

ALETHEIC VENGEANCE<sup>1</sup>

Kevin Scharp

The Ohio State University

Before you set out for revenge, first dig two graves.  
– attributed to Confucius

0. INTRODUCTION

Thinking about truth can be more dangerous than it looks. Of course, our concept of truth is the source of one of the most frustrating and impenetrable paradoxes humans have ever contemplated, the liar paradox, but that is just the beginning of its treachery. In an effort to understand why one of the most beloved and revered members of our conceptual repertoire could cause us so much trouble, philosophers have for centuries proposed “solutions” to the liar paradox. However, it seems that our concept of truth takes offense to our efforts to understand it because it appears to retaliate against those who propose “solutions” to the liar. It takes its revenge on us by creating new paradoxes from our own attempts to find resolution. That is, most proposed solutions to the liar paradox give rise to new, more insidious paradoxes—often called *revenge paradoxes*. For our attempts at understanding, truth rewards us with inconsistent theories, untenable logics, and a deep feeling of bewilderment. It is as if our concept of truth lashes out at us because it wants to remain a mystery. After a few run-ins with truth, many philosophers have the good sense to keep their distance. Far from being the serene, profound concept most people take it to be, those of us who think much about the liar paradox know truth to be a vengeful bully—a conceptual misanthrope.

---

<sup>1</sup> I use ‘aletheic’ as an adjective meaning *pertaining to truth*.

Why has truth treated us in this way? And is there anything we can do about its misdeeds? I suggest that part of the blame falls on us, for there is a reason it is angry with us; truth is a bully, but it isn't a sociopath. Its wrath is partly a result of our insensitivity. We have tried to impose our conceptual will on truth; we have been unwilling to accept it for what it is. We have unreflectively assumed that all concepts are healthy in a certain sense, and in doing so, we have discriminated against truth. In short, we have treated truth as if it is a normal, healthy concept, when in fact, it is defective and its flaw is inherent. All the paradoxes associated with truth arise from this misunderstanding. As with most bullies, truth's misdeeds are cries for help. However, once we understand its specific defect, we should also recognize that there is no place in our conceptual repertoire for truth. Truth cannot be rehabilitated. Instead, it is time for truth to retire and for us to replace it with one or more healthy concepts that perform its role without causing us trouble. Only by adopting this strategy for handling truth can we finally put an end to its reign of terror.<sup>2</sup>

Accepting that truth is a particular type of defective concept and that we should no longer employ it does not relieve us of our explanatory responsibility. We are still in desperate need of an acceptable theory of truth and an account of the liar paradox, but before we can settle on the best explanation, we need a better grasp of how it has mistreated us. In particular, we need a better understanding of the revenge paradoxes.

I begin by distinguishing between two types of revenge paradoxes: inconsistency problems and self-refutation problems. These problems reinforce one another in the sense that attempts to avoid one tend to bring on the other. In the next section, I argue that if a theory of truth validates the truth rules (i.e., certain intuitively plausible rules governing the use of truth expressions), then either it is restricted from applying to certain languages, which renders it unacceptable, or it faces either an inconsistency problem or a self-refutation problem. Section three is where I use the

---

<sup>2</sup> In Schiffer's terminology, I am calling for an unhappy face solution to the liar paradox; see Schiffer (2003).

revenge paradox phenomenon to justify the theory of truth and the approach to the liar paradox I endorse.<sup>3</sup> I argue that truth displays a particular type of defect: it is an inconsistent concept; roughly, an inconsistent concept has incompatible rules governing the way it should be employed. I present three arguments for theories of truth on which truth is an inconsistent concept. The first argument is an abductive argument: if we accept that truth is an inconsistent concept, then we can explain the pattern of our failures to understand it. That is, the best explanation of why revenge paradoxes occur depends on the claim that truth is an inconsistent concept. The second argument is that if we assume that truth is a consistent concept that obeys the truth rules, then our only options are unacceptable theories. The third argument is that if we accept that truth is an inconsistent concept and we have a proper understanding of how to explain such concepts, then we can construct a theory of truth that: (i) implies that truth obeys the truth rules, (ii) avoids both types of revenge paradoxes, and (iii) does not have to be restricted in any way. Finally, in section four, I present an overview of the theory of truth and the approach I endorse to the liar paradox.

Before I move on to the substantive parts of the paper, I want to make a methodological point. I am concerned with the liar paradox as it arises in natural languages. Most of the work done by analytic philosophers on the liar paradox focuses on technicalities associated with constructing artificial languages; this is unfortunate. Of course, there is an important place for technical work on truth.<sup>4</sup> However, there is a tendency to get lost in the technical details and ignore how they relate to natural language. For example, one can construct a theory of truth and an artificial language such that the artificial language contains sentences that give rise to the liar paradox, and the theory of truth can handle any sentence that belongs to the language. One can even construct one's theory so that it is expressible in the artificial language without giving rise to revenge paradoxes in that

---

<sup>3</sup> I use both 'theory of truth' and 'approach to the liar paradox' in a loose way to include any set of claims about truth and any set of claims about how to deal with the paradox, respectively.

<sup>4</sup> For example, it is important for someone who wants to be able to do mathematical logic in an expressively rich language without worrying about the liar paradox.

language. That is, we can present an artificial language and a theory of truth for that language such that the theory is expressible in the language and applies to the entire language; hence, we no longer need a substantive object language / metalanguage distinction for certain theories of truth.<sup>5</sup>

Although this is a huge accomplishment, I am not concerned with a project of this type because it does not, by itself, constitute an acceptable approach to the liar paradox as it occurs in natural language. To constitute an approach to the liar as it occurs in natural language, a proponent of such a theory would have to claim that natural languages are relevantly similar to the artificial language, or that we should change our natural languages so that they are relevantly similar to the artificial language. It turns out that neither of these claims are tenable because these theories still give rise to revenge paradoxes when applied to anything like natural languages (more on this issue below).

Thus, there is more to an acceptable approach to the liar paradox (as it occurs in natural languages) than a theory of truth whose metalanguage is (or is a sublanguage of) its object language. Indeed, it seems to me that the biggest myth associated with contemporary work on the liar paradox is that a theory of truth is “revenge immune” if and only if it does not require a distinction between object language and metalanguage.<sup>6</sup> I agree that a theory of truth should apply to the language in which it is formulated, but an acceptable approach to the liar that works for natural languages requires much more than this.

## 1. REVENGE PARADOXES

Let us first take a look at the liar paradox.<sup>7</sup> The liar paradox involves sentences like the following (which I call a *liar sentence*):

---

<sup>5</sup> See McGee (1991), Field (2003a, 2003b, 2005a, 2005b, Forthcoming), and Maudlin (2004) for examples; I am not accusing any of these theorists in particular of getting lost in technical details.

<sup>6</sup> For an extended discussion of this point, see Scharp (TI).

<sup>7</sup> The liar is the most familiar paradox associated with truth, but there are others: the Curry paradox and the Yablo paradox. The Curry paradox is that, from intuitive assumptions, one can use the sentence ‘if this sentence is true, then god exists’ to derive that god exists (or any other absurdity). The Yablo paradox is that from intuitive assumptions, one

(1) (1) is false.

The paradox is that from intuitively plausible assumptions via intuitively plausible inferences, one can derive that (1) is both true and false. In fact, there are many different ways to derive this conclusion.<sup>8</sup> The most popular ones depend on T-sentences (i.e., sentences of the form:  $\langle p \rangle$  is true if and only if  $p$ )<sup>9</sup>, but I prefer one based on what I call the *truth rules*:

- (i) *ascending truth rule*:  $\langle \langle p \rangle$  is true  $\rangle$  follows from  $\langle p \rangle$ .
- (ii) *descending truth rule*:  $\langle p \rangle$  follows from  $\langle \langle p \rangle$  is true  $\rangle$ .
- (iii) *substitution rule*: two names that refer to  $\langle p \rangle$  are intersubstitutable in extensional occurrences of  $\langle \langle p \rangle$  is true  $\rangle$  without changing its truth-value.

The argument also depends on some of the inference rules of classical logic. On the one hand, assume that (1) is true. If (1) is true, then '(1) is false' is true (by substitution). If '(1) is false' is true, then (1) is false (by descending). Thus, if (1) is true, then (1) is false. On the other hand, assume that (1) is false. If (1) is false, then '(1) is false' is true (by ascending). If '(1) is false' is true, then (1) is true (by substitution). Thus, if (1) is false, then (1) is true. Therefore, (1) is true if and only if (1) is false. It follows that (1) is both true and false. Anyone who endorses a theory of truth that applies to a language with sentences like (1) must reject one of the premises, reject one of the inferences, or accept the conclusion.

For most of this paper, I assume the *principle of mono-aletheism*: no sentence is both true and false; however, I discuss approaches to the liar paradox based on rejecting it in section three.<sup>10</sup>

---

can prove contradictory consequences concerning an infinite descending sequence of sentences  $s_1, s_2, \dots$ , where  $s_1$  is 'for all  $i > 1, s_i$  is false' and  $s_2$  is 'for all  $i > 2, s_i$  is false', etc. See van Bentham (1978), Meyer, Routley, and Dunn (1979), Hazen (1990), Beall (1999), Field (2001, 2002, 2003a, 2003b, 2005b), for discussion of the Curry paradox; see Yablo (1985, 1993c), Hardy (1995), Tennant (1995), Priest (1997), Sorenson (1998), Beall (1999, 2001), Leitgeb (2002), Bueno and Colyvan (2003a, 2003b), and Ketland (2004) for discussion of the Yablo paradox.

<sup>8</sup> See Maudlin (2004) for discussion.

<sup>9</sup> ' $\langle \rangle$ ' and ' $\langle \rangle$ ' are angle quotes; ' $p$ ' serves as a sentential variable that can be replaced by a sentence, and ' $\langle p \rangle$ ' is the quote-name of such a sentence. I also use ' $p$ ' as a logical constant (e.g., in ' $p$  is true'). Note that these uses are distinct: an occurrence of ' $p$ ' cannot be both a sentential variable and a constant.

<sup>10</sup> Likewise, I assume the more complex versions of mono-aletheism: no sentence is both true and neither true nor false, and no sentence is both false and neither true nor false.

I also assume that *falsity* is defined in the usual way in terms of truth: the extension of falsity is just the anti-extension of truth, and the anti-extension of falsity is just the extension of truth.

Furthermore, I follow most of those who work on the liar paradox in assuming that sentences are primary truth bearers. I should note that the arguments I present do not depend on any particular choice of primary truth bearers.<sup>11</sup>

As I mentioned, a revenge paradox for a theory of truth T often involves a sentence that contains an expression used by T to classify liar sentences. Let us consider an example. Let T be a theory of truth that implies that truth expressions are partially defined predicates. That is, T implies that some sentences containing truth predicates are truth-value gaps—they are neither in the extension of ‘true’ nor in the anti-extension of ‘true’. Assume as well that T validates the truth rules and the other rules involved in the derivation of the liar paradox. Thus, ‘(1) is true if and only if (1) is false’ follows from T. However, no contradiction follows from ‘(1) is true if and only if (1) is false’ because we are working in a three-valued scheme. Indeed, T implies that (1) is a gap. Hence, (1) is not paradoxical for T. We can say that (1) is *pseudo-paradoxical* for T (i.e., (1) has traditionally been involved in a liar paradox, but it poses no problem for T).

So far so good for T; however there is trouble on the horizon. Consider another sentence:

(2) (2) is either false or a gap.

Notice that (2) contains ‘gap’, which is used by T to classify (1). Using the same resources needed to derive ‘(1) is true if and only if (1) is false’, we can derive ‘(2) is true if and only if (2) is either false or a gap’. On the one hand, assume that (2) is true. If (2) is true, then ‘(2) is false or a gap’ is true (by substitution). If ‘(2) is false or a gap’ is true, then (2) is false or a gap (by descending).

---

<sup>11</sup> When I say that sentences are *primary* truth bearers, I mean that I take sentential truth (i.e., the truth of a sentence) to be explanatorily primary; propositional truth, doxastic truth, etc. should be explained in terms of sentential truth. I adopt this view here because the vast majority of those who offer approaches to the liar paradox accept it. Moreover, it seems to me that approaches to the liar that depend on any particular choice of primary truth bearers (e.g., Glanzberg (2004)) face revenge paradoxes of their own.

Thus, if (2) is true, then (2) is false or a gap. On the other hand, assume that (2) is false or a gap. If (2) is false or a gap, then ‘(2) is false or a gap’ is true (by ascending). If ‘(2) is false or a gap’ is true, then (2) is true (by substitution). Thus, if (2) is false or a gap, then (2) is true. Therefore, (2) is true if and only if (2) is false or a gap. Given that (2) is either true, false, or a gap, a contradiction follows. Thus, T is inconsistent. Although T can handle sentences like (1), it cannot handle sentences like (2); (2) constitutes a revenge paradox for T.<sup>12</sup> That is our first example.

Let us consider how T might be altered to accommodate sentences like (2). One way to do so is to alter the logic we use so that the theory still validates the truth rules and still implies that (2) is a gap, but now the theory implies that ‘(2) is true if and only if (2) is false or a gap’ is a gap as well. Let us call this theory T’. Now we cannot derive a contradiction from ‘(2) is true if and only if (2) is either false or a gap’. However, this sentence poses another problem for T’. Namely, T’ implies that (2) is true if and only if (2) is either false or a gap; hence, T’ has ‘(2) is true if and only if (2) is either false or a gap’ as a consequence. However, T’ implies that ‘(2) is true if and only if (2) is either false or a gap’ is a gap; that is, T’ implies that ‘(2) is true if and only if (2) is either false or a gap’ is neither true nor false. Therefore, T’ implies that one of its consequences is not true. Consequently, T’ is self-refuting—it implies that it is not true.<sup>13, 14</sup> That is our second example.

Let us consider a way of altering the theory so that it is not self-refuting. We need a way of characterizing (2) and ‘(2) is true if and only if (2) is either false or a gap’ that does not result in the theory having a consequence that it labels untrue. One way to do this is to accept that (1) is a truth-value gap, but stipulate that the truth-value gaphood predicate itself is partially defined (i.e., the

---

<sup>12</sup> For an example of a theory like T, see Kripke (1975).

<sup>13</sup> For an example of a theory like T’, see Maudlin (2004).

<sup>14</sup> In the last two sentences of this paragraph, I am assuming that one can use ‘not’ in such a way that ‘p is not true’ follows from ‘p is neither true nor false’. I take it for granted that this is a legitimate use of ‘not’. In logic, a sentential operator with this property is often called *exclusion negation*. Thus, I am assuming that sometimes the English word ‘not’ expresses something like exclusion negation. Note that this assumption need not commit me to the claim that ‘not’ is ambiguous; see Horn (1989) and Atlas (1989) for discussion. When I am worried about misunderstandings, I use ‘Xnot’ to express exclusion negation.

gaphood predicate has gaps—*gaphood gaps*). Let  $T''$  be such a theory.  $T''$  implies that (1) is a truth-value gap.  $T''$  also implies that (2) is true if and only if (2) is either false or a truth-value gap. However,  $T''$  can be constructed so that it implies that ‘(2) is true if and only if (2) is false or a truth-value gap’ is true; the reason is that  $T''$  does not imply that (2) is either true, false, or a truth-value gap. Indeed  $T''$  implies that (2) is a gaphood gap. Of course, one can construct a new problematic sentence for  $T''$ :

(3) (3) is either false, a truth-value gap, or a gaphood gap.

