not reading the chapter is equivalent to reading the chapter. So if you are not reading this chapter but also not-not reading this chapter, then you are both reading and not reading this chapter. Neither one *nor* the other simplifies, in this case, to both one *and* the other[46]. Thus it follows that

[46] It might well be argued that the use of the principle of double negation to simplify not not-α to α is actually not necessary for the contradiction and the Type D description to obtain. The support for this claim is that a change which is both not not-α and not-α is just as much a contradiction as one which is both α and not-α. I think this argument is unsound. Here's why: in examining whether a contradictory description of the transition state obtains,

in this transition there must be a point of perfect symmetry at which you are *both* reading *and* not reading this chapter; s occupies both s_0 (represented by a sentence α) and s_1 (represented by the negation of α, namely not-α) – quite clearly a contradiction, and apparently a true and unavoidable one.

While this is the most straight-forward way of getting the true contradiction Priest is after, there is another way of getting the contradiction – and one which more easily generalizes and conforms to the pattern evident in the other two main arguments examined so far – even on the assumption that s occupies a Type C state at time t_0. Suppose that we manage to assign some interesting and mysterious alternative value to describe a system s when it occupies this point of perfect symmetry between the state s_0 and the state s_1 at time t_0. Let's call this third value 'NFM' for 'no fact of the matter'. We might make this stipulation by arguing that it is not just an arbitrary stipulation but rather a justified one. It is a justified stipulation, we argue, because there is no fact of the matter which makes s in s_0 at t_0, but also no fact of the matter which makes s in s_1 at t_0, just as we might argue that a sentence such as 'The present King of France is bald' fits into a gap between truth and falsity since there is no fact of the matter which makes it either true or false. Now, Priest argues against this stipulation as follows: {the fact that (there is no fact of the matter which makes α true)} is precisely the fact which makes not-α true (Priest 2006a pp 64-6). Thus there is a fact of the matter which makes not-α the case, thus assigning the sentence a

---

we can either use classical logic or not use classical logic. If we are using classical logic, then we should use and apply all its principles and rules of inference consistently, and examine the results when the subject of the application is something like a semantic paradox. If we are not using classical logic, then we need a theoretical justification for using any principle or rule of inference in classical logic. Other than the rejection of a given principle or rule of inference (in this case, the principle of double negation) from the calculus of classical logic, there is nothing else known about this alternative system of logic in use by the person doing the rejecting. The point is not that this person is deprived of any theoretical justification for appealing to any principle or rule of classical logic. The person could, theoretically, have any sort of alternative logic they like which might have considerable overlap with classical logic. But that person needs to justify any instances of this overlap, independent of the justification used by the system of classical logic, showing why each principle and/or rule is in her logic – whether it overlaps with classical logic or not – as opposed to the principle of double negation. Thus it becomes theoretically unclear what status the conjunction not-α and not not-α genuinely takes on.

definite value (false, not-α) rather than no fact of the matter either way. The mere fact that a sentences fails to be true provides sufficient ground for the truth of the negation of that sentence, or so Priest argues. I think there might well be serious problems with this argument against what Priest calls the ‘central rationale’ for the existence of truth-value gaps; for instance, it appears completely arbitrary to stipulate that {the fact that (there is no fact of the matter which makes α true)} is more decisive than {the fact that (there is no fact of the matter which makes not-α true)}. Also, it seems very odd to presuppose that the fact that there is no fact of the matter is just as much a definite and sharply defined fact as the definite and sharply defined fact that one plus one equals two (and the applicable definite fact that the sentence ‘One plus one equals two’ is true), or the fact that snow is white (and the applicable definite fact that the sentence ‘Snow is white’ is true).

Even if there is a serious problem with Priest’s argument, though, there seems a more relevant and less objectionable way of getting the true contradictions Priest is after: in line with the pattern clearly evident in the reoccurrence of the vagueness problem in higher order vagueness, and in the reoccurrence of the liar paradox in the liar’s revenge, the Type C description of the state of system *s* at time t_0, however underdetermined or construed, must *also* have *sharply defined boundaries*. So if *s* is described as Type C or ‘NFM’ at t_0, then are the same four possibilities (Type A through Type D) for the description of *s* in its transition from s_0 to its transition state, and from its transition state to s_1. Claiming in response that before the symmetry point, the person is reading this chapter and after the symmetry point the person is not reading this chapter does not seem to help matters much, since there must also be a transition between the reading state and the symmetry point, and a transition between the symmetry point and the not-reading state. Through induction, then, *s* is now described as always in a Type C or NFM state. There are, of course, many concrete instantiations of this generalization, since virtually *any* system in *any* state at *any* time can be substituted in place for the system s in state s_0 before time t_0 and in state s_1 after t_0. So the problem generalizes, and spreads uninterrupted indeterminacy all over its domain – motion, change, and time. The under-determined gap, just as in higher order vagueness, expands indefinitely, so that you are never actually reading this chapter (Type A description), and nor are you ever *not* reading this chapter (Type B description). But of course, there *is* a definite time at which you are reading this chapter, just as there is a definite time at which you are not reading this chapter. Similarly, there are definitely men who are bald, and there are definitely men who are not bald.

But as I noted in a footnote immediately after my exposition of higher-order vagueness, so far we have continuous indeterminacy established by the typical approach to stipulating an under-determined value – Type C or borderline or whatever – to things like transition points and borderline cases. It seems that Priest does not yet have his true contradictions in two out of the three areas so far discussed. It might seem, therefore, that the reoccurrence of the liar paradox in the liar's revenge does not quite follow this same pattern, given that the strategy of stipulating some alternative semantic value for liar sentences seems to directly result in semantic contradiction; the result appears to be inconsistency rather than continuous indeterminacy. That is, liar sentences can simply take on the new semantic value and thus create a new contradiction; what seems to expand is a *glut involving* truth and falsity, rather than an *ever increasing gap between* truth and falsity. The only real inconsistency, one might argue, is between the fact that in the world there *are* in fact verifiable, real instances of the definite predicate in question and the negation of that predicate (rather than just borderline cases) and the fact that higher order vagueness leaves us with solely descriptions of borderline cases. Similarly, the only real inconsistency is between the fact that in the world there *are* in fact definite states of systems in time and definite negations of states of systems in time, and the fact that assigning Type C descriptions to transition states leaves us with continuous and constant indeterminacy for states of systems in time. This appearance, however, is specious for a number of reasons. To understand these reasons, we should begin by noting that the liar's revenge problem leaves us not just with dialetheism as an option, but also with constant semantic ascension as a second option (but admittedly an equally bizarre one). As Tim Maudlin notes:

> We need a new semantic category, call it Weird, which then allows for the construction of the new problematic sentence 'F-or-O-or-W is either false or Other or Weird', and off we go on another cycle. In each turn of the crank we add a new semantic category, and then add that new category as an alternative in a new problematic sentence. Approaching the situation in this way, we seem to reach a conclusion: for any given set of possible semantic values, if they are mutually exclusive, then they are not jointly exhaustive. Supposing that no sentence can have more than one value implies that at least one sentence has a semantic value not on the list. The list must grow, without end. (Maudlin 2007, from Beall 2007 p 185)

Theoretically, there are an infinite number of alternative semantic values which can be introduced to avoid the dual values truth and falsity; once we run out of names for these values like 'Meaningless', 'Gappy,', 'Other' and 'Weird', we can simply start numbering each new semantic value. As Maudlin notes, the list of semantic values will expand continuously and infinitely. The problem is, of course, that even for the 917,716,491st new semantic value introduced to assign liar sentences some special category outside or beyond or in between truth and falsity, there must also be a negation of that 917,716,491st new semantic value. Moreover, that semantic value is, by definition, *not* the semantic value we call falsity, and certainly not the semantic value we call truth. Thus there will also necessarily be a liar's revenge sentence available for that 917,716,491st new semantic value. We should recall that a liar's revenge sentence ψ states the negation of not just its truth in general but also affirms as alternatives those properties ($p_1, p_2, p_{3 \ldots} p_n$) of Γ which are used in the relevant semantic system to assign a novel truth value to liar sentences: $\psi = \neg\Gamma<\psi>> \vee \Gamma_{p1}<\psi> \vee \Gamma_{p2}<\psi> \ldots \vee \Gamma_{pn}<\psi>$. Thus one of the main topological features of the structures involved in the pattern created by the reoccurrence of the liar paradox does in fact directly match up with one of the two main topological features of the structures involved in the pattern created by the reoccurrence of the vagueness problem *and* the reoccurrence of the sharp transition problem: the inconsistency is *continuous;* the inconsistency spreads throughout the entire applicable domain of its context (in this case, the entire domain of possible semantic values beyond truth and falsity).