However,  $T''$  can follow the same strategy to handle (3) by positing a hierarchy of gaphood predicates, each of which is partially defined. On this account (1) is a truth-value gap, (2) is a gaphood gap, (3) is a gaphood-hood gap, etc. In this way,  $T''$  avoids labeling any of its consequences untrue.<sup>15</sup>

The problem with  $T''$  is that if it applies to a language that contains a completely defined truth-value gaphood predicate, then  $T''$  is inconsistent because it implies that a sentence of this language like (2) (i.e., a sentence that attributes either falsity or truth-value gaphood to itself—where the truth-value gaphood predicate is completely defined) is true if and only if it is either false or a truth-value gap. Thus,  $T''$  faces an inconsistency problem. The progression from  $T$  to  $T'$  to  $T''$  illustrates the fact that there is something like an oscillation between the two kinds of revenge paradoxes—attempts to avoid one tend to bring on the other.<sup>16</sup>

It is my view that one must distinguish between these two types of revenge paradoxes in order to understand our current predicament regarding truth. In short, there are two broad trends when it comes to theories of truth designed to handle sentences like (1). Some theories can handle sentences like (1), but they still have inconsistent consequences for other sentences (e.g., (2)). That is, they do not provide a way of solving all other paradoxes associated with truth that are

---

<sup>15</sup> For an example of a theory like  $T''$ , see Field (2003a, 2003b, 2005a, 2005b, Forthcoming).

<sup>16</sup> I borrow the term ‘oscillation’ from McDowell (1994).

structurally identical to the liar. This is the inconsistency problem. Other theories can handle sentences like (1), but they imply that they have the same status (i.e., being untrue) as (1). Because few, if any, theories of truth that are designed to handle sentences like (1) imply that sentences like (1) are true, a theory of truth that implies that it has the same status as a liar sentence implies that it is untrue. This is the self-refutation problem.<sup>17</sup>

The inconsistency problem arises when a theory of truth handles some versions of the liar paradox, but not all of them. There are many different versions of the liar; some versions involve concepts that are often used to classify sentences that figure in other versions. This should not come as a surprise given the prominence of views on which sentences that figure in liar paradoxes are defective in a way that renders them neither true nor false. Once one has a term for the third status, one has a new version of the liar paradox. The most common response to the inconsistency problem is to restrict the theory so that it does not apply to such sentences.

On the other hand, the self-refutation problem arises in connection with the consequences of a theory of truth. The liar paradox is unlike other paradoxes (e.g., Russell's paradox, Grelling's paradox, etc.) in that it concerns truth, which applies to things that can participate in inferential relations (e.g., sentences, propositions, etc.). In other cases, the paradoxical items (e.g., sets, predicates, etc.) are not the type of thing that can be the consequence of a theory. However, for truth, the paradoxical items are sentences, which can be consequences of a theory. A theory of truth that is designed to deal with the liar paradox has to classify many paradoxical sentences like (1). It turns out that for many theories of truth, no matter what they say about such sentences, some of these sentences are going to be consequences of the theory.

---

<sup>17</sup> Both the Curry paradox and the Yablo paradox depend on the truth rules as well (i.e., they require all three rules for their construction), and approaches to each one generate revenge paradoxes in the same way that approaches to the liar paradox generate revenge paradoxes; thus, one can use structurally analogous arguments to the ones in this paper to argue for conclusions pertaining to the Curry and the Yablo that are analogous to the conclusions I draw pertaining to the liar. It is my view that all three paradoxes (i.e., the liar, the Curry, and the Yablo) are manifestations of the defectiveness of our concept of truth. The approach to truth that I offer solves all three without generating revenge paradoxes of any kind.

This way of formulating the self-refutation problem is somewhat misleading because it makes it seem as though the class of paradoxical sentences is fixed. In fact, we should think of paradoxicality as relative to a theory of truth, and we should distinguish between paradoxicality and pseudo-paradoxicality. A sentence is *paradoxical* for a theory of truth if and only if the theory of truth either has contradictory consequences for it or has consequences for it that the theory implies are untrue. A sentence is *pseudo-paradoxical* for a theory of truth if and only if it is not paradoxical for the theory in question, but it is paradoxical for the naïve theory of truth, which implies that truth is completely defined and obeys all the principles we commonly take truth to obey (e.g., the truth rules, rules about how it interacts with sentential operators, etc.). The sentences that are pseudo-paradoxical for a theory of truth (e.g., (1)) figure in the liar paradox, but the theory can handle them. The sentences that are paradoxical for a theory of truth figure in revenge paradoxes for the theory, which the theory cannot handle. The class of sentences that are pseudo-paradoxical for a theory of truth and the class of sentences that are paradoxical for a theory of truth depend on the way the theory of truth classifies the liar (e.g., as gappy, as indeterminate, as uncathegorical, etc.). For example, the revision theory of truth implies that the liar is uncathegorical. Thus, a revenge paradox for it concerns the sentence:

(4) (4) is either false or uncathegorical.

That is, sentence (4) is paradoxical for the revision theory of truth. However, an indeterminacy theory of truth (i.e., one on which the liar is indeterminate) has no problem with (4) because it does not imply that the liar is uncathegorical. Thus, (4) is not paradoxical for an indeterminacy theory of truth. It turns out that if a theory of truth validates the truth rules, then no matter what it says about the liar sentence, the set of sentences that are pseudo-paradoxical for it will include some of its own consequences (unless it is restricted so that it does not apply to them). Given that an acceptable theory of truth does not imply that pseudo-paradoxical sentences are true, one can either restrict

one's theory so that it does not apply to the pseudo-paradoxical sentences that are outside its scope, or one can bite the bullet and accept that one's theory implies that some of its consequences are untrue.

## 2. THE REVENGE ARGUMENT

In this section, I present a criticism of theories of truth that offer approaches to the liar paradox on which truth is pretty much as we take it to be. In particular, it is a criticism of theories of truth that validate the truth rules (i.e., theories of truth that imply that the truth rules are valid for some class of sentences that includes some liar sentences). I argue that any theory of truth that implies that the truth rules are valid is either: (i) inconsistent, (ii) self-refuting, or (iii) restricted so that it does not apply to certain sentences that contain truth predicates. Although there are many theories of truth that offer approaches to the liar paradox on which one or more of the truth rules are not valid, I do not address them here.<sup>18</sup> It is my view that the truth rules are constitutive of our concept of truth—any theory of truth that implies that truth does not obey them is unacceptable. Of course, that claim is not intended to be a criticism. However, one can develop it into a criticism that shows these theories to be unacceptable, but I don't have the space to do so here.<sup>19</sup>

Assume that  $T$  is a theory of truth and  $T$  implies that the truth rules are valid for a class of sentences that includes (1). Assume also that  $T$  implies that truth predicates are univocal, invariant, non-circular, etc; in short, truth predicates do not have any "hidden" semantic features that render the reasoning in the liar paradox invalid. Note that this assumption does not add much because theories of truth that imply that truth predicates have "hidden" semantic features don't usually

---

<sup>18</sup> For example, theories of truth that treat natural language truth predicates as context-dependent (e.g., Parsons (1974), Burge (1979), Barwise and Etchemendy (1987), Gaifman (1992, 2000), Koons (1992, 2000), Simmons (1993), and Glanzberg (2004)), theories of truth that treat truth as a circular concept (e.g., Herzberger (1982a, 1982b), and Gupta and Belnap (1993)), theories of truth that reject the substitution rule (e.g., Skyrms (1982), and theories of truth that reject the ascending rule (e.g., Feferman (1982)).

<sup>19</sup> See Scharp (AP) for this criticism.

validate the truth rules (at least, I am unaware of any that do).<sup>20</sup> Finally, assume that T applies to a language that contains liar sentences. Let a *liar sentence* be any sentence that attributes falsity and only falsity to itself. Thus, sentence (1), ‘this sentence is false’, and ‘the sentence named by the third singular term used in the sixth sentence of the second paragraph of the second section of “Aletheic Vengeance” is false’, are liar sentences.

There is plenty to say about what languages and sentences are, and about what it is for a theory to apply to a particular language or to a particular sentence, but I want to leave them at an intuitive level. In addition, there is plenty to say about the conditions under which a language contains a liar sentence; however, it is my view that because of the prevalence of *empirically paradoxical sentences* (i.e., sentences that are paradoxical because of some empirical facts) and *inter-linguistic truth attributions* (i.e., sentences of one language that attribute truth or falsity to sentences of other languages) it is impossible to provide a non-circular account of the conditions under which such sentences arise in natural languages.<sup>21</sup>

There are very few choices for the way in which a theory of truth that validates the truth rules classifies a liar sentence. The theory can imply that liar sentences are false or the theory can imply that liar sentences are true, but given that the theory implies that the truth rules are valid, a theory of either type implies that liar sentences are true if and only if they are false; for a theory that classifies the liar as true or false, ‘(1) is true if and only if (1) is false’ is a contradiction. Therefore, an acceptable theory of truth that validates the truth rules will not classify liar sentences as true or false.

Instead of classifying (1) as true or as false, T can classify (1) as neither true nor false. We English speakers find this a natural description of the case. That is, we find it natural to say that

---

<sup>20</sup> For examples of theories of truth that imply that truth predicates have “hidden” semantic features, see Burge (1979), Gupta and Belnap (1993), and Williamson (2000).

<sup>21</sup> On the former, see Scharp (RB) and on the latter see Scharp (TI) and Eklund (Forthcoming).

certain things are neither true nor false (e.g., acorns). We also find it natural to say that such things are not true and not false. Here we are using ‘not’, but this use is distinct from the use of ‘not’ in ‘a sentence that is not true is false’; the former ‘not’ expresses exclusion negation, while the latter expresses choice negation.<sup>22</sup> In English, we sometimes use ‘not’ to express exclusion negation as above. Other times we use ‘not’ to express choice negation. I assume that it can express exclusion negation and that it can express choice negation.

If T implies that the liar is a truth-value gap, then we can construct another sentence that causes problems for T. Sentence (2) (i.e., ‘(2) is either false or a gap’) is classified as a gap by T and, hence, it is a consequence of T; that is, if T implies that (2) is a gap, then T implies that (2) is either false or a gap. Thus, T has (2) as a consequence and T implies that (2) is untrue.<sup>23</sup> Thus, T faces a self-refutation problem.

In addition, T faces an inconsistency problem. Recall that if T validates the truth rules, then ‘(1) is true if and only if (1) is false’ is a consequence of T. The approach to the liar on which truth is treated as a partially defined concept (i.e., on which (1) is a gap) handles (1) because both (1) and ‘(1) is true if and only if (1) is false’ are gaps. The catch is that ‘(1) is true if and only if (1) is false’ is not a genuine contradiction according to this theory; that is, one cannot prove ‘(1) is true and (1) is not true’ from ‘(1) is true if and only if (1) is false’ because the proof depends on the principle of bivalence for (1) (i.e., (1) is either true or false), which the partiality approach denies. However, the partiality approach cannot rely on the same trick when it comes to sentence (2). The sentence ‘(2) is true if and only if (2) is either false or a gap’ follows from T by an argument that is structurally identical to the one that shows ‘(1) is true if and only if (1) is false’ follows from T. Given that a sentence is either true, false, or a gap, one can derive that (2) is both true and either false or a gap

---

<sup>22</sup> See footnote 14.

<sup>23</sup> This argument depends on the standard rule of disjunction introduction.

from this consequence. Hence, '(2) is true if and only if (2) is either false or a gap' is a consequence of T and it is a genuine contradiction. Thus, T faces an inconsistency problem.

In summary, if T validates the truth rules and T implies that liar sentences have status  $\Delta$ , where a sentence is  $\Delta$  only if it is Xnot true, then there are three options for T:

- (i) T implies that (2'), '(2') is either false or  $\Delta$ ', is true.
- (ii) T implies that (2') is false.
- (iii) T implies that (2') is  $\Delta$ .

On any of these options, T implies '(2') is true if and only if (2') is false or  $\Delta$ '. If T classifies this sentence as true, then T is inconsistent. If T classifies it as false, then T is self-refuting. If T classifies it as gappy, then T is self-refuting. Therefore, T is either inconsistent or self-refuting.

Given the massive amount of work on the liar paradox and the ridiculously sophisticated logical tools that have been marshaled to combat it, the reader *should* be skeptical when presented with such a simple argument that is touted as a refutation of most prima facie plausible approaches to the liar paradox. Although the distinction between the inconsistency problem and the self-refutation problem is new, the revenge argument should not come as a surprise to any of the veterans of our battles with the liar paradox. In fact, most of them have been hard at work devising plans to avoid arguments like this one. So why have I presented it as a central insight into the nature of truth and the liar paradox? It might seem like I am throwing a rock at an army battalion; if so, read on—it turns out that the appearance of a battalion is nothing but a mirage and a thrown rock is a fine way to expose it as such. That is, the real insight is that there is no acceptable way of avoiding the revenge argument; thus, the real work is done in the objections and replies, to which I now turn.

*Objection 1:* One can avoid the liar paradox and both types of revenge paradoxes by assuming that paradoxical sentences are meaningless or ill-formed.

*Reply 1:* Strictly speaking, this is not an objection to the revenge argument—it offers an approach to the liar that would take care of the liar and all the revenge paradoxes. However, it is instructive to see why this sort of approach fails. The most obvious problem with this objection is that there is no independent reason to think that paradoxical sentences are meaningless or ungrammatical. In fact, if one were to adopt such an account, one would have to reject our most popular theories of meaningfulness and theories of grammar. There is another reason to reject these accounts: paradoxicality is not determined by meaning and grammar. That is, one can specify two sentence tokens of the same type that have the same sentential meanings, the same subsentential meanings for their subsentential parts, and the same referents for their singular terms, but one is paradoxical and the other is not. Paradoxicality can depend on virtually any fact one can imagine, while meaningfulness and grammaticality do not. Thus, if one accepts that paradoxical sentences are meaningless or ungrammatical, then one has to accept that whether a sentence is meaningful or grammatical can depend on virtually any fact that one can imagine, which is radically implausible.<sup>24</sup>

*Objection 2:* The revenge argument involves not just inferences licensed by the truth rules, but inferences of classical logic as well. If one endorses a non-classical logic as part of one's approach to the liar paradox, then one can avoid both types of revenge paradoxes.

*Reply 2:* I do not deny that many approaches to the liar paradox involve non-classical logics. However, using this move to block the revenge argument has several problems. One problem is that the classical inference rules needed to derive the troublesome conclusion (i.e., '(2) is true if and only if (2) is either false or gappy') are minimal. All one really needs is a conditional that obeys the natural deduction inference rule of conditional proof (alternatively, a conditional for which one can prove a deduction theorem); a conditional with this property is required if anything like everyday

---

<sup>24</sup> See Kripke (1975) for a similar point; see also Scharp (RB) for discussion.

reasoning is possible in the language. These inference rules are going to be valid for any natural language; thus, a theory of truth that rejects them will not apply to natural languages.

Moreover, a theory of truth that avoids the revenge argument by denying one of the inference rules involved (except for the truth rules of course) would still have to be restricted so that it does not apply to languages for which the inference rule in question is valid. Let  $T$  be a theory of truth that applies only to languages for which a certain inference rule  $R$  involved in the revenge argument is invalid. Let  $L$  be a language for which  $R$  is valid and let  $L$  contain a truth expression. Of course,  $T$  does not apply to  $L$ ; thus,  $T$  is restricted so that it does not apply to some languages that contain truth expressions. A theory of truth that avoids the liar paradox or the revenge paradoxes only by denying certain inference rules of classical logic is a theory that is restricted so that it does not apply to certain languages containing truth expressions. Thus, the conclusion of the revenge argument (i.e., a theory of truth that validates the truth rules is either inconsistent, self-refuting, or restricted) withstands this objection.<sup>25</sup>

*Objection 3:* Instead of treating the expression for truth value gaps as completely defined, which is an assumption needed to derive the inconsistency problem in the revenge argument, one can assume that ‘gappy’ is itself gappy. Then one can treat (1) as a truth-value gap and one can treat (2) as a gaphood gap. Indeed, one can define a hierarchy of partially defined gaphood predicates that can be used to classify all the sentences that seem to give rise to liar paradoxes, and one can do so without facing either a self-refutation problem or an inconsistency problem.