The other apparent point of distinction is the more significant one: the pattern created by the reoccurrence of the vagueness problem and the sharp transition problem features *indeterminacy,* and this main feature seems to be contrasted with the *inconsistency* created by the reoccurrence of the liar's revenge. This distinction is, I think, just as specious as the continuity distinction examined and discarded in the previous paragraph. To make this as clear as possible, though, it helps a great deal to lay out the four distinct categories for descriptive analysis of 1) a given liar sentence; 2) borderline case in a vague predicate, and 3) transition state at t_0 of a system *s:*

Four distinct categories for descriptive analysis of a given liar sentence *L (where L is a liar sentence which states its own falsity/untruth; or a sentence* α where α states $\neg\alpha$)

A: Given liar sentence *L* is true; this category is here described by a sentence α.

B: *L* is not true; this category is described by $\neg\alpha$

C: *L* is neither true nor not true; described, therefore, as $\neg\alpha$ and $\neg\neg\alpha$

D: *L* is both true and not true; described as α and $\neg\alpha$

Four distinct categories for descriptive analysis of a system *s* in a transition state at a time t_0 when it is in a state s_0 before t_0 (described by α) and in another state s_1 after t_0 described by $\neg\alpha$)

A: At time t_0, system *s* occupies state s_0 only; described by α

B: *s* occupies state s_1 only; described by $\neg\alpha$

C: s occupies neither s_0 nor s_1; described, therefore, by $\neg\alpha$ and $\neg\neg\alpha$

D: s occupies both s_0 and s_1; described by $\neg\alpha$ and $\neg\alpha$

<u>Four distinct categories for description of a borderline case involved in the transition from the range of a predicate (*Def-P*) and the range of the predicate's negation (*Def not-P*)</u>

A: Borderline case stipulated as *Def-P*; described by α

B: Borderline case stipulated as *Def not-P*; described by $\neg\alpha$

C: Borderline case stipulated as neither *Def-P* nor *Def not-P*; described by $\neg\alpha$ and $\neg\neg\alpha$

D: Borderline case stipulated as both *Def-P* and *Def not-P*; described by α and $\neg\alpha$

It's clear, then, that common throughout all three of these central dialetheist arguments claiming to establish radically counterintuitive conclusions is the general assumption that stipulating some value equivalent to indeterminacy beyond the original two supposedly exclusive two values is no less contradictory than stipulating a conjunction of the two values (despite their apparent exclusiveness). This is because indeterminacy is, by definition, *not* determinacy; its limits *are* sharply defined, even if its content is taken to be, by definition, *not* sharply defined. So a borderline case of baldness is, by definition, *not* in the range of what *is* baldness, and correspondingly (also by definition), not in the range of what is *not* baldness. A borderline case of baldness is, therefore, *both* in the range of baldness *and* in the range of not-baldness. It is, in any system of logic, difficult to dispute the claim that $\neg\alpha$ & $\neg\neg\alpha$ is any less contradictory or inconsistent than α & $\neg\alpha$. But in the context of the calculus of classical logic, it is *impossible* to dispute this claim. Thus all three dialetheist arguments seem to show that the application of classical logic to these three phenomena spread *continuous inconsistency* throughout each of their respective domains – motion, vague predicates, and semantics.

Hold on just a second, though. The discussion so far has been predicated upon an absolutely fundamental and perhaps tenuous assumption. The assumption is this: between a thing and what that thing is not, between a predicate and its negation, between a sentence α and $\neg\alpha$, between any true sentence and its negation, between a state of a system in time and the different state occurring *immediately* after that state, there is *anything*. I think there is something rather odd about this assumption. Why is there something odd? Dummett articulates the point quite neatly:

> A statement...divides all states of affairs into just *two* classes. For a given state of affairs, either the statement is used in such a way that a man who asserted it but envisaged that state of affairs as a possibility would be held to have spoken misleadingly, or the assertion of the statement would not be taken as expressing the speaker's exclusion of that possibility. If a state of affairs of the first kind obtains, the statement is false; if all actual states of affairs are of the second kind, it is true. (Dummett 1959, p 8 of reprint. Italics original.)

To talk about a class of states of affairs which mysteriously exists in between the only two possible classes – that class in which what the statement states *is* the case, making the statement true, or that class in which what the statement states is not the case, making it false – is to try to create and fill a gap where there simply is *not* a gap. What a thing is not, both formally/logically and metaphysically, is anything and, more pertinently, *everything* which it is not. To presuppose that there is something outside of 1) and 2) where 1) is a specific thing x and 2) is everything else besides that specific thing $\neg x$, is to *presuppose* that a thing can be both what it is and what it is not. It is not surprising, then, that odd and contradictory results follow from arguments which take this presupposition for granted. The liar's revenge phenomenon is simply the result of attempts at solving the liar paradox which falsely presuppose that truth and the negation of truth are not exhaustive of the domain of semantic values. But the liar's revenge shows us that the negation of truth includes *all* semantic values other than truth – gappyness, meaninglessness, indeterminacy, context-sensitive assignments, semantic value in a higher meta-language as opposed to a lower object language, and so on. The same holds true in the case of vagueness and higher order vagueness; there is no such thing, in classical logic, as a borderline case which is defined as neither the given predicate nor that given predicate's negation. This is true of the relationship between what is a borderline case of a bald man and what is not a borderline case of a bald man, just as it is true of the relationship between what is the very last point in time at which your eyes are fixed on this page and the very first point at which your eyes are no longer on this page. The bottom line, then, is that there is not a gap between what a thing *is* and what it is *not*, whether this gap is considered to have metaphysical content, logical content, or any other sort of content.

So what are the implications of attacking this common presupposition? Well, let's start with vagueness. It eliminates the true contradictions Priest wants to get through a glutty solution to the vagueness problem. More specifically, acknowledging precisely why Priest doesn't get these true contradictions results in a solution to the vagueness problem in line with what is known as 'epistemicism' – the view that there *is* a sharp and precise boundary between virtually all vague predicates and their negation, but that this sharp and precise boundary is often *unknowable,* or at least very *difficult* to know.[47] It is not the aim of this chapter – or this thesis – to dive into the debate about whether epistemicism is too counter-intuitive to be believed by a rational person (or accepted by a rational philosopher). This issue raises deep questions which cannot be answered here. But I will say this: while it might seem a very strange belief indeed that there is a precise and specific number of hairs on a man's head which is exactly *one more* than the number which would make him bald, this belief is in my view much easier to swallow and incorporate than many of the bizarre consequences which follow from an unsolved vagueness problem, namely continuous inconsistency spread throughout descriptions (both colloquial and philosophical) of much of the empirical world. Thus all so-called borderline cases of vagueness

[47] For examples of proponents and explanations of – and an impressive assortment of arguments for – this view in line with classical logic, see Williamson (1994) and Sorensen (2001).

are actually either a Type A or Type B description; either *Def-P* or *Def not-P*; either in the range of what is the given predicate or in the range of that predicate's negation.

As for the implications of acknowledging the problematic presupposition in the dialetheist argument focusing upon transition states in motion in the empirical world, we seem led to a natural hypothesis: there is, by definition, no such thing as a further point – 'transition point' – between the state in time you occupy at which you are reading this chapter and the state in time you occupy at the first moment at which you are *not* (no longer) reading this chapter. There is no such thing as a *single* point at which you go from reading this chapter to not reading this chapter. Instead, we might hypothesize there are *two* points which constitute the transition from the original state to the next: the very last point which is the last at which you *are* reading this chapter, and the very first point which is the first at which you are *not* reading this chapter. An alternative hypothesis, based on the idea that time is a continuum, is that there is no symmetry point at all: there is no such thing as a least upper bound to the Type A state and no greatest lower bound to the Type B state. On this solution, then, between any two given points there are other points as well. In this case as well, then, Priest fails to get the true contradictions he wants. All states of any given system at any point in time are either Type A or Type B descriptions.

But what about the implications of this dismantled presupposition for the liar paradox? Well, as previously noted, it explains precisely why the liar's revenge problem crops up over and over again: the negation of truth includes *all* semantic values other than truth. So we understand why it is completely useless to stipulate some semantic value somehow beyond truth and falsity/untruth, (call this value 'Other'), and consign liar sentences to 'Other': while liar's revenge sentences seem to say something special and separate from the original liar sentence, namely *I am whatever truth is not*[48] as opposed to *I am false*, the original liar sentence implicitly expresses the same thing. Whatever truth is not, whether that is called meaninglessness or gappyness or ungroundedness or simply just falsity, all these things are, by acceptance of a trivial tautology, *not* truth. Therefore, we can eliminate the possibility of a Type C description for liar sentences. The problem is, of course, that if we stipulate that liar sentences fit into Type A descriptions, then they also fit into Type B descriptions. If we stipulate that they fit into Type B descriptions, then they also fit into Type A descriptions. It follows, therefore, (unless one subscribes to the approach outlined in Chapter Two), that liar sentences fit into Type D descriptions.