*Reply 3:* I agree that one can provide a theory of truth of this sort and an artificial language with predicates like these such that none of the sentences of the language give rise to revenge paradoxes for the theory.<sup>26</sup> However, that does not constitute an acceptable approach to the liar

---

<sup>25</sup> I suppose that one could deny that such languages exist, but it would follow that there is no language such that it is one in which we can reason normally and it contains a truth expression; that seems radically counterintuitive to me.

<sup>26</sup> See Field (2003a, 2003b, 2005a, 2005b, Forthcoming) for an example of such a theory.

paradox because the theory has to be restricted so that it does not apply to languages that contain completely defined gaphood predicates; otherwise, sentences like (2) in which completely defined gaphood predicates occur are paradoxical for the theory (I discussed this problem in section one).

The objector might respond by claiming that if there were a language that obeys the logic posited by the theory of truth in question and contains a truth predicate and a completely defined indeterminacy predicate, then it would be trivial in the sense that anything would be derivable in it; thus, if the theory is right, then languages with truth predicates don't have completely defined indeterminacy predicates. My reply is that we can just stipulate that some language contains a truth predicate and a completely defined indeterminacy predicate; thus, the theory of truth in question won't apply to this language because whatever the right logic is for the language, it won't be the one posited by the theory. It is common to assume that we can stipulate the syntactic features of a language and that we can stipulate that a certain word expresses a certain concept. Thus, it makes sense to think that we can stipulate that a language has a truth expression and a completely defined indeterminacy predicate.<sup>27</sup>

*Objection 4:* One can arrive at a satisfactory theory of truth by restricting it so that it does not apply to sentences that give rise to inconsistency problems or self-refutation problems. The most familiar strategy of this type is to assume that the theory is formulated in one language (the metalanguage) and applies only to languages that are expressively weaker in certain ways (the object languages). A theory of this type is restricted from applying to languages that have the expressive resources required to formulate the theory. However, some more recent theories of truth do not appeal to the distinction between object language and metalanguage, but they still have to be restricted to avoid revenge paradoxes. For example, Field's theory of truth applies unproblematically to certain artificial languages that have the resources to formulate the theory;

---

<sup>27</sup> Some philosophers deny that we have this stipulative power, but I cannot take issue with them here; see Williamson (1997).

however, it does not apply to languages that contain completely defined indeterminacy predicates (the linguistic expressions that give rise to revenge paradoxes for Field's theory—completely defined indeterminacy predicates—aren't required to formulate the theory).

*Reply 4:* Although almost everyone who presents an approach to the liar resorts to this move in one way or another, it is unacceptable because it results in a theory of truth that does not even apply to all truth expressions. Desperation has overwhelmed common sense in this case. It is as if these philosophers are saying “look everyone, I have come up with a theory of chairs!” When a critic objects, “your theory implies that *this* chair is both black and not black—that result refutes your theory,” the theorist responds, “oh, my theory doesn't apply to *that* chair.” We all *should* agree that this response is totally unacceptable. It often comes as a shock to those outside philosophical logic that this sort of move is tolerated for truth theorists who offer approaches to the liar.

It seems to me that this response to theories of truth that are restricted to avoid revenge paradoxes should be sufficient. However, I expect that those philosophers who have been hardened by combat with the liar will be deaf to this sort of criticism. In a companion paper, I argue that theories of truth that have been restricted to avoid revenge paradoxes are unacceptable.<sup>28</sup> There I argue that if T is a theory of truth that is restricted to avoid revenge paradoxes and L is a natural language, then there are sentences of L that give rise to revenge paradoxes for T; thus, if T applies to these sentences of L, then T is either inconsistent or self-refuting. The key to the argument is constructing a sentence of L that attributes truth indirectly to a sentence of some other language that gives rise to a revenge paradox for L; that is, if a theory of truth faces revenge paradoxes, then one can “import” one of these revenge paradoxes into a natural language. It is important that one can construct such a sentence even if one can appeal only to language specific concepts of truth (e.g.,

---

<sup>28</sup> See Scharp (IT).

truth-in-L, truth-in-English). Therefore, if a theory of truth is restricted to avoid revenge paradoxes, then it does not successfully apply to natural languages.

*Objection 5:* All the linguistic expressions used to construct revenge paradoxes are meaningless. The newest generation of approaches to the liar paradox show that one can construct a theory of truth that applies to the language in which it is formulated and classifies all the sentences of that language without giving rise to revenge paradoxes.<sup>29</sup> Of course, these theories seem to face revenge paradoxes when applied to languages containing other linguistic resources. However, one can treat these linguistic expressions as meaningless and avoid the revenge paradoxes altogether. If one takes this path, then one does not even have to restrict one's theory of truth.

*Reply 5:* Some philosophers do try to avoid restricting their theories of truth by claiming that the resources that give rise to the revenge paradoxes are meaningless or unintelligible.<sup>30</sup> I call this the *unintelligibility maneuver*. My view is that it is unacceptable to assume that these linguistic expressions are meaningless. As I have said, for a language that contains truth value gaps, one can define two sentential operators that behave like classical negation:

<u>p</u>	<u>~ p</u>	<u>¬ p</u>
T	F	F
F	T	T
G	G	T

The first one ('~') is choice negation and the second ('¬') is exclusion negation. Both can be expressed in English.<sup>31</sup> Even if we assume that some sentences are neither true nor false, one might infer from the claim that a sentence is not true that the sentence is false. If this inference is appropriate, then the negation involved is choice negation. On the other hand, one might infer from the claim that a sentence is neither true nor false that the sentence is not true. If this inference is

---

<sup>29</sup> See McGee (1991), Field (2002, 2003a, 2003b, 2004, 2005a, 2005b, Forthcoming), and Maudlin (2004) for examples.

<sup>30</sup> See Parsons (1984), Priest (1990), and Tappenden (1999), who claim that there is no such thing as exclusion negation; see also Maudlin (2004), who claims that there are no non-monotonic sentential operators whatsoever.

<sup>31</sup> For evidence of this claim see Atlas (1989) and Horn (1989).

appropriate, then the negation involved in the conclusion is exclusion negation. A theory like McGee's or Field's or Maudlin's has trouble with sentences that express exclusion negation for two reasons. First, these theories are fixed point theories and so apply only to languages that do not contain non-monotonic sentential operators.<sup>32</sup> However, exclusion negation is non-monotonic. Thus, they do not even return results for languages that express exclusion negation. Second, one can easily extrapolate to determine the results they would return if they were capable of returning results. The sentence

(5) (5) is Xnot true.

poses a problem. It means something like (5) *has a status other than that of being true*. One can use (5) to generate a revenge paradox for most theories of truth that imply that (5) is Xnot true. Likewise, one can generate revenge paradoxes using completely defined gaphood predicates (as I did in the revenge argument), paradoxicality predicates, groundedness predicates, certain conditionals, quantification over hierarchies of predicates, and so on. In order to pursue the strategy advocated in the objection, one would have to claim that all these linguistic items are meaningless.

I take it for granted that if there is an established practice of using a linguistic expression, then that linguistic expression is meaningful.<sup>33</sup> For each of the linguistic expressions that are labeled unintelligible by these theorists, there is an established practice of using them. Moreover, these linguistic expressions belong to some natural languages, including English. Furthermore, anyone who claims that the linguistic expressions involved in revenge paradoxes are meaningless

---

<sup>32</sup> In a three-valued scheme, a sentential operator is *monotonic* if and only if for a sentence containing that sentential operator, changing a component of that sentence from a gap to a truth-value (i.e., from a gap to true or from a gap to false) never results in changing the sentence from one truth-value to the other or from a truth-value to a gap (i.e., from true to false, from false to true, from true to a gap, or from false to a gap). Intuitively, one can "fill in" the gaps in the components without changing the truth-value of the compound. See Gupta and Martin (1984) who show that by using a weak Kleene scheme, one can arrive at fixed points even though one's language contains certain non-monotonic operators. However, exclusion negation is not among them.

<sup>33</sup> Even linguistic expressions like 'tonk' are meaningful; they just express inconsistent concepts; see Prior (1960) for a discussion of 'tonk'. Of course, Wittgenstein (1923) is infamous for claiming that many seemingly meaningful words are "nonsense" (*unsinnig*); I don't have the space to discuss the relation between his views on language and the claim on which this footnote comments.

will have to claim that logicians and linguists have been wasting their time studying exclusion negation, other non-monotonic sentential operators, and the rest of the “outlaw” linguistic expressions.

In addition, the “outlaw” linguistic expressions serve an important explanatory role. If we decided that they are all meaningless and gave them up, then we would rob natural languages of important expressive resources. For example, if an object,  $A$ , is in neither the extension nor the anti-extension of a predicate,  $\phi$ , then we need a way of expressing this fact. One way of doing so is to say that  $A$  is  $X$ not  $\phi$  and  $A$  is  $X$ not  $[\sim \phi]$ . Another is to say that  $A$  is a  $\phi$ -value gap. If the theorists in question are right that the “outlaw” linguistic expressions are meaningless, then we have no way of expressing these facts.

Finally, simply claiming that the linguistic resources in question are meaningless is not enough to avoid the revenge argument. One would have to provide an independent argument for this claim (e.g., something other than, “that’s the only way to avoid the liar paradox”). No such argument has been forthcoming and it seems it would be impossible to present one whose premises were more plausible than the claim that these items are meaningful.<sup>34</sup>

Analytic philosophy has a long history of claiming that certain linguistic expressions that figure in established linguistic practices are meaningless. It is high time for us to realize that we need no longer resort to this kind of move; we can provide an approach to the liar paradox without it.

*Objection 6:* Theories of truth that are restricted to avoid revenge paradoxes can be thought of as revisionary, not descriptive. That is, one can treat them as prescribing how we should use truth predicates instead of describing how we do use them. If one treats a restricted theory in this way, then a proponent of such a theory advocates eliminating from natural language the resources

---

<sup>34</sup> See Eklund (Forthcoming) for another criticism of the unintelligibility maneuver that focuses on exclusion negation.

that contribute to revenge paradoxes (e.g., non-monotonic sentential operators, gapless gaphood predicates, etc.).

*Reply 6:* I think it is fine to treat these theories as revisionary, and there is an important place for such theories.<sup>35</sup> However, qua revisionary theories of truth, they face several problems. First, they say nothing about natural languages as they are now. Thus, they do not really provide approaches to the liar paradox at all. Second, even as revisionary theories of truth, they fail. As long as one has a truth predicate in the language that obeys the truth rules, one can “import” the revenge paradoxes into the language even though the language does not have the “outlaw” linguistic resources. Assume that T is a theory of truth that offers an approach to the liar paradox and that T is a revisionary theory of truth—it implies that we should change English so that the linguistic items involved in revenge paradoxes (e.g., exclusion negation, other non-monotonic operators, gapless gaphood predicates, etc.) are no longer part of English. Call the new language English\*. Assume also that T validates the truth rules. I am willing to grant that T might be expressible English\* and that T can adequately classify all the sentences of English\* that involve truth attributions to sentences of English\*.<sup>36</sup> The problem arises with sentences of English\* that attribute truth to sentences of other languages. Given that the revenge argument is correct (i.e., sound), T faces revenge paradoxes when applied to other languages. Let L be a language that contains the resources needed to construct a revenge paradox for T. Although T is restricted so that it does not apply to L, we know that if it did apply to L, then it would be inconsistent or self-refuting because L has sentences that would give rise to revenge paradoxes for T. Call such sentences *potentially paradoxical for T*. I argue that English\* contains potentially paradoxical sentences for T; hence, to avoid inconsistency or self-refutation, T has to be restricted so that it does not apply to certain

---

<sup>35</sup> For example, these theories can prescribe the parts of language that are acceptable for formulating arguments in mathematical logic.

<sup>36</sup> Field’s theory is a good example of a theory like T. Others include Maudlin’s and McGee’s.

sentences of English\*. Moreover, there is no apriori way to specify which sentences of English\* are potentially paradoxical for T. Thus, even though English\* does not contain any of the explicitly problematic linguistic resources, T is unable to apply to all the sentences of English\*.

Consider the inconsistency case. Let  $\phi$  be a sentence of L that is potentially paradoxical for T, let the potential paradox be an inconsistency problem, and let English\* have a name ' $\phi$ ' for  $\phi$ . Let  $\psi$  be the English\* sentence, ' $\phi$  is true'. If  $\phi$  figures in an inconsistency problem for T, then  $\psi$  figures in an inconsistency problem for T. Therefore, if T does not face an inconsistency problem, then it is restricted so that it does not apply to  $\psi$ .

In order to justify the claim that if  $\phi$  is potentially paradoxical for T, then  $\psi$  is too, we need to consider an example (remember that whether a sentence is paradoxical for a theory of truth depends on the features of the theory). Assume that T is a version of Kripke's theory of truth. A revenge liar for Kripke's theory is:

(2) (2) is either false or a gap.

Assume that (2) is a sentence of L, that English\* does not contain sentences like (2) because English\* does not have a gaphood predicate, and that T is restricted so that it does not apply to L. If T did apply to L, then T would imply that (2) is both true and either false or a gap. However, as long as English\* contains a name, '(2)' for (2), English\* contains:

(6) (2) is true.

If T applies to (6), then T implies that (6) is both true and either false or a gap. Assume that (6) is true. If (6) is true, then '(2) is true' is true (substitution). If '(2) is true' is true, then (2) is true (descending). If (2) is true, then '(2) is either false or a gap' is true (substitution). If '(2) is either false or a gap' is true, then (2) is either false or a gap (descending). If (2) is either false or a gap,

then ‘(2) is true’ is either false or a gap (ascending).<sup>37</sup> If ‘(2) is true’ is either false or a gap, then (6) is either false or a gap (substitution). Hence, if (6) is true, then (6) is either false or a gap. Assume that (6) is either false or a gap. If (6) is either false or a gap, then ‘(2) is true’ is either false or a gap (substitution). If ‘(2) is true’ is either false or a gap, then (2) is either false or a gap (descending).<sup>38</sup> If (2) is either false or a gap, then ‘(2) is either false or a gap’ is true (ascending). If ‘(2) is either false or a gap’ is true, then (2) is true (substitution). If (2) is true, then ‘(2) is true’ is true (ascending). If ‘(2) is true’ is true, then (6) is true (substitution). Hence, if (6) is either false or a gap, then (6) is true. Therefore, (6) is true if and only if (6) is either false or a gap. It follows that (6) is both true and either false or a gap. The self-refutation case is analogous.<sup>39, 40</sup>

One might stipulate that English\* does not have names for sentences of L, so this way of “importing” revenge paradoxical sentences into English\* doesn’t work. The problem with this response to the argument is that there are plenty of ways English\* might have to refer to sentences that are potentially paradoxical for T. For example, English\* might have a definite description that picks out a potentially paradoxical sentence for T, a sentence of English\* containing a demonstrative might refer to a potentially paradoxical sentence for T, a pronoun of English\* might

---

<sup>37</sup> It might seem that this inference is not licensed by the truth rules and instead requires a rule specific to weak truth. In three-valued settings, it is common to distinguish between weak truth and strong truth, where *weak truth* obeys the rule ‘if p is gappy, then ‘p is true’ is gappy’, but *strong truth* obeys the rule ‘if p is gappy, then ‘p is true’ is false’. As far as I know, no one has noticed that one can justify the weak truth rule by appeal to the truth rules and the claim that they can be applied not only to an entire formula, but to the subformulas of a formula. We can represent the weak truth rule as: from  $\langle\langle p \rangle \text{ is gappy} \rangle$  infer  $\langle\langle p \rangle \text{ is true} \rangle \text{ is gappy}$ . Given our definition of the gaphood predicate, this rule becomes: from  $\langle \neg \langle\langle p \rangle \text{ is true} \rangle \wedge \neg \sim \langle\langle p \rangle \text{ is true} \rangle \rangle$  infer  $\langle \neg \langle\langle\langle p \rangle \text{ is true} \rangle \text{ is true} \rangle \wedge \neg \sim \langle\langle\langle p \rangle \text{ is true} \rangle \text{ is true} \rangle \rangle$ . We can derive this rule using the ascending truth rule on subformulas:

1.  $\neg \langle\langle p \rangle \text{ is true} \rangle \wedge \neg \sim \langle\langle p \rangle \text{ is true} \rangle$
2.  $\neg \langle\langle\langle p \rangle \text{ is true} \rangle \text{ is true} \rangle \wedge \neg \sim \langle\langle p \rangle \text{ is true} \rangle$
3.  $\neg \langle\langle\langle p \rangle \text{ is true} \rangle \text{ is true} \rangle \wedge \neg \sim \langle\langle\langle p \rangle \text{ is true} \rangle \text{ is true} \rangle$

It is common practice to apply an inference rule to subformulas so long as it is an equivalence rule. Given that I have endorsed both the ascending and descending truth rules, we have essentially an equivalence rule. There are plenty of issues associated with this move (e.g., freestanding vs. ingredient content, categorical vs. hypothetical inference rules, etc.), but I do not have the space to discuss them.

<sup>38</sup> Again, this inference requires using the descending truth rule on a subformula.

<sup>39</sup> See Eklund (Forthcoming) for a different sort of “importation” argument.

<sup>40</sup> This argument highlights the fact that if a truth expression obeys the truth rules and applies to sentences of languages other than the one to which it belongs, then applying the descending truth rule can take one from a sentence of one language to a sentence of another. Although this consequence doesn’t seem problematic to me, I don’t have the space to defend it here.

refer to a potentially paradoxical sentence for T, a quantifier of English might range over a potentially paradoxical sentence for T, etc. There is no way to specify which of English\*'s linguistic resources might refer to a potentially paradoxical sentence of T because it might depend on all sorts of empirical facts (e.g., ‘the last sentence Russell uttered is true’ might be potentially paradoxical for T if the empirical facts turn out unfavorably). In addition, even if English\* has no way to refer directly to a potentially paradoxical sentence for T, we can still “import” a revenge paradox into English\* by considering a sentence of English\* that attributes truth to a sentence  $\phi$  of some language L, where  $\phi$  attributes truth to a sentence  $\theta$ , which is potentially paradoxical for T. If one wants to ensure that no sentence of English\* indirectly attributes truth to a sentence that is paradoxical for T, then one has to eliminate from English most of the linguistic resources we assume natural languages have. If instead one wants to ensure that T does not apply to any potentially paradoxical sentences for T, then one has to restrict T so that it does not apply to most of the sentences of English\* that contain truth predicates. Either way, T fails to offer an acceptable approach to the liar paradox.

*Objection 7:* Some of the replies I have offered (i.e., 2 and 6) depend on inter-linguistic truth attributions (i.e., sentences of the form ‘p is true’ where ‘p’ names a sentence of some other language). There are two problems with this move. First, all the theorists being discussed provide theories for language-specific truth predicates. A *language-specific truth predicate* (an *LS truth predicate*) is satisfied only by true sentences of a particular language. For example, ‘true-in-English’ is an LS truth predicate: ‘p is true-in-English’ is true if p is a true sentence of English, and it is false if either p is a false sentence of English or p is a sentence of some other language. Most of these theorists claim that natural language truth predicates can be explained in terms of LS truth predicates. The arguments given so far fail for LS truth predicates and, hence, for natural language truth predicates (as long as we explain the latter in terms of the former). For example, if T is a

theory of truth-in-L for a particular language L, and T offers a partiality approach to the liar paradox, then one might think that T would face a revenge paradox from the following sentence of some other language, L:

(2'') (2'') is either false-in-L or gappy-in-L.

However, the reasoning employed in the revenge argument to derive either an inconsistency problem or a self-refutation problem fails for (2''). Indeed (2'') is either false-in-L or gappy-in-L.<sup>41</sup> Therefore, none of the arguments offered so far even threaten the theories of truth that have been discussed.

Second, I have assumed both that sentences are primary truth bearers and that natural language truth predicates are unrestricted (i.e., that they are not LS truth predicates). However, these two commitments are incompatible. Consider a sentence that is a member of English and German (e.g., in English, 'Kripke rang' means that Kripke rang, and in German, it means that Kripke wrestled). If the truth conditions for the sentence qua member of English are satisfied and the truth conditions qua member of German are not, then the sentence is both true and false (i.e., it is true qua member of English and false qua member of German). Of course, no sentence can be both true and false. If I want to keep the claim that sentences are truth bearers, then I have to admit that natural language truth predicates are language-specific. It seems that I do not even grasp the rules of the game.

*Reply 7:* I have three replies to this objection. First, it is acceptable to claim that natural language truth predicates are unrestricted and choose sentences as primary truth bearers as long as one is careful in specifying what sentences are. It is common to distinguish sentence tokens from sentence types; *sentences tokens* are physical objects while *sentence types* are abstract entities.

Beyond this claim, there is little agreement on how to individuate sentence tokens, how to

---

<sup>41</sup> Of course, '(2'') is true-in-L' does not follow from '(2'') is false-in-L or gappy-in-L' and the fact that (2'') is '(2'') is false-in-L or gappy-in-L'; for that result, (2'') would have to be a sentence of L.

individuate sentence types, how to explain sentence types, or how to explain the relation between sentence tokens and sentence types. One option is to take sentence tokens to be pairs of possible physical objects and contexts, where the context is sufficient to determine the syntactic and semantic features of the sentence token. One can then take ‘true’ to express an unrestricted notion of truth and treat sentence tokens as primary truth bearers without running into trouble. Another option is to take sentence types to be individuated on the basis of semantic features. One can then take ‘true’ to express an unrestricted notion of truth and treat sentence types as primary truth bearers without difficulty. There are many other options as well.<sup>42</sup>

Second, I agree that almost everyone who presents a theory of truth that offers an approach to the liar paradox addresses only LS truth predicates and assumes that natural language truth predicates can be explained in terms of LS truth predicates. However, natural language truth predicates cannot be explained in terms of LS truth predicates; of course, there is no way I can provide a good justification for this claim here. I do want to point out several problems with it. One is that those who claim that natural language truth predicates can be explained in terms of LS truth predicates rarely say how this is to be done. Those who do usually either claim that natural language truth predicates are ambiguous and can take on the meaning of any of the LS truth predicates or that a truth predicate of a natural language L is synonymous with ‘translatable into a sentence of L that is true-in-L’. Neither of these approaches is plausible. First, they imply that some warranted assertions of blind truth attributions are unwarranted (e.g., Cletus says that some sentence is true without knowing the language to which it belongs).<sup>43</sup> Second, they cannot handle

---

<sup>42</sup> See Kaplan (1973, 1990), Hugly and Sayward (1981), Simons (1982), Bromberger (1989), Wetzel (1993), Horwich (1998: 98-103), Cappelen (1999), Dummett (1999), Szabó (1999), and Truncellito (2000) for discussion of the distinction between types and tokens. I prefer treating sentence tokens (i.e., pairs of possible objects and contexts) as primary truth bearers, so long as one individuates contexts finely enough. However, there are some tricky issues here having to do with the fact that the syntactic and semantic features of a truth bearer are not always sufficient to determine whether it is paradoxical and the fact that paradoxicality affects truth value (on some views); see Scharp (RB) for discussion.

<sup>43</sup> See Scharp (FTT) for a defense of this criticism.

multiple-target truth attributions (e.g., when Cletus says ‘everything Brandine uttered yesterday is true’ and Brandine spoke in several different languages on the day in question).<sup>44</sup> Third, the ambiguity approach faces the problem that ‘true’ fails all the standard ambiguity tests used by linguists, and the translation approach cannot handle truth attributions to some sentences of other languages that contain LS truth predicates (e.g., ‘‘Schnee ist weiss’ ist wahr-auf-Deutsch’ isn’t translatable into English unless English has ‘true-in-German’; thus, a translation predicate and ‘true-in-English’ aren’t sufficient).<sup>45</sup> Finally, they do not validate the truth rules (e.g., the ascending truth rule for an LS truth predicate is:  $\langle\langle p \rangle \text{ is true-in-}L\rangle$  follows from  $\langle p \rangle$  and  $\langle\langle p \rangle \text{ is in } L\rangle$ , but not from  $\langle p \rangle$  alone).

Third, all the arguments offered so far that depend on inter-linguistic truth attributions can be altered to accommodate the claim that natural language truth predicates can be explained in terms of LS truth predicates. For example, even if we assume that natural language truth predicates are LS truth predicates, one can still “import” a sentence that is potentially paradoxical for a theory of truth into revised natural language to which the theory applies—it just takes a bit more work. Assume that T is a version of Kripke’s theory of truth that is intended to explain truth-in-English\* and that T is restricted to English\*.<sup>46</sup> English\* is like English except that it does not contain any of the resources that can be used to construct revenge paradoxes for T. Let L be a language that does contain resources that could be used to construct revenge paradoxes for T (i.e., there are sentences of L that are potentially paradoxical for T—if T had applied to L, then these sentences would be paradoxical for T). Let  $\phi$  be one of these sentences. The direct way of “importing” a potentially paradoxical sentence into English\* no longer works. That is, ‘ $\phi$  is true-in-English\*’ is not paradoxical for T because it is false-in-English\*. However, consider another language, M such that:

<sup>44</sup> See Scharp (FTT) for a defense of this criticism.

<sup>45</sup> See Scharp (TI) for a defense of this criticism.

<sup>46</sup> This restriction is not trivial because languages other than English\* contain truth-in-English\* predicates (e.g., German has ‘wahr-auf-Englisch\*’). Assume that sentences of languages other than English\* are false-in-English\*.

(i) M has a sentence,  $\psi$  that is a translation of  $\phi$ , (ii)  $\psi$  has certain specifiable empirical features (e.g., it is the only sentence written on a certain blackboard), and (iii) M has a sentence  $\theta$  that attributes truth-in-M to  $\psi$  by appeal to its empirical features (e.g., ‘the sentence on the blackboard is true-in-M). Assume that: (i) English\* has a sentence  $\sigma$  that is a translation of  $\theta$ , and (ii) that English\* has a sentence  $\zeta$  that attributes truth-in-English\* to  $\sigma$ .<sup>47</sup> Although it might not seem like it, if T applies to  $\zeta$ , then  $\zeta$  is paradoxical for T. Here is the argument. If  $\phi$  is paradoxical for a theory of truth-in-L that applies to  $\phi$ , then  $\theta$  is paradoxical for a theory of truth-in-M that applies to  $\theta$ . If  $\theta$  is paradoxical for a theory of truth-in-M that applies to  $\theta$ , then  $\zeta$  is paradoxical for a theory of truth-in-English\* that applies to  $\zeta$ .<sup>48</sup> The argument relies on a combination of inter-linguistic truth attributions and empirically paradoxical sentences to show that it is impossible to eliminate from a natural language all the sentences that are potentially paradoxical for a given theory of truth (as long as the natural language has the features we take all natural languages to have).

*Objection 8:* Although the inconsistency problem is a genuine problem for a theory of truth, the self-refutation problem does not really pose a threat. In particular, it is acceptable for a theory of truth to imply that some of its consequences are gappy. Of course, the theory will end up being gappy as well, but that is acceptable too. In order for this sort of approach to be workable, one must accept that gappy sentences can be assertible, but that can be accommodated as well. Given that the other options are either a restricted (hence, unacceptable) theory or an inconsistent (hence, unacceptable) theory, a gappy theory doesn’t seem so bad.<sup>49</sup>

---

<sup>47</sup> All English\* needs in order to have a translation of  $\theta$  is a truth-in-M predicate and a way of specifying the empirical features of  $\psi$ , which it must have if it is to be a workable natural language. If English\* does not have a truth-in-M predicate, then all English\* truth attributions to sentences of M that express truth-in-M are false-in-English—an intolerable result for a descriptive theory.

<sup>48</sup> See Scharp (TI) for a more detailed version of this argument.

<sup>49</sup> See Maudlin (2004) for an example of a theory of this type.

*Reply 8:* There are several problems with this sort of theory. The most significant problem is that we are no longer able to use truth in our assessment of a theory. When one presents a theory, one wants one's theory to be correct—to get it right. That is, one wants one's theory to be true. Presumably, a theorist who offers the type of theory in question has the same desire. If asked, “is your theory correct?,” he would probably say, “yes.” However, given that the theory is gappy by its own lights, the theorist is unable to make the move from “I have it right” to “my theory is true.” We use truth as a measure of theoretical correctness and a theory of the type in question does not respect this aspect of our use of truth.

The second problem is that the approach to the liar paradox that involves accepting a theory of truth that implies that it is gappy depends on either restricting the theory so that it does not apply to certain languages or denying that certain linguistic resources are meaningful (i.e., an unintelligibility maneuver). Recall that a theory for which the self-refutation problem occurs implies that (1) (a liar sentence) is gappy, that (2) (a revenge liar sentence) is gappy, and that ‘(2) is true if and only if (2) is either false or gappy’ (a consequence of the theory) is gappy.

The problem occurs with certain sentential operators that force one into a two-valued scheme. Recall the sentence:

(5) (5) is Xnot true.

Just as in the case of a liar sentence, if we assume that (5) is true or that (5) is false, then we get a contradiction. If we assume that (5) is gappy, then we can derive that (5) is false. Thus, an approach of this sort cannot handle sentences like (5). The only options are to restrict the theory so that it does not apply to sentences like (5) or claim that the linguistic resources required to construct sentences like (5) (e.g., exclusion negation) are meaningless. As I have argued, neither option is remotely plausible.

### 3. INCONSISTENCY ARGUMENTS

In this section, I present three arguments for an approach to truth on which truth is an inconsistent concept. They are presented in the order of increasing specificity for their conclusions. The first argument is an abductive argument: an inconsistency theory of truth provides the best explanation of all the paradoxes associated with truth (including all the various revenge paradoxes). This argument supports any theory of truth that implies that truth is an inconsistent concept. The second argument is based on the revenge argument given in the last section. It supports most inconsistency theories of truth, but not all. The third argument appeals to the claim that a particular strategy for explaining our inconsistent concept of truth avoids all the revenge paradoxes and, hence, allows for a theory of truth that does not have to be restricted. This argument supports only inconsistency theories of truth that meet very specific conditions. In the following section, I present an overview of one such theory.

Before presenting those arguments, I want to discuss concepts and what I mean when I say that truth is an inconsistent concept. Roughly, I think of concept application on the lines of belief formation or assertion. For example, I apply the concept **scab**<sup>50</sup> to some object  $\alpha$  if I am prepared to assert ‘ $\alpha$  is a scab’ or I believe that  $\alpha$  is a scab. An important part of my account of concepts is that there are rules that govern the employment of a concept. A person who possesses a certain concept and is committed to employing it is committed to following the rules for the employment of that concept. One such rule is that the concept **scab** should be applied to scabs and it should not be applied to things that are not scabs.<sup>51</sup> One can think of these rules as *constitutive* in the sense that if a person utters a word that expresses the concept **scab**, then that person is committed to following the rules for the employment of **scab**. By ‘committed to following the rules’ I do not mean that the

---

<sup>50</sup> I use bold type as a convention for the names of concepts.