It would seem that my analysis of these central dialetheist arguments shows that the motivation for dialetheism is much more parochial than Priest claims, and thus the consequences of the position not nearly as radical as we might assume. After all, according to my analysis, the only legitimate true contradictions Priest has derived are those which arise from very strange sentences which say something *about the tools we use* to evaluate whether *other* sentences which talk *about the world* are sentences which say of what is that it is or of what is not that it is not. The other true contradictions, those which involve sentences which actually talk about things in the world, are illegitimate. The issue, however, is that even if it can be shown that the only sorts of true contradictions which Priest is licensed to assert arise not from purely descriptive analysis of concrete and familiar things like motion, time, change, and vague predicates, but rather from a select few contorted and obscure sentences, there is a principle of classical logic which implies that in that case every sentence is both true and false. Here's why: *ex falso quodlibet*, more accessibly known simply as 'explosion', stipulates that from one logically false sentence, all

[48] It is very straightforward to recast this mold for typical liar's revenge sentences in a way which eliminates the self-referential aspect. Take, for instance the following pair of sentences: 'The next sentence is whatever truth is not. The previous sentence is true.'

other true and false sentences follow (otherwise known as trivialism). That is, explosion states that for any accepted logical falsehood, such as the acceptance of a contradiction $\varphi \land \neg\varphi$ (of which $T<L> \land \neg T<L>$ is a good example) any other sentence ψ must also be accepted:

$$\{\varphi \land \neg\varphi\} \rightarrow \psi$$

The proof is very straightforward: Given a contradiction $\varphi \land \neg\varphi$, simplify the conjunction to isolate the first conjunct φ and the second conjunct $\neg\varphi$. Introduce a disjunction using the isolated first conjunct between that sentence and any other sentence: $\varphi \lor \psi$. Apply disjunctive syllogism to $\varphi \lor \psi$ using the isolated second conjunct $\neg\varphi$ to derive ψ. We've shown that the truth and falsity of *any* sentence (no matter how apparently nonsensical or banal) follows from the acceptance of a single contradiction – QED.[49]

Now there are, I think, three fairly reasonable ways to proceed at this junction: first, there is Priest's strategy of forming a new, paraconsistent system of logic which throws out principles such as disjunctive syllogism, thus making it impossible to prove theorems of classical logic such as explosion (the proof given above). It seems a bit odd to throw out classical logic and construct a whole new system of logic which accommodates the existence of true contradictions solely because of a few pathological, obscure, contorted sentences; the motivation for this move, though technically logically warranted, seems both parochial and abstract. Moreover, there is then a considerable burden of proof on that new paraconsistent system of logic to demarcate which contradictions are true and false and which are false only. Second, there is the possibility of simply running with the consequences of the liar paradox coupled with explosion, accepting that every sentence is both true and false. While this second option may seem patently insane

[49] This proof owes its origin to C.I. Lewis (1932).

and/or ridiculous, it is actually much more difficult to refute trivialism than we might expect.[50] Trivialism is usually dismissed as flagrantly nonsensical or absurd simply because it appears so, but deeper analysis shows that it resists all obvious objections. If nothing else, though, the fact that trivialism is lately being seriously considered by some prominent philosophers is an enormous testament to the power of the liar paradox. Of course, trivialism is not, in logician-speak, technically an alternative to dialetheism, since dialetheism does not seem to intrinsically rule out *any* contradictions. But any rational dialetheist will maintain that many contradictions are false only (rather than both true and false); in fact, a dialetheist who adopts dialetheism at the expense of non-dialetheism seems to be intrinsically implying that at least one contradiction is false only, namely, the contradiction involving dialetheism and non-dialetheism. Third, we might continue to search for a completely satisfactory solution to paradoxes such as the liar, while genuinely assuming that there *is* such a solution. In the meantime, it does not become immediately permissible to start assuming that inconsistency and contradiction spread all throughout our descriptions of the entire empirical and theoretical world. The liar paradox is a powerful problem of philosophy, but it is not so powerful as to persuade us to abandon all attempts at reasoning carefully, consistently and clearly. Rather, we should be quite content with restricting inconsistency to obscure, contorted sentences.