<sup>51</sup> I am not committed to explaining concepts in terms of such rules; see Davidson (1982) for a criticism of those who favor such an explanatory strategy.

person actually acts in accordance with these rules or that he explicitly endorses them; rather, I mean that the person *ought* to follow them—he is obligated to follow them—whether he explicitly endorses them or not (of course, it is not the case that everyone actually obeys the constitutive principles for a given concept—some disobey out of ignorance, others do so on purpose).<sup>52</sup>

An *inconsistent concept* is one whose constitutive rules are incompatible in the sense that they dictate that the concept both applies and disapplies to some entities.<sup>53</sup> The rules for the employment of an inconsistent concept impose conflicting commitments on the employers of that concept. Thus, the employer of an inconsistent concept cannot follow the rules for the application of that concept in all circumstances.<sup>54</sup> Consider an example:

(7a) ‘rable’ applies to x if x is a table.

(7b) ‘rable’ disapplies to x if x is red.<sup>55</sup>

**Rable** is an inconsistent concept. Someone who possesses **rable** might run into difficulty employing it because it both applies and disapplies to red tables. When confronted with a red table, an employer of **rable** will be unable to satisfy the demands it places on her. It might seem that someone could employ **rable** without trouble as long as she avoids red tables. However, even if an employer of **rable** never encounters a red table, the concept still poses a problem for her because inconsistent concepts pose a normative problem for their employers. Someone who chooses to employ **rable** *should* apply it to tables and *should* disapply it to red things. These are conceptual

---

<sup>52</sup> My commitment to constitutive rules for concepts places me in the tradition of meaning-constitutive accounts of concepts. However, not all the members of this tradition agree on rules as the relevant constitutive element. Some constitutive accounts choose the possession of propositional attitudes, the truth of theories, the validity of implications, etc. See Peacocke (1992) for an example; however, Peacocke treats a concept’s constitutive principles as those possessors of the concept are committed to following, whereas I treat them as those employers of a concept are committed to following. One can possess a concept without employing it; see below for the role this claim plays in my account.

<sup>53</sup> I have yet to find a good antonym for ‘applies’.

<sup>54</sup> See Chihara (1979, 1984), Yablo (1993a), and Eklund (2002) for similar views on inconsistent concepts.

<sup>55</sup> I say that a concept *applies* to the members of its extension and *disapplies* to the members of its anti-extension.

norms to which the employer has decided to bind herself. Thus, an employer of **rable** has committed herself to obeying incompatible rules even if she never encounters a red table.

Inconsistency theories of truth imply that truth is an inconsistent concept. That is, they imply that truth has certain constitutive rules that govern its employment and these rules are incompatible in the sense that they dictate that truth both applies and disapplies to certain items. I am not concerned with arguing about which rules are constitutive for truth, but it seems to me that the truth rules are constitutive of it along with the principle of mono-aletheism (i.e., no truth bearer is both true and false simultaneously). Liar sentences are among the items to which truth both applies and disapplies by virtue of the fact that the rules governing its employment are incompatible.

These sketchy remarks about truth as an inconsistent concept are bound to raise more questions than they answer, but they will have to do for now. A complete theory of inconsistent concepts including a logic, a semantic theory, and a pragmatic theory is beyond the scope of this paper. Likewise, the details of how such a theory should be applied to the case of truth will have to wait for another occasion. (However, I briefly mention some of my views on both matters in section four).

The first argument is that if one decides to treat truth as an inconsistent concept, then one has available a satisfying explanation of the current situation in truth studies. That is, one can explain why other theories of truth face revenge paradoxes, both inconsistency problems and self-refutation problems. No other theory of truth has managed to do this.<sup>56</sup>

The explanation for why theories of truth that imply truth is a consistent concept face revenge paradoxes or self-refutation problems is straightforward. Our concept of truth is inconsistent in the sense that its constitutive principles, the truth rules, are incompatible. That is,

---

<sup>56</sup> See Field (Forthcoming) and Glanzberg (2005) for the only alternative explanations of which I am aware; I consider them in a reply to an objection below.

there are objects that these rules classify as both true and not true. We can call the set of such objects the *overdetermination set* for truth. All the paradoxical sentences considered so far are members of the overdetermination set for truth.<sup>57</sup> Any theory of truth that implies that truth is a consistent concept and that includes these principles is inconsistent and can be rendered consistent only by restricting it. If a theory of truth implies that some of the sentences in the overdetermined set for truth are gaps, then its fate depends on which of these sentences it classifies as gaps. Recall that many of the members of the overdetermination set for truth are truth attributions, and no matter what truth status (e.g., true, false, gappy, etc.) one assigns them, they are consequences of the assignment. No matter whether one's theory of truth classifies these paradoxical sentences as true, false, or gappy, some of these paradoxical sentences are consequences of the theory. Thus, if a theory of truth implies that all the sentences in the overdetermined set for truth are gaps, then the theory implies that some of its consequences are gaps. On the other hand, if a theory of truth does not classify some of these sentences as gaps, then the truth rules imply that they are both true and not true. On the first option, the theory is self-refuting, while on the second, it faces an inconsistency problem. Therefore, both types of revenge paradoxes can be explained if we assume that truth is an inconsistent concept.

In section two, I argued that theories of truth that validate the truth rules face revenge paradoxes. If we admit that truth is an inconsistent concept, then we can explain why this occurs. Therefore, by accepting that truth is an inconsistent concept, we arrive at a deeper explanation for why theories of truth that validate the truth rules fail are unacceptable.

The following is a summary of the first argument for treating truth as an inconsistent concept:

---

<sup>57</sup> It seems to me that truth-tellers (e.g., sentence  $\tau$ , ' $\tau$  is true', is a truth-teller) are in the *underdetermination* set for truth, but these sentences are not paradoxical and none of my claims or arguments hang on this opinion.

- (i) If we assume that truth is an inconsistent concept, then we can explain the presence of the liar paradox and the presence of revenge paradoxes.
  - (ii) The inconsistency explanation of the liar paradox and the revenge paradoxes is better than any of the others.
- ∴ (iii) Probably, truth is an inconsistent concept.

Only by admitting that truth is an inconsistent concept can we satisfactorily explain the most significant feature of our long battle with the liar paradox.

The second argument for the claim that truth is an inconsistent concept depends on the revenge argument from the previous section: if a theory of truth implies that the truth rules are valid, then it is inconsistent, self-refuting, or restricted. Given that any theory of truth that is inconsistent, self-refuting, or restricted is unacceptable, any theory of truth that implies that the truth rules are valid is unacceptable. That is the conclusion of the revenge argument. I have assumed that we should accept that the truth rules are constitutive of truth—that they govern the way truth should be employed. Again, this claim counts as an assumption in this paper because I do not have the space to argue for it here; however, there are good reasons to accept it. Thus, on my view: (i) if a theory of truth is acceptable, then it implies that the truth rules are constitutive for truth, and (ii) if a theory of truth validates the truth rules, then it is unacceptable. There seems to be very little wiggle room here. However, there is an additional assumption that connects the two conditionals: if a theory of truth implies that the truth rules are constitutive of truth, then it implies that they are valid. I reject this claim. I accept that if a concept is not defective (i.e., its constitutive principles are compatible with one another and empirical facts), then its constitutive principles are valid (or true). However, for inconsistent concepts, it is not so straightforward. One of the key aspects of the account of inconsistent concepts I endorse is that one or more of the constitutive rules governing the

concept are invalid.<sup>58</sup> Recall that the constitutive principles for a concept are those that one commits oneself to following if one is committed to employing the concept. However, inconsistent concepts should not be employed. Of course, they are still possessed, but there are plenty of concepts we possess even though we do not employ them. Only if one is committed to employing a concept does one assume that its constitutive inference rules are valid. Thus, I accept that the truth rules are constitutive of truth, but it is not the case that they are all valid. This result allows anyone with a suitable account of inconsistent concepts to avoid the revenge argument. The following is a summary of the second argument for treating truth as an inconsistent concept:

- (i) If a theory of truth is acceptable, then it implies that truth obeys the truth rules.
- (ii) If a theory of truth implies that truth is a consistent concept and it implies that truth obeys the truth rules, then it implies that the truth rules are valid.
- (iii) If a theory of truth implies that the truth rules are valid, then it is inconsistent, self-refuting, or restricted.
- (iv) If a theory of truth is inconsistent, self-refuting or restricted, then it is unacceptable.
- ∴ (v) If a theory of truth is acceptable, then it does not imply that truth is a consistent concept.

Only an inconsistency theory of truth on which inference rules that are constitutive for a concept need not be valid has a chance of being an acceptable theory of truth.

The third argument justifies only a small number of inconsistency theories of truth. In short, the argument is: if one assumes that truth is an inconsistent concept, then one can construct a theory of truth that does not face any revenge paradoxes and so does not have to be restricted in any way. Given that there are no other theories of truth that are really revenge-free, the fact that an inconsistency theory of truth can accomplish this feat is a strong reason to accept it. If I am right

---

<sup>58</sup> This claim is a result of the fact that I explain inconsistent concepts in terms of confused concepts. Thus, I claim that one should explain an inconsistent concept by appeal to a set of component concepts. Each component concept obeys only some of the constitutive principles for the inconsistent concept in question. An inference rule for the inconsistent concept is valid if and only if it is valid for each of the component concepts.

that theories of truth that are restricted to avoid revenge paradoxes are unacceptable and that no other purportedly acceptable theories of truth avoid the revenge paradoxes, then the theory I endorse is the only acceptable theory of truth to have ever been proposed. It should not come as a shock that I cannot provide a convincing argument for the claim that the theory I endorse is genuinely revenge-free—for I do not even have the space to present the theory in detail. Moreover, I am unable to fully justify my claim that all the other purportedly acceptable theories of truth do have to be restricted in some way. Nevertheless, the discussion in the preceding section should motivate (to some degree) the claim about the other theories, while the discussion in the following section should motivate (to some degree) the claim about my own theory.

The following is a summary of the third argument for treating truth as an inconsistent concept:

- (i) If a theory of truth is acceptable, then it is not restricted.
  - (ii) A particular inconsistency theory of truth *T* does not face revenge paradoxes and so is not restricted.
  - (iii) Of the proposed theories of truth, those other than *T* are either antecedently implausible or restricted.
- ∴ (iv) Of the proposed theories of truth, only *T* is acceptable.

Only the theory of truth I propose (or one that is relevantly similar to it) is acceptable.

*Objection 1:* There are no inconsistent concepts. Any attempt to introduce a term that behaves according to incompatible rules fails to introduce a meaningful term at all. Thus, it is impossible that a term obeys incompatible rules of employment. One reason for thinking this is that interpretation requires one to use the logic one endorses when interpreting another. Thus, it is inappropriate to ever attribute an inconsistent concept to someone, since the interpreter would have to attribute something that defies the logic she endorses.<sup>59</sup> Moreover, even if one could introduce a

---

<sup>59</sup> One can find a similar objection in Stebins (1992).

term that obeys incompatible rules, it would be overdetermined for every item, so it would be unemployable.<sup>60</sup>

*Reply 1:* First, the claim that we interpret others as if they endorse our logical standards is simply false. If it were true then there would be no distinction between criticizing someone for failing to follow an inference rule she endorses and criticizing someone for endorsing the wrong inference rule. It is obvious that there is such a distinction and it plays an important role in philosophical discussions. Second, charity can cut both ways. One might simply introduce an inconsistent concept, begin using it, and describe it as inconsistent (I did this with the concept **rable**). It seems to me that it would be quite difficult to go on interpreting someone who does this as if they had misunderstood their own stipulative definition and their claims about it. Indeed, one might give an account of all the relevant factors in charitable interpretation and present two situations, one in which the weighted sum of all the factors is higher than that of the second, while in the first one attributes an inconsistent concept, but in the second one does not. The point here is that attributing an inconsistent concept is sometimes the most charitable thing to do. No matter what constraints one imposes on charitable interpretation (except of course, a conceptual consistency constraint), there will be situations in which it is more charitable to attribute an inconsistent concept.

I agree that a major problem for a theory of inconsistent concepts is showing that a concept can be both inconsistent and employable (i.e., not overdetermined for every item). The theory I endorse accomplishes this, in part, by employing a relevance logic to evaluate arguments with

---

<sup>60</sup> See Gupta and Belnap (1993: 13-15) for this objection; see also Chihara (1984) for discussion.

sentences containing truth predicates; however, a complete account is beyond the scope of this paper.<sup>61</sup>

*Objection 2:* The account of inconsistent concepts depends on the claim that each concept has some constitutive principles. However, there are good reasons to reject this claim, which include: (i) it commits one to analytic truths,<sup>62</sup> and (ii) it implies that there is a well-defined distinction between changes in meaning and changes in belief.<sup>63</sup>

*Reply 2:* I agree that many meaning-constitutive accounts of concepts face those problems, but mine does not. I do not claim that the meaning-constitutive principles concern concept *possession*; rather, they concern concept *employment*. One can possess a concept without being committed to employing it. Indeed, I suggest that although we possess the concept of truth, we should not employ it. On the account I offer, simply possessing a concept does not commit one to its meaning-constitutive principles at all; only if one is committed to employing a concept is one committed to its meaning-constitutive principles. Without this distinction, it seems impossible to give an adequate account of inconsistent concepts.

Although other views on meaning-constitutive principles do imply that these principles are analytic (i.e., true or valid by virtue of their meaning alone), the one I offer does not. Indeed, it is my view that *any* sentence that expresses an inconsistent concept has no truth value (because I think that truth itself is inconsistent, when I use ‘truth value’ I am appealing to the concepts of truth I offer as replacements for our inconsistent concept).<sup>64</sup> Thus, no meaning-constitutive principles are

---

<sup>61</sup> Note that one needs a two part solution to the problem of overdetermination—one needs some sort of paraconsistent logic on which it is not the case that everything follows from a contradiction, and one needs a logic on which it is not the case that all the constitutive rules for the concept in question are valid.

<sup>62</sup> See Fodor and Lepore (1994).

<sup>63</sup> See Davidson (1974).

<sup>64</sup> Otherwise, one would not be able to distinguish between concept possession and concept employment. For example, if some sentence *p* that expresses an inconsistent concept has a truth value, then one should be able to assert that it has that truth value. If one can assert that *p* has a certain truth value, then (because of the truth rules) one is committed to either *p* or its negation. Thus, if *p* has a truth value, then it is impossible to avoid employing the inconsistent concept it expresses.

analytic and some are not even true or valid.<sup>65</sup> Moreover, one can endorse a meaning-constitutive account of concepts and admit that there is no principled way of distinguishing between meaning change and belief change. The reason usually given for this criticism is that if one's word means X, then one has to believe the associated meaning-constitutive principle. However, the account I offer is externalist in a certain sense; if one employs a certain concept, then one is committed to its constitutive principles, but one need not have the associated beliefs. That is, we can imagine a situation in which a person employs the concept *elm* and the concept *beech* and she is committed to the respective constitutive principles governing those concepts, but she doesn't have any beliefs that distinguish elms from beeches. Indeed, one can employ a certain concept even though one denies its constitutive principles (e.g., theorists who offer approaches to the liar paradox on which the truth rules are not constitutive).

*Objection 3:* There are other inconsistency theories of truth. Why aren't they at least as good as the one I offer?

*Reply 3:* There are several inconsistency theories of truth on the market and I cannot do justice to all of them here; I confine my comments to pointing out differences between my theory and the rest and explaining why I prefer my account. None of what follows should be treated as real criticism. Yablo has a theory of inconsistent concepts that is based on an account of circularly defined concepts.<sup>66</sup> Eklund has a theory of inconsistent concepts as well, but he focuses on entire languages (I take it that an inconsistent language is just a language that has a term that expresses an inconsistent concept). He suggests that by virtue of our semantic competence, we accept the constitutive principles of an inconsistent concept and he offers a semantics for inconsistent languages on which an acceptable assignment of semantic values to expressions of a language, L, is

---

<sup>65</sup> One might worry that if a concept is not defective, then its meaning-constitutive principles are analytic. On the contrary, in the case of a non-defective concept, a meaning-constitutive principle is true or valid, but it is true or valid by virtue of both its meaning and the fact that the concept is not defective (which depends on empirical facts).

<sup>66</sup> Yablo (1993a, 1993b).

one that makes true a weighted majority of the constitutive principles for L.<sup>67</sup> The biggest problem I have with the theories Yablo and Eklund propose is that they require that we continue to employ our inconsistent concept of truth. That is, they imply that it is acceptable to employ an inconsistent concept even after one has discovered that it is inconsistent. Indeed, both Yablo's theory and Eklund's theory appeal to the concept of truth. Because their theories appeal to the inconsistent concept of truth, anyone who accepts either theory has to employ an inconsistent concept. I think that is unacceptable. It is my view that inconsistent concepts should be replaced with consistent ones; they aren't fit for employment. The rationale for this view is simple: other things considered, one should avoid undertaking incompatible commitments. Thus, if one discovers that a concept is inconsistent, one should stop employing it if one can.

In addition, the fact that these theories appeal to the inconsistent concept of truth causes problems for them that are similar to the revenge paradoxes. For example, if a philosopher accepts Eklund's theory of truth, then she also commits herself to using truth according to principles that are incompatible. Thus, she commits herself to both applying and disapplying truth to certain sentences. That is similar to the inconsistency problem. Moreover, Eklund's theory still has to classify paradoxical sentences and some of these will be consequences of the theory. If it classifies them as false, then his theory has false consequences. If it classifies them as gaps, his theory has gappy consequences. Either way, his theory is self-refuting. Similar remarks hold for Yablo's theory.

Another theory that might seem like an option is dialetheism—the view that some contradictions (e.g., '(1) is true and (1) is false') are true. Given what I have said about inconsistency theories of truth, dialetheism need not count as one. That is, the dialetheist can just deny that mono-aletheism (i.e., no truth bearer is true and false simultaneously) is a constitutive

---

<sup>67</sup> Eklund (2002, Forthcoming). See also Azzouni (2003) and Patterson (2006) for similar theories; I can't give them all the space they deserve.

principle for truth. According to this version of dialetheism, the rules governing our concept of truth are compatible (as far as I understand it, all contemporary dialetheists fall into this category); they just imply that truth and falsity overlap in some cases.<sup>68</sup> However, one might use dialetheism as a part of a theory of inconsistent concepts. On this version of dialetheism, mono-aletheism is a constitutive principle for truth; hence, the constitutive principles for truth are incompatible. As a theory of inconsistent concepts, dialetheism implies that some sentences expressing inconsistent concepts are both true and false (e.g., ‘the red table is rable’). It is unclear to me whether the two versions of dialetheism (i.e., truth is a consistent concept that doesn’t obey mono-aletheism vs. truth is an inconsistent concept that does obey mono-aletheism) are different in a substantive way. Instead of worrying about this issue, let me say that I reject the first version of dialetheism because it does not respect one of the constitutive principles for truth (as I said in section two, this isn’t an objection by itself, but it does point the way toward one). Although the second version of dialetheism (as a theory of inconsistent concepts), avoids this problem, it still requires one to accept that some truth-bearers are both true and false (which seems like a consequence we should avoid if we can). Moreover, this version of dialetheism has the same problem as other theories of inconsistent concepts: because it appeals to the inconsistent concept of truth, anyone who accepts the theory has to employ an inconsistent concept. It also cannot deal with certain non-monotonic sentential operators, like exclusion negation.

*Objection 4:* There are at least two well-developed explanations of the revenge paradox phenomenon, one from Hartry Field and one from Michael Glanzberg. Why is the explanation I offer superior to the ones they offer?

*Reply 4:* Let me begin the reply by saying that Glanzberg and Field each have interesting, complex, and subtle approaches to the liar paradox and there is no way I can do justice to either of

---

<sup>68</sup> See Priest (1987, 2006), and the papers in Priest, Beall, and Armour-Garb (2004).

them in this reply. One should not take what I write here to be something like definitive criticisms; instead, I aim to point out relevant differences between our respective explanations.

I begin with Glanzberg, who offers a context-dependence approach to the liar. However, instead of claiming that truth predicates are explicitly context dependent, Glanzberg argues that sentences that contain truth predicates display an implicit context dependence that is due to the presence of quantification. Glanzberg offers a theory of background domains of propositions for the quantifiers involved, which includes an infinite hierarchy of domains and no “biggest” domain.<sup>69</sup>

If one accepts Glanzberg’s theory, then one has to admit that there is no unrestricted quantification. Indeed, one has to accept that we can express the notion of truth-in-a-context and we can even quantify over contexts to a limited degree, but we cannot express an unrestricted notion of truth. “One way or another, hierarchical theories all require that speakers cannot in any one instance express the entirety of a unified concept of truth,” (Glanzberg 2004: 289). He argues that the sort of fragmentation we see in our concept of truth is familiar to us (i.e., it occurs in the concept of mathematical proof as well) and that it occurs because truth fails to be closed under reflection.

Glanzberg’s defense of this feature is based on the idea that any characterization of truth permits one to reflect on the truth of the characterization, and this reflection both shows that the initial characterization is inadequate and points the way toward a stronger one. This process of reflection is unending; hence the infinite hierarchy of contexts.<sup>70</sup> The motivation for this view comes from what has been called the strong liar reasoning. Consider the following sentence:

---

<sup>69</sup> Glanzberg (2001, 2004, 2005).

<sup>70</sup> This view about the relation between reflection and revenge seems to stem from some of Kripke’s remarks: “Such semantical notions as ‘grounded,’ ‘paradoxical,’ etc. belong to the metalanguage. This situation seems to me to be intuitively acceptable; in contrast to the notion of truth, none of these notions is to be found in natural language in its pristine purity, before philosophers reflect on its semantics (in particular, the semantic paradoxes). If we give up the goal of a universal language, models of the type presented in this paper are plausible as models of natural language at a stage before we reflect on the generation process associated with the concept of truth, the stage which continues in the daily life of nonphilosophical speakers” (Kripke 1975: 714).

(8) (8) is not true.

The partiality approach to the liar implies that (8) is gappy. We know that if a sentence is gappy, then it is not true. Thus, the partiality approach implies that (8) is not true. Hence, the partiality approach implies that ‘(8) is not true’ is true; therefore, it implies that (8) is true.<sup>71</sup> It is by reflection on the way the approach classifies (8) that drives us to conclude that (8) is true after all. The claim, (if the partiality approach implies that  $\langle p \rangle$  is true, then  $\langle p \rangle$  is true), is similar to what is called a reflection principle. It states something about a formal theory that cannot be captured by the formal theory on pain of contradiction.<sup>72</sup>

Glanzberg argues that one can begin with a basic formal theory of truth, formulate a reflection principle for that theory, which illustrates the theory’s inadequacy, and arrive at a new formalization of the theory that effectively incorporates the reflection principle. We can continue this process to arrive at a transfinite hierarchy of formal theories of truth, which is analogous to the hierarchy of contexts for truth attributions. He claims that truth is a *Kreiselian concept* in this sense: any formal theory of truth points the way to a stronger formal theory, and the process of theory construction is unending.<sup>73</sup>

Glanzberg’s point is that what seem to be revenge paradoxes are really just the effects of the Kreiselian aspect of truth. A theory of truth should not be expected to treat as true the claim that its consequences are true. Nor should a theory of truth be found lacking if the result of conjoining a reflection principle to it results in an inconsistent theory. These phenomena are just consequences of the fact that truth is a Kreiselian concept.

There are several places at which I disagree with Glanzberg’s analysis. The first is that I do not find the strong liar reasoning compelling. Because strong truth does not have truth value gaps,

---

<sup>71</sup> See Burge (1979) for discussion of the strong liar reasoning.

<sup>72</sup> See Feferman (1991) for an overview of reflection principles.

<sup>73</sup> Glanzberg (2005).

and the strong liar reasoning involves a move from  $\langle\langle p \rangle \text{ is gappy} \rangle$  to  $\langle\langle p \rangle \text{ is not true} \rangle$ , (8) should be read as: (8) is Xnot weak true. We already know that partiality approaches have troubling handling sentences like this, but the trouble has nothing to do with reflection on how the theory of truth in question classifies (8). In fact, most partiality approaches to the liar are based on fixed-point constructions and so have no consequences for sentences like (8) at all. Thus, there is no reason to think that one derives a contradiction only by assuming that the theory implies that (8) is gappy. Therefore, the view that reflection on the dictates of a theory of truth has any special role to play in reasoning about the status of paradoxical sentences seems to be a mistake.

A second issue is that there seems to be no reason to think that the concepts described by each of the formal theories Glanzberg identifies have anything in common. In particular, I see no reason to think that they are all “more or less” theories of truth. However, Glanzberg claims that each of the formal theories provides a rough characterization of the unified concept of truth. The problem with this view is that on Glanzberg’s own account, it is impossible to express the unrestricted notion of truth each of these theories is supposed to describe. Thus, we have this concept of truth, but we can never actually use it. That sounds fairly counterintuitive. Moreover, if Glanzberg is right, then it is impossible to arrive at a theory of truth that correctly and completely describes our concept of truth. The best we can achieve is stronger and stronger theories that are always lacking.

On the other hand, Field presents a sophisticated version of the partiality approach that is designed to handle not just liar sentences but certain revenge sentences as well. Field shows how to introduce a conditional into a three-valued scheme that obeys most of the intuitive principles we take to govern conditionals and acts just like a material conditional in classical contexts (i.e., those for which excluded middle is assumed). It is hard to overestimate the significance of this accomplishment; for the first time, we have a three-valued logic with a real conditional—a three-

valued logic in which we can reason. Field then defines a determinacy operator in terms of the conditional. He claims that liar sentences are indeterminate (not determinately true and not determinately false). Given what I have said already about revenge paradoxes, it should be obvious that a sentence like the following poses a problem for an approach like Field's:

(6) (6) is either false or indeterminate.

Field's determinacy operator iterates non-trivially, so he can say that (6) is not determinately determinately true and not determinately false (determinate determinate truth is weaker than determinate truth, but not so for determinate falsity). We can say that a liar sentence is indeterminate<sub>0</sub> (i.e., it is not determinately true and not determinately false) and that (6) is indeterminate<sub>1</sub> (i.e., it is not determinately determinately true and not determinately false). Field shows how to construct a hierarchy of determinacy operators (by iterations of the primitive one defined in terms of the conditional), which are used to define a hierarchy of determinate truth predicates and a hierarchy of indeterminateness predicates. As a final touch, he proves that his theory of truth is expressible in the artificial language he constructs (which has the expressive resources to handle set theory) without giving rise to revenge paradoxes. Thus, his theory does not appeal to a distinction between object language and metalanguage; it can be applied to the very language in which it is formulated. Most impressive.<sup>74</sup>

In section two on the revenge argument, I argued that partiality approaches to the liar paradox face revenge paradoxes. How does Field avoid that argument? He rejects the excluded middle for all of the indeterminateness predicates he defines. That is, he rejects '(6) is indeterminate<sub>0</sub> or (6) is not indeterminate<sub>0</sub>'. Thus, all of his indeterminacy predicates are analogous to truth (on his view) in that they are partially defined. Of course, as one goes further and further up the hierarchy of indeterminacy predicates, one gets closer and closer to a completely defined

---

<sup>74</sup> See Field (2003a, 2003b, 2005a, 2005b, Forthcoming).

indeterminateness predicate, but Field argues convincingly that one never gets there. That is, it is impossible to define a completely defined indeterminacy predicate in the language he constructs.

Thus, he never has to deal with a real revenge paradox, like:

(6\*) (6\*) is either false or indeterminate\*,

where ‘indeterminante\*’ is completely defined (i.e., every sentence is either indeterminate\* or not indeterminate\*). Field’s theory clearly cannot handle sentences like (6\*). What does he say about them?

Field argues (convincingly in my view) that one need not use any such linguistic expression to formulate his theory (his semantic theory uses a completely defined notion of semantic value, but it is relative to a model and so cannot be used to construct a revenge paradox). In addition, he argues against the claim that the artificial language he constructs avoids revenge paradoxes only because it has expressive limitations. Instead, he claims that sentences like (6\*) are unintelligible. Indeed, Field argues, any linguistic expression that is not in his artificial language and seems to give rise to a revenge paradox (e.g., ‘indeterminate\*’) is unintelligible; however, by ‘unintelligible’ he does not mean *meaningless*:

I don’t want to deny that we have these notions; but not every notion we have is ultimately intelligible when examined closely. A large part of the response to the counterintuitiveness qualm will be an argument, in Part Four, that the notion of “the” hierarchy of iterations of D has a kind of inherent vagueness that casts doubt on there being a well-behaved notion of “ $D\alpha$ -true for every  $\alpha$ ”; and without that there is no reason to suppose that there is a well-behaved notion of “determinately true in every reasonable sense of that term”. The apparent clarity of such notions is an illusion. (Field Forthcoming c: §11).

Field then argues that there is no way to extrapolate from the hierarchy of determinateness predicates to define a well-behaved (i.e., intelligible) notion of hyper-determinateness.<sup>75</sup>

I am willing to admit that if we have only the resources provided by Field’s artificial language, then we will be unable to define a well-behaved notion of hyper-determinateness.

---

<sup>75</sup> Field also discusses what he calls “model-theoretic revenge,” but it is distinct from the sort of worry that I’ve pressed in this paper; see Field (Forthcoming: §9).

However, it seems to me that this point does little to quell the revenge worries. The problem is *not*: how can we use the resources Field gives us to generate a paradox his theory cannot handle? The problem is: we have a notion of determinateness that obeys excluded middle (i.e., what Field calls hyper-determinateness) and one cannot express this notion in Field's artificial language. Thus, Field avoids revenge only by an expressive limitation on his language.

I assume that it is obvious how Field would respond. He would probably claim that his artificial language can express any *intelligible* notion we have. Furthermore, he might continue, the problem of revenge paradoxes that I keep pressing is a problem that arises only when truth, which is intelligible, is combined with other resources (e.g., exclusion negation, other non-monotonic sentential operators, hyper-determinateness operators, etc.), which are not intelligible. Thus, it is not that truth is responsible for these revenge paradoxes; rather, truth has been keeping company with some troublemakers who are responsible. I address this reply in the next objection.

Before doing so, a summary is in order. I have argued that there are two kinds of revenge paradoxes: self-refutation problems and inconsistency problems. Glanzberg addresses self-refutation problems, which confront theories of truth that imply that they are  $X$ not true. He argues that this kind of revenge paradox has its source in the fact that truth is a Kreiselian concept (i.e., it is not closed under reflection). However, Glanzberg does not explain or even address inconsistency problems, and there are good reasons to doubt his explanation of the self-refutation problem. On the other hand, Field addresses inconsistency problems and argues that these sorts of revenge paradoxes arise when what are ultimately unintelligible—read that as inconsistent or not well-defined—concepts (e.g., hyper-determinateness) are combined with truth. However, Field does not explain or even address self-refutation problems, and there are good reasons to doubt his explanation of the inconsistency problem. In contrast to both Glanzberg and Field, I offer an

explanation of both types of revenge paradoxes, and my explanation of each type is superior to the one offered by Glanzberg and by Field, respectively.

*Objection 5:* As Field suggests, we should assume that the other items involved in the revenge paradoxes are defective instead of assuming that truth is defective. Given the importance and centrality of truth and the relative unimportance of these other items, we should prefer a theory on which truth is an acceptable concept and the others are defective.