[50] For examples of unsuccessful arguments against trivialism, see Priest (2000) and Kroon (2003).

Bibliography

Avicenna. *La Métaphysique du Shifā Livres I à V*, Georges Anawati, ed., Paris: Vrin, 1978.

Barnes, Jonathan, 1978. *The Presocratic Philosophers (Two Volumes)*. Routledge and Kegan Paul.

—— 1984. *The Complete Works of Aristotle*, Volumes I and II, Princeton: Princeton University Press.

Barwise, Jon and Etchemendy, John, 1987, *The Liar*, Oxford: Oxford University Press.

Beall, JC (ed.), 2008, *Revenge of the Liar*, Oxford: Oxford University Press.

—— 2009, *Spandrels of Truth*, Oxford: Oxford University Press.

Boolos, George, 1991. 'Zooming Down the Slippery Slope'. Nous, 25: 695-706.

Bonevac, Daniel, 1987. *Deduction*, Mayfield Press.

Brody, Baruch, 1967. 'Logical Terms, Glossary of,' in Paul Edwards (ed.), *The Encyclopedia of Philosophy*, Macmillan and Free Press.

Cicero, *On Academic Scepticism* (Hackett, 2006). [Original title: *Academica*]

Clark, Michael, 2002, *Paradoxes from A to Z*. London: Routledge.

Cohen, S. Marc. 1986. "Aristotle on the Principle of Non-Contradiction." *Canadian Journal of Philosophy* 16 no. 3: 359–370.

Colyvan, M., 2009, "Vagueness and Truth", in H. Dyke (ed.), *From Truth to Reality: New Essays in Logic and Metaphysics*, Oxford: Routledge, pp. 29–40.

Crivelli, Paolo. 'Aristotle on the Liar', *Topoi: An International Review of Philosophy* 23. 1 (2004): 61-70.

DeMorgan, Augustus, 1846. 'On the Syllogism: I: On the Structure of the Syllogism', trans. Camb. Phil. Soc. VIII, pp. 379-408, repr. In Peter Heath (ed.), *On the Syllogism and Other Logical Writings*, Routledge and Kegan Paul, 1966.

Diogenes Laertius, *The Lives and Opinions of Eminent Philosophers, II*. Trans. C.D. Yonge 1853. Online: http://classicpersuasion.org/pw/diogenes/index.htm

—— *The Lives and Opinions of Eminent Philosophers, II*. Trans. R.D. Hicks 1925. Harvard University Press, Loeb Classical Library.

Dummett, M. (1959) 'Truth', Proceedings of the Aristotelian Society 59, 141-62. Reprinted in Dummett (1978).

——1975. 'Wang's paradox', *Synthese*30: 301–24; reprinted in his *Truth and Other Enigmas*, Also reprinted in Keefe and Smith (1996).

Dummett, M., 1978. *Truth and Other Enigmas*, Duckworth.

Field, H., 2003. ‘No fact of the matter’, *Australasian Journal of Philosophy*, 81: 457–480.
——2008. *Saving Truth From Paradox,* Oxford: Oxford University Press.

Fitch, F.B., 1964. “Universal Metalanguages for Philosophy”, *Review of Metaphysics* 17: 396-402.

Forbes, Graeme, 1994. *Modern Logic*, Oxford University Press.

Frankfurt, Harry G. 1964. ‘The logic of omnipotence’. *Philosophical Review* 73 (2): 262-263.

Grim, Patrick, 1979. ‘Plantinga’s God and Other Monstrosities’, Religious Studies 15: 91-7.
——1981. ‘Platinga, Hartshorne, and the Ontological Argument’, Sophia 20:12-16.
——1991. *The Incomplete Universe: Totality, Knowledge, and Truth*, MIT Press.
——2003. ‘What is a contradiction?’ from *The Law of Non Contradiction: New Philosophical Essays* (2003), Oxford, edited by JC Beall, Graham Priest, Bradley Armour-Garb

Gupta, Anil, 1982, “Truth and paradox”, *Journal of Philosophical Logic*, 11: 1–60. Reprinted in Martin (1984).

Haack, Susan, 1978. *Philosophy of Logics*, Cambridge University Press.