*Reply 5:* There are several issues to consider when deciding which linguistic items should be blamed for the paradox. One issue involves the sort of explanation we get. The objector suggests that we should blame exclusion negation, and blame all the other non-monotonic sentential operators, and blame the conditional, and blame completely defined gaphood predicates, and blame idempotent determinacy operators, and blame quantification over partially defined gaphood predicates, and blame paradoxicality predicates, and blame groundedness predicates, and blame truth expressions that are not language-specific, ... the list goes on. I suggest that we should blame truth. That's it. Thus, my explanation is much simpler. It is also much more plausible. We can construct artificial languages that contain the outlaw linguistic expressions and they are perfectly well-behaved as long as they do not contain truth predicates (or related semantic terms). Of course, we can also construct artificial languages with truth predicates that are perfectly well behaved as long as they do not contain the outlaw linguistic expressions. However, the difference is that there are many different ways to construct revenge paradoxes; one involves truth and exclusion negation, one involves truth and another non-monotonic sentential operator, one involves truth and the conditional, one involves truth and an idempotent determinacy operator, etc. Exclusion negation is not involved in each case, nor are any of the other outlaw linguistic expressions. However, truth is involved every time. Truth is the only suspect that has no alibi—it is present at every crime scene;

none of the others is. It does not take a Holmes, or a Spade, or a Columbo to identify the perpetrator; even a Wiggum could get this one right.

However, there is another, even more compelling reason to prefer my approach. As I said in reply to objections six and seven of section two, as long as one has a truth predicate that obeys the truth rules (even a language-specific one) and some minimal resources (e.g., common descriptions), one can “import” a revenge paradox into the language by way of inter-linguistic truth attributions. Field avoids this problem only because he does not consider other languages at all. Thus, even if we follow Field’s advice (i.e., blame the outlaw linguistic resources for the liar and revenge paradoxes and revise our language so that it does not contain any outlaw linguistic resources), then we still have to restrict the theory so that it does not apply to certain sentences of the revised language. Thus, Field’s strategy does not really secure a language and a theory such that the theory both applies to and is expressible in the language. Hence, the “blame everything but truth” strategy does not work. Not only is my strategy simpler and more plausible, it is the only one that works.

#### 4. THE WAY FORWARD

Inconsistent concepts are rarely a topic of contemporary philosophical discussions, and there are, to my knowledge, no detailed theories of inconsistent concepts.<sup>76</sup> Given that the central claim of the account of truth I offer is that truth is an inconsistent concept, I am badly in need of an adequate theory of inconsistent concepts. My strategy is to first construct such a theory and then apply it to the case of truth.

A central claim of the account of inconsistent concepts I offer is that inconsistent concepts are confused. A concept is confused if and only if anyone who employs it fails to properly

---

<sup>76</sup> Perhaps Yablo (1993b) and Eklund (2002) should count, but there are good reasons to be dissatisfied with them (expressed in the previous section); see also Gupta (1999) on misconceptions; see also Field (1973, 1974), Priest (1995), Allen (2001), Camp (2002), and Scharp (2005b) on confused concepts; see also Dummett (1973: 454-5) and Williamson (2003, Forthcoming) on defective concepts.

distinguish between two or more entities. A classic example is the concept of mass as it was employed in Newtonian mechanics. In Newtonian mechanics, physical objects have a single physical quantity: mass. According to this theory, the concept of mass obeys the two laws (which are considered equally fundamental): (i)  $\text{mass} = \text{momentum} / \text{velocity}$ , and (ii) the mass of an object is the same in all reference frames. However, in relativistic mechanics, physical objects have two different “kinds” of mass: proper mass and relativistic mass. An object’s proper mass is its total energy divided by the square of the speed of light; an object’s *relativistic mass* is its non-kinetic energy divided by the square of the speed of light. Although  $\text{relativistic mass} = \text{momentum} / \text{velocity}$ , the relativistic mass of an object is not the same in all reference frames. Contrariwise,  $\text{proper mass} \neq \text{momentum} / \text{velocity}$ , but the proper mass of an object is the same in all reference frames. Thus, relativistic mass obeys one of the laws for mass, and proper mass obeys the other. A person who employs the concept of mass thinks that there is one thing (mass), when there are really two (relativistic mass and proper mass); an employer of the concept of mass is committed to the two laws given above, which are incompatible.<sup>77</sup>

If inconsistent concepts are confused, then for every inconsistent concept, there are two or more *component concepts* that are not being distinguished properly. In the mass example, relativistic mass and proper mass are the component concepts of the inconsistent concept of mass. If truth is an inconsistent concept, and all inconsistent concepts are confused, then an important question is: what are the component concepts of truth? I give my answer to this question below; for now it is enough to know that I propose two component concepts for truth.

A second central claim of the theory of inconsistent concepts I offer is that inconsistent concepts should not be employed. There is an important distinction between concept employment and concept possession. I possess many concepts that I do not employ (e.g., the concept of mass).

---

<sup>77</sup> This example is prominent in Field (1973, 1974). See Camp (2002) and Scharp (2005b) for discussion.

If one employs an inconsistent concept, then one is committed to its constitutive principles, which are incompatible. Because one should avoid undertaking incompatible commitments, one should not employ an inconsistent concept. Instead of employing an inconsistent concept, one should employ its component concepts. I call this the *replacement policy* for handling inconsistent concepts. Because I advocate the replacement policy, I often use the term ‘replacement concept’ for each of the component concepts of an inconsistent concept.

So the two central claims of the theory of inconsistent concepts are: (i) inconsistent concepts are confused, and (ii) one should replace an inconsistent concept with its component concepts in the repertoire of concepts one employs. We can get a general picture of how this account applies to truth. First, truth is a confused concept, which implies that there are components of truth. Second, if there are two component concepts for truth, then we need three theories of truth: one for each component concept and one for our inconsistent concept of truth. Third, we should no longer employ our inconsistent concept of truth. Instead, we should employ the component concepts. Fourth, it follows that the three theories of truth should *not* appeal to our inconsistent concept of truth; indeed, no theory should appeal to our inconsistent concept of truth. Instead, the theory of our inconsistent concept of truth should appeal to the component concepts of truth and the theory for each component should be explained by appeal to whatever other concepts are appropriate (as long as they aren’t, in turn, explained in terms of truth).

The theory of our inconsistent concept of truth should explain how we should interpret discourse that expresses this inconsistent concept; in particular it should specify: (i) how we should understand arguments that express truth (i.e., a logic for truth), (ii) how we should understand sentences that express truth (i.e., a semantic theory for truth), and (iii) how we should understand speech acts that express truth (i.e., a pragmatic theory for truth). It is my view that an adequate account of confusion should give us a good idea of how to construct these theories. In particular, I

advocate a logic for truth that is a type of relevance logic, a semantic theory for truth that is based on an inferential role theory of meaning, and pragmatic theory for truth that is based on a scorekeeping theory of linguistic usage.<sup>78</sup>

When considering an inconsistent concept like truth, it is essential to distinguish between several sets of rules for using it. First, there are the incompatible rules that are constitutive of truth (i.e., the truth rules). Those who employ truth *try* to follow these rules. Second, there are the rules stipulated by the logic, the semantic theory, and the pragmatic theory for truth. An interpreter who knows that truth is inconsistent treats those who employ it as if they are bound by these rules. Third, there are the rules stipulating that truth should not be employed at all. Those who know that truth is inconsistent are bound by these rules.

Although there are many complicated issues involved in deciding on the best account of the components of truth, it seems to me that there are there are two fundamental components: *ascending truth* and *descending truth*.<sup>79</sup> Both of them obey versions of mono-aletheism (i.e., no truth bearer is both ascending true and ascending false, and no truth bearer is both descending true and descending false) and substitution (i.e., any singular term that is coreferential with  $\langle\langle p \rangle\rangle$  can be substituted in  $\langle\langle p \rangle$  is ascending true $\rangle$  or  $\langle\langle p \rangle$  is descending true $\rangle$  without changing the ascending truth value or the descending truth value). Ascending truth obeys a version of the ascending truth rule for every truth apt truth bearer (i.e.,  $\langle\langle p \rangle$  is ascending true $\rangle$  follows from  $\langle p \rangle$ ), while descending truth obeys a version of the descending truth rule for every truth-apt truth bearer (i.e.,  $\langle p \rangle$  follows from  $\langle\langle p \rangle$  is descending true $\rangle$ ). However, ascending truth obeys a version of the descending truth rule (i.e.,  $\langle p \rangle$  follows from  $\langle\langle p \rangle$  is ascending true $\rangle$ ) for some truth-apt truth bearers, but not all of them. Likewise,

---

<sup>78</sup> I do not have the space to discuss these theories here; see Scharp (2005a) for details and references.

<sup>79</sup> Depending on one's views on the distinction between weak truth and strong truth, one's views on deflationism, and one's views on truth bearers, truth aptness, and truth values, one might want to distinguish between more than two components of truth. For simplicity, I discuss just two.

descending truth obeys a version of the ascending truth rule (i.e.,  $\langle\langle p \rangle\rangle$  is descending true) follows from  $\langle p \rangle$ ) for some truth-apt truth bearers, but not all of them. There are precise ways of characterizing these restrictions, but roughly, if one substitutes a truth predicate for the ascending truth predicates and the descending truth predicates in a sentence and the result is paradoxical for a theory of truth that validates the truth rules, then the original sentence does not obey either the descending rule for ascending truth or the ascending rule for descending truth. Call these sentences *pathological*. It turns out that pathological sentences are ascending true and descending false.

Ascending truth and descending truth are both partially defined in the sense that their extensions and anti-extensions are not exhaustive of the set of declarative sentences; there are ascending truth gaps and descending truth gaps, which are the same. Although I want to permit a range of views on the gaps, there is group of sentences that definitely count as ascending and descending truth gaps: the sentences that express our inconsistent concept of truth. Indeed, it is my view that any sentence that expresses an inconsistent concept is an ascending gap and a descending gap (because they are the same, I will just use ‘gap’ from here on). The reason is simple. If a sentence has a particular truth value, then it is acceptable to assert that it has this truth value. If one asserts that a sentence that expresses an inconsistent concept is true (false), then one is committed to the proposition expressed by the sentence in question (its negation) because of the truth rules. If one is committed to a proposition containing an inconsistent concept, then one is committed to the constitutive principles for that inconsistent concept.<sup>80</sup> Therefore, if sentences that express an inconsistent concept have truth values, then there is no way to avoid employing that inconsistent concept. Because it is essential to distinguish between concept employment and concept possession when it comes to inconsistent concepts, every sentence that expresses an inconsistent concept is gappy.

---

<sup>80</sup> This is not quite right; I can endorse the proposition expressed by ‘acorns are Xnot true’ without committing myself to employing the concept of truth, but this sort of example is rare.

I want to turn to the approach to the liar paradox and the revenge paradoxes. Ascending truth and descending truth differ on the pathological sentences. Consider the following two sentences:

( $\alpha$ ) ( $\alpha$ ) is ascending false.

( $\delta$ ) ( $\delta$ ) is descending false.

These sentences are pathological. However, they are not paradoxical—it is *not* the case that either the theory of ascending truth or the theory of descending truth has the following consequences: (i) ( $\alpha$ ) is both ascending true and ascending false, (ii) ( $\alpha$ ) is both descending true and descending false, (iii) ( $\delta$ ) is both ascending true and ascending false, and (iv) ( $\delta$ ) is both descending true and descending false. Indeed, these theories imply that ( $\alpha$ ) and ( $\delta$ ) are both ascending true and descending false. Consider the analogs of the liar reasoning for ( $\alpha$ ) and ( $\delta$ ) (with ‘AT’ in for ‘ascending truth’ and ‘DT’ in for ‘descending truth’):

( $\alpha$ ) is AT (assumption)	( $\delta$ ) is DT (assumption)
‘( $\alpha$ ) is AF’ is AT (substitution)	‘( $\delta$ ) is DF’ is DT (substitution)
* ( $\alpha$ ) is AF (descending)	( $\delta$ ) is DF (descending)
( $\alpha$ ) is AF (assumption)	( $\delta$ ) is DF (assumption)
‘( $\alpha$ ) is AF’ is AT (ascending)	* ‘( $\delta$ ) is DF’ is DT (ascending)
( $\alpha$ ) is AT (substitution)	( $\delta$ ) is DT (substitution)
$\therefore$ ( $\alpha$ ) is AT iff ( $\alpha$ ) AF	$\therefore$ ( $\delta$ ) is DT iff ( $\delta$ ) is DF

Neither of these arguments is valid. In the argument concerning ( $\alpha$ ), the third step (marked with a ‘\*’) is invalid, and in the argument concerning ( $\delta$ ), the fifth step (marked with a ‘\*’) is invalid. Of course, one can prove that both ( $\alpha$ ) and ( $\delta$ ) are AT and DF, but that is not a contradiction.

The details of how the inconsistency theory works will have to wait for some other occasion, but roughly, one uses the theory of ascending truth and the theory of descending truth to assign semantic values to the sentences containing ‘true’ and then one uses a special logic to evaluate the arguments containing these sentences. The semantic values have an epistemic interpretation—they are similar to the “told true”, “told false”, “told neither”, and “told both” values that are familiar from 4-valued semantics for some relevance logics (except that there are more than just four options in the case of truth). The logic used to evaluate the arguments is a type of relevance logic that is appropriate for confused concepts.<sup>81</sup> Once one has an account of inferential correctness from the logic, one can use an inferential role theory of meaning to assign meanings to the sentences that contain ‘true’ based on their inferential roles; one can also use a scorekeeping theory of speech acts to assign pragmatic features to utterances of sentences that contain ‘true’.<sup>82</sup>

Does the inconsistency theory of truth face revenge paradoxes? I cannot guarantee that it does not. However, given that my analyses of self-refutation problems and inconsistency problems are correct, I can argue that it does not. First, notice that the inconsistency theory of truth does not imply that the constitutive principles for truth are true or valid. In addition, this theory of truth does not employ the inconsistent concept of truth. The logic does not use truth-values and it does not explain validity in terms of truth preservation; instead, it uses epistemically interpreted semantic values and it explains validity in terms of profitability preservation. The pragmatic theory does not explain assertibility in terms of truth; instead, it uses the same resources as the logic. The semantic theory does not assign truth conditions to the sentences in its scope; instead, it assigns inferential

---

<sup>81</sup> See Belnap (1976, 1977) for more on the 4-valued semantics and the epistemic interpretation of the semantic values; see also Camp (2002) for the use of this logic on confused expressions. Note that the logic is based on profitability preservation instead of truth preservation. This is essential if one wants to avoid appealing to the inconsistent concept of truth in one’s theory of the inconsistent concept of truth.

<sup>82</sup> See Lewis (1979) for an example of this kind of theory.

roles. Thus, one can accept this theory of truth without employing the inconsistent concept of truth it describes.

My explanation of the inconsistency problems is that they occur for theories of truth that both validate the truth rules and classify some of the paradoxical sentences as defective, where the class of defective sentences does not include all the paradoxical ones. However, the inconsistency theory of truth I endorse classifies all the sentences that express the inconsistent concept of truth as defective. Thus, it does not face any inconsistency problems. The theories of the component concepts of truth do not respect the truth rules; hence they do not give rise to inconsistency problems.

Self-refutation problems occur for theories of truth that validate the truth rules and classify all the paradoxical sentences as defective, including some of their own consequences. However, the inconsistency theory of truth has no paradoxical sentences as consequences because it does not employ the inconsistent concept of truth. It does not classify the sentences in its scope as true or false. Instead, it implies that all sentences are in the range of inapplicability for the inconsistent concept of truth (i.e., truth has empty extension and empty anti-extension). The theory of the component concepts classifies all the sentences that express the inconsistent concept of truth as gaps. Still, one might wonder whether the theory of ascending truth or the theory of descending truth faces a self-refutation problem. Call the theory of ascending truth  $T_A$  and the theory of descending truth  $T_D$ . It might seem that both theories are pathological (i.e., they are both AT and DF). Even if this were true, it would not constitute a self-refutation problem because pathological sentences can be acceptable. However, neither theory is pathological. Consider  $T_A$ . It implies that  $(\alpha)$  is AT and DF. One might be tempted to infer from the claim that  $T_A$  implies that  $(\alpha)$  is AT, that  $T_A$  implies  $(\alpha)$ . However, the rule, from  $\langle\langle p \rangle \text{ is AT} \rangle$  infer  $\langle p \rangle$ , is invalid for pathological sentences. Now consider  $T_D$  and the following argument:

- (i)  $T_D$  implies that  $(\delta)$  is AT and DF.
- (ii)  $(\delta)$  is ‘ $(\delta)$  is DF’.
- (iii) Hence,  $(\delta)$  is a consequence of  $T_D$ .
- $\therefore$  (iv)  $T_D$  is AT and DF.