Halldén, S., 1949. *The Logic of Nonsense*, Uppsala: Uppsala Universitets Arsskrift..

Kahane, Howard , 1995. *Logic and Contemporary Rhetoric,* 7th edn., Wadsworth.

Keefe, R., 2000. *Theories of Vagueness*, Cambridge: Cambridge University

Kneale and Kneale, 1962, *The Development of Logic*. London: Oxford University Press.

Kleene, S.C. 1938, “On notation for ordinal numbers”, *Journal Symbolic Logic*, 3: 150–155.

Kripke, Saul, 1975, “Outline of a theory of truth”, *Journal of Philosophy,* 72: 690–716. Reprinted in Martin (1984).

Lewis, C.I. 1932. “Alternative Systems of Logic”, *The Monist*, 42: 481–507. Reprinted in Lewis (1970), pp. 400–419.

Long, A.A. and Sedley, D.N. 1987. *The Hellenistic Philosophers Volume 1*. Cambridge University Press.

Lukasiewicz, Jan., 1951. *Aristotle's Syllogistic*. Oxford: Oxford University Press.

Maudlin, T., 2007. ‘Reducing Revenge to Discomfort’, *Revenge of the Liar*, pp. 184-196.

McGee, V., 1991. *Truth, Vagueness and Paradox*. Indianapolis: Hackett.

Mignucci, Mario, 1999 K. Algra et al, *The Cambridge History of Hellenistic Philosophy*: Cambridge: Cambridge University Press.

Mills, Eugene. 1998. ‘A simple solution to the Liar’, *Philosophical Studies* 89, 197-212.

Moline, Jon, 1969. ‘Aristotle, Eubulides, and the Sorites’. *Mind: A Quarterly Review of Philosophy 78* (Jul 1969): 393-407.

Price, Huw, 1990. “Why ‘Not’?”, Mind, 44: 221-38

Priest, G., 1987, *In Contradiction*, Dordrecht: Martinus Nijhoff.

—— 1995, *Beyond the Limits of Thought*, Cambridge: Cambridge University Press, 2nd expanded edition, Oxford: Oxford University Press, 2002.

—— 1998. ‘What’s So Bad about Contradictions?’ Journal of Philosophy, 45: 410-26.

——2002, “The Hooded Man”, *Journal of Philosophical Logic* 31 (5): 445-467.

—— 2003, “To Be and Not to Be: That Is the Answer. On Aristotle on the Law of Non-Contradiction”, *Philosophiegeschichte und Logische Analyse*, 1: 91–130.

—— 2006a, *In Contradiction*, Oxford: Oxford University Press.

—— 2006b, *Doubt Truth to Be a Liar*, Oxford: Oxford University Press.

—— 2007, ‘Revenge, Field, and ZF’, *Revenge of the Liar*, pp. 225-233.

—— 2010, “Inclosures, Vagueness, and Self-Reference”, *Notre Dame Journal of Formal Logic*, 51: 69–84.

——2013. ‘Dialetheism’, *Stanford Encyclopedia of Philosophy*.

Priest, G., JC Beall, and B. Armour-Garb (eds.), 2004, *The Law of Non-Contradiction. New Philosophical Essays*, Oxford: Oxford University Press. -Priest (2003).

Prior, A. N., 1955. “Curry's Paradox and 3-Valued Logic”, *Australasian Journal of Philosophy* 33:177-82.

—— 1961. “On A Family of Paradoxes”. *The Notre Dame Journal of Formal Logic*, II, pp. 16–32.

——1967. *Past, Present and Future*. Oxford: Clarendon Press

——1976. *Papers in Logic and Ethics*. London: Duckworth. (Edited by Geach, P.T., Kenny, A.J.P.)

Quine, W.V.O. 1959. *Methods of Logic*, Holt, Rinehart, & Winston.

——1981. ‘What price bivalence?’ *Journal of Philosophy*, 77: 90–5.

Reiger, Adam. The Liar, the Strengthened Liar, and Bivalence. *Erkenntnis* 54 (2):195-203 (2001)

Rescher, N., 2009. *Unknowability: an enquiry into the limits of knowledge* , New York: Lexington Books.

Routley and Routley, 1985. ‘Negation and Contradiction’, *Revista Colombiana de Matematicas*, 19: 201-31.

Ripley, D., 2012a, “Sorting Out the Sorites”, in K. Tanaka, F. Berto, E. Mares and F. Paoli (eds.), *Paraconsistency: Logic and Applications*, Dordrecht: Springer, 2012, pp. 327–45.

––– 2012b, “Paradoxes and Failures of Cut”, *Australasian Journal of Philosophy*, 91: 139–64.

Russell, B., 1923. ‘Vagueness’, *The Australian Journal of Philosophy and Psychology*, 1: 84–92. Reprinted in Keefe and Smith (1996).

Sainsbury, R. M., 1995, *Paradoxes*, Cambridge: Cambridge University Press.

Scharp, Kevin, 2007. ‘Aletheic Vengeance’, *Revenge of the Liar*, pp. 272-319.

Simmons, Keith. 1987. ‘On a Medieval Solution to the Liar Paradox’, *History and Philosophy of Logic* 8, 121-140.

––– 1993. *Universality and the Liar*, Cambridge: Cambridge University Press.

Strawson, P.F. 1952. ‘Logical Appraisal’, ch.1 of Introduction of Logical Theory (London: Methuen), repr. in R.I.G. Hughes (ed.), *A Philosophical Companion to First-Order Logic*, Hackett, 1993.

Sennet, Adam, 2011. ‘Ambiguity’. *Stanford Encyclopedia of Philosophy.* Online: http://plato.stanford.edu/entries/ambiguity/

Sorensen, Roy, 1988. *Blindspots*, Oxford: Clarendon Press.

––– 2001. *Vagueness and Contradiction*, Oxford: Oxford University Press.

Tarski, Alfred. 1933. *Pojęcie prawdy w językach nauk dedukcyjnych*, Warsaw: Nakładem Towarzystwa Naukowego Warszawskiego.

––– 1935. “Der Warheitsbegriff in den Formalisierten Sprachen”. *Studia Logica*: 1: 261-405. English translation by J.H. Woodger, “The Concept of Truth in Formalized Languages”, in Tarski [1956, pp. 152-278].

––– 1936. “O gruntowania naukiowej semantyki”. *Przeglad Filozoficzny* 39: 50-57. Englsih translation by J.H. Woodger, “The Establishment of Scientific Semantics”, in Tarski [1956, pp. 401-8].

––– 1944. “The Semantic Conception of Truth and the Foundations of Semantics”. *Philosophy and Phenomenological Research* 4: 341-76.

––– 1956. *Logic, Semantics, Mathematics*. Edited and translated by J. H. Woodger. Oxford: Oxford University Press.

––– 1969. “Truth and Proof”. *Scientific American* 220, no. 6: 63-77.

––– 1983a. *Logic, Semantics, Metamathematics*, second edition, ed. by J. Corcoran. Indianapolis: Hackett.

––– 1983b. “The Concept of Truth in Formalized Languages”, translation of Tarski 1935 by J.H.

Woodger in Tarski 1983a, pp. 152–278.

Tappenden, J., 1993. ‘The liar and sorites paradoxes: towards a unified treatment’, *Journal of Philosophy*, 90: 551–77.

Tye, M., 1994. ‘Sorites paradoxes and the semantics of vagueness’, in J. Tomberlin (ed.), *Philosophical Perspectives: Logic and Language*, Atascadero, California: Ridgeview. Reprinted in Keefe and Smith (1996).

Unger, Peter. (1979). ‘There Are No Ordinary Things’, Synthese, 41, pp. 117-54

Weber, Z., 2010a, “A Paraconsistent Model of Vagueness”, *Mind*, 119: 1026–45.

Wheeler, Samuel C., 1983, ‘Megarian Paradoxes as Eleatic Arguments’, *American Philosophical Quarterly* 20/3: 287-95.

Williamson, Timothy (1994). *Vagueness*, London: Routledge.
––– (1999). ‘On the Structure of Higher-Order Vagueness’, Mind, 108, pp. 193-200.
––– 2000. *Knowledge and its Limits*, Oxford: Oxford University Press.

Wittgenstein, L. 1961. *Notebooks 1914-16*. Oxford: Blackwell.

Yablo, Stephen, 1985, Yablo, “Truth and reflection”,*Journal of Philosophical Logic*, 14(3): 297–349.
––– 1993, “Paradox without self-reference”, *Analysis*, 53: 251–252.
––– 2006, “Circularity and Paradox”, in Bolander, Hendricks, Pedersen (eds.), *Self-reference*, Stanford: CSLI Publications.