The problem with this argument is that pathologicity is not preserved by the consequence relation. One can define validity in terms of AT and DT: an argument is *valid* if it preserves DT and the absence of AF (i.e., if the premises are DT, then the conclusion is DT, and if the conclusion is AF, then one of the premises is AF). ‘ $(\delta)$  is DF’ is a consequence of  $T_D$ , and ‘ $(\delta)$  is DF’ is  $(\delta)$ ; hence,  $(\delta)$  is a consequence of  $T_D$ . However, in order to show that  $T_D$  is pathological, one would have to show that ‘ $(\delta)$  is DT’ is a consequence of  $T_D$ . However, ‘ $(\delta)$  is DT’ is not a consequence of  $T_D$  because ‘ $(\delta)$  is DF’ is a consequence of  $T_D$ . Therefore, neither  $T_A$  nor  $T_D$  is pathological.

I want to be clear about my suggestion: I am not expecting people to stop using ‘true’ and start using ‘descending true’ and ‘ascending true’. It is crucial that people be able to continue using ‘true’, but because of the division of linguistic labor and their propensity to defer to the experts in cases where it matters, their use of ‘true’ will no longer express our inconsistent concept of truth, but rather a general concept (i.e., ‘true’ will mean something like *either descending true or ascending true*).

## 5. CONCLUSION

I offer a look back and a look forward. First, there are two types of revenge paradoxes: inconsistency problems and self-refutation problems. Second, any theory of truth that offers an approach to the liar paradox and validates the truth rules is inconsistent, self-refuting, or restricted (that is the conclusion of the revenge argument). Third, truth is an inconsistent concept—a claim

justified by three arguments: (i) an inconsistency theory of truth provides the best explanation for revenge paradoxes, (ii) only an inconsistency theory of truth that implies that the truth rules are constitutive but invalid can avoid revenge paradoxes, and (iii) only the particular inconsistency theory of truth I offer has been able to avoid revenge paradoxes. Fourth, the best inconsistency theory of truth treats truth as a confused concept, identifies ascending and descending truth as its components, and does not appeal to our inconsistent concept of truth at all (indeed, it implies that our concept of truth should not be employed).

If these points are correct, then there is much to be done. First, we need to rethink our views on the nature of truth. Deflationists, correspondence theorists, and the rest treat truth as a consistent concept. If truth is an inconsistent concept, then all familiar views on the nature of truth are unacceptable. Second, we need a good theory of inconsistent concepts, which should include a logic, a semantic theory, and a pragmatic theory; if I am right that inconsistent concepts are confused, then a major part of a theory of inconsistent concepts will be a theory of confusion. Third, we need a theory of the replacement concepts for truth (i.e., ascending truth and descending truth), which should include a logic, a semantic theory, and a pragmatic theory. Fourth, we need to use the theory of the replacement concepts for truth and the theory of inconsistent concepts to arrive at a theory of our inconsistent concept of truth. Fifth, we need to start thinking about the nature of the replacement concepts. It seems to me that a broadly deflationist account will work best, but that is little more than an educated guess at this point. Sixth, we need to start thinking about consequences of this account of truth for other concepts that are typically explained by appeal to truth, including: validity, consistency, completeness, necessity, knowledge, meaning, and assertion.<sup>83</sup> Finally, we need to consider whether analogous approaches work for other paradoxes,

---

<sup>83</sup> I am especially interested in the effects the theory has on the fundamental theorems of mathematical logic (e.g., Gödel's theorems).

including: the paradoxes of predication, the paradoxes of reference, the set-theoretic paradoxes, the paradoxes of vagueness, and the paradoxes of the infinite.<sup>84, 85</sup>

---

<sup>84</sup> My suggestions for how to pursue most of the issues mentioned in this paragraph can be found in Scharp (2005a).

<sup>85</sup> Thanks to Bob Brandom, Hartry Field, Anil Gupta, John McDowell, Stewart Shapiro, Neil Tennant, Matti Eklund, Doug Patterson, Eric Carter, and A. Duncan Kerr for conversations on these issues and comments on earlier drafts.

WORKS CITED

- Allen, Martin. (2001). "A Paraconsistent-Preservationist Treatment of a Common Confusion Concerning Predicate Extensions." In *Logical Consequence: Rival Approaches* (Proc. 1999 Conf. of the Society for Exact Philosophy), John Woods & Bryson Brown (eds.), Middlesex: Hermes Science.
- Atlas, Jay. (1989). *Philosophy without Ambiguity*. Oxford: Oxford University Press.
- Azzouni, Jody. (2003). "The Strengthened Liar, the Expressive Strength of Natural Languages, and Regimentation," *The Philosophical Forum* 34: 329-50.
- Barwise, Jon and Etchemendy, John. (1987). *The Liar: An Essay on Truth and Circularity*. Oxford: Oxford University Press.
- Beall, JC. (1999). "Completing Sorensen's Menu: A Non-Modal Yabloesque Curry," *Mind* 108: 737-739.
- Beall, JC. (2001). "Is Yablo's Paradox Non-Circular?" *Analysis* 61: 176-187.
- Beall, JC (ed.). (2003). *Liars and Heaps: New Essays on Paradox*. Oxford: Clarendon Press.
- Belnap, Nuel. (1976). "How a computer should think," in *Contemporary Aspects of Philosophy*, G. Ryle (ed.), London: Oriel Press.
- . (1977). "A Useful Four-Valued Logic," in *Modern Uses of Multiple-Valued Logic*, J. M. Dunn and G. Epstein (eds). Dordrecht: D. Reidel.
- Bromberger, Sylvain. (1989). "Types and Tokens in Linguistics," in *Reflections on Chomsky*, A. George (ed.), Oxford: Blackwell.
- Bueno, Otávio and Colyvan, Mark. (2003a). "Paradox without satisfaction," *Analysis* 63: 152–6.
- . (2003b). "Yablo's paradox and referring to infinite objects," *Australasian Journal of Philosophy* 81: 402–12.
- Burge, Tyler. (1979). "Semantical Paradox," *The Journal of Philosophy* 76: 169-198.
- Camp Jr., Joseph L. (2002). *Confusion: A Study in the Theory of Knowledge*. Cambridge: Harvard University Press.
- Cappelen, Herman. (1999). "Intentions in Words," *Noûs* 33: 92-102.
- Chapuis, André and Gupta, Anil (eds.). (2000). *Circularity, Definition and Truth*. New Delhi: Indian Council of Philosophical Research.
- Chihara, Charles. (1979). "The Semantic Paradoxes: A Diagnostic Investigation," *Philosophical Review* 88: 590-618.
- . (1984). "The Semantic Paradoxes: Some Second Thoughts," *Philosophical Studies* 45: 223-229.
- Davidson, Donald. (1974). "Belief and the Basis of Meaning," in Davidson (1984).
- . (1982). "Communication and Convention," in Davidson (1984).
- . (1984). *Inquiries into Truth and Interpretation*. Oxford: Oxford University Press.
- Dummett, Michael. (1973). *Frege: Philosophy of Language*, Cambridge: Harvard University Press.
- . (1999). "Of What Kind of Thing is Truth a Property?" in *Truth*, Blackburn and Simmons (eds.), Oxford: Oxford University Press..
- Eklund, Matti. (2002). "Inconsistent Languages," *Philosophy and Phenomenological Research* 64: 251-275.
- . (Forthcoming). "The Liar Paradox, Expressibility, and Possible Languages," in *Revenge Paradoxes*, JC Beall (ed.), Oxford: Oxford University Press.
- Feferman, Solomon. (1982). "Toward Useful Type-Free Theories I," *The Journal of Symbolic Logic* 49: 75-111.

- . (1991). “Reflecting on Incompleteness,” *Journal of Symbolic Logic* 56: 1-49.
- Field, Hartry (1973). “Theory Change and the Indeterminacy of Reference,” in Field (2001a).
- . (1974). “Quine and the Correspondence Theory,” in Field (2001a).
- . (2002). “Saving the Truth Schema From Paradox,” *Journal of Philosophical Logic* 31: 1-27.
- . (2003a). “A Revenge-Immune Solution to the Semantic Paradoxes,” *Journal of Philosophical Logic* 32: 139-177.
- . (2003b). “The Semantic Paradoxes and the Paradoxes of Vagueness,” in Beall (2003).
- . (2003c). “No Fact of the Matter,” *Australasian Journal of Philosophy* 81: 457-480.
- . (2004). “The Consistency of the Naïve Theory of Properties,” *The Philosophical Quarterly* 54: 78-104.
- . (2005a). “Is the Liar Sentence Both True and False?” in *Deflationism and Paradox*, JC Beall and B. Armour-Garb (eds.), Oxford: Oxford University Press.
- . (2005b). “Variations on a Theme by Yablo,” in *Deflationism and Paradox*, JC Beall and B. Armour-Garb (eds.), Oxford: Oxford University Press.
- . (Forthcoming). “Solving the Paradoxes, Escaping Revenge,” in *Revenge Paradoxes*, JC Beall (ed.), Oxford: Oxford University Press.
- Fodor, Jerry and Lepore, Ernest. (1994). *Holism: A Shopper’s Guide*, Cambridge: MIT Press.
- Gaifman, Haim. (1992). “Pointers to Truth,” *The Journal of Philosophy* 89: 223-261.
- . (2000). “Pointers to Propositions,” in Chapuis and Gupta (2000).
- Glanzberg, Michael. (2001). “The Liar in Context,” *Philosophical Studies* 103: 217-251.
- . (2004). “A Contextual-Hierarchical Approach to Truth and the Liar Paradox,” *Journal of Philosophical Logic* 33: 27-88.
- . (2005). “Truth, Reflection, and Hierarchies,” *Synthese* 142: 289-315.
- Gupta, Anil. (1999). “Meaning and Misconceptions,” in *Language, Logic, and Concepts*. R. Jackendoff, P. Bloom and K. Wynn (eds.), Cambridge: MIT Press.
- Gupta, Anil, and Belnap, Nuel. (1993). *The Revision Theory of Truth*. Cambridge: MIT Press.
- Gupta, Anil, and Martin, Robert L. (1984). “A Fixed Point Theorem for the Weak Kleene Valuation Scheme,” *Journal of Philosophical Logic* 13: 131-135.
- Hardy, James. (1995). “Is Yablo’s Paradox Liar-Like,” *Analysis* 55: 197-198.
- Hazen, Allen. (1990). “A Variation on a Paradox,” *Analysis* 50: 7-8.
- Herzberger, Hans G. (1982a). “Naïve Semantics and the Liar Paradox,” *The Journal of Philosophy* 79: 479-497.
- Herzberger, Hans G. (1982b). “Notes on Naïve Semantics,” *Journal of Philosophical Logic* 11: 61-102.
- Horn, Lawrence. (1989). *A Natural History of Negation*. Stanford: CSLI Publications.
- Horwich, Paul. (1998). *Truth*. 2<sup>nd</sup> ed. Oxford: Clarendon Press.
- Hugly, Philip and Sayward, Charles. (1981). “Expressions and Tokens,” *Analysis* 41: 181-189.
- Kaplan, David. (1973). “Bob and Carol and Ted and Alice,” *Approaches to Natural Language*, Hintikka et Ketland (2004, 2005)
- . (1990). “Words,” *The Aristotelian Society, Supplementary Volume*, 64: 93-119.
- Ketland, Jeffrey. (2004). “Bueno and Colyvan on Yablo’s Paradox,” *Analysis* 64: 165-72.
- Koons, Robert. (1992). *Paradoxes of Belief and Strategic Rationality*. Cambridge: Cambridge University Press.
- . (2000). “Circularity and Hierarchy,” in Chapuis and Gupta (2000).
- Kripke, Saul. (1975). “Outline of a Theory of Truth,” *The Journal of Philosophy* 72: 690-716.

- Leitgeb, Hannes. (2002). "What is a Self-Referential Sentence? Critical Remarks on the Alleged (Non-)Circularity of Yablo's Paradox," *Logique et Analyse* 177–178: 3–14.
- Lewis, David. (1979). "Scorekeeping in a Language Game," in *Philosophical Papers*, vol. 1. Oxford: Oxford University Press. 1983.
- Martin, R. M. (ed.). (1984). *Recent Essays on Truth and the Liar Paradox*. Oxford: Clarendon Press.
- Maudlin, Tim. (2004). *Truth and Paradox: Solving the Riddles*. Oxford: Oxford University Press.
- McDowell, John. (1994). *Mind and World*, Cambridge: Harvard University Press.
- McGee, Vann. (1991). *Truth, Vagueness, and Paradox*. Indianapolis: Hackett.
- Meyer, Robert K., Routley, Richard, and Dunn, J. Michael. (1979). "Curry's Paradox," *Analysis* 39: 124-128.
- Parsons, Charles. (1974). "The Liar Paradox," *Journal of Philosophical Logic* 3: 381-412.
- Parsons, Terence. (1984). "Assertion, Denial, and the Liar Paradox," *Journal of Philosophical Logic* 13: 137-152.
- Patterson, Doug. (2006). "Tarski, the Liar, and Inconsistent Languages," *The Monist* 89:
- Peacocke, Christopher. (1992). *A Study of Concepts*. Cambridge: MIT Press.
- Priest, Graham. (1990). "Boolean Negation and All That," *Journal of Philosophical Logic* 19: 201-215.
- . (1995). "Multiple Denotation, Ambiguity and the Strange Case of the Missing Amoeba," *Logique et Analyse*, 38: 361-73.
- . (1997). "Yablo's Paradox," *Analysis* 57: 236-242.
- . (2006). *Doubt Truth to be a Liar*. Oxford: Oxford University Press.
- Priest, Graham, Beall, JC, and Armour-Garb, Brad (eds.). (2004). *The Law of Non-Contradiction: New Philosophical Essays*, Oxford: Oxford University Press.
- Prior, A. N. (1960). "The Runabout Inference Ticket,"
- Quine, W. V. (1960). *Word and Object*. Cambridge: MIT Press.
- Scharp, Kevin. (2005a). "Truth and Aletheic Paradox," Ph.D. dissertation, University of Pittsburgh.
- . (2005b). "Scorekeeping in a Defective Language Game," *Pragmatics and Cognition* 13: 203-226.
- . (FTT). "Fragmentary Theories of Truth," in preparation.
- . (TI). "Truth and Internalizability," in preparation.
- . (RB). "Risky Business: Truth and Paradoxicality," in preparation.
- . (AP). "The Aletheic Problem," in preparation.
- Schiffer, Stephen. (2003). *The Things We Mean*. Oxford: Oxford University Press.
- Simmons, Keith. (1993). *Universality and the Liar: An Essay on Truth and the Diagonal Argument*. Cambridge: Cambridge University Press.
- Simons, Peter. (1982). "Token Resistance," *Analysis* 42: 195-203.
- Skyrms, Brian. (1982). "Intensional Aspects of Semantical Self-Reference," in Martin (1984).
- Sorensen, Roy A. (1998). "Yablo's Paradox and Kindred Infinite Liars," *Mind* 107: 137-156.
- Stebbins, Sarah. (1992). "A Minimal Theory of Truth," *Philosophical Studies* 66: 109-137.
- Szabó, Zoltán Gendler. (1999). "Expressions and their Representations," *Philosophical Quarterly* 49: 145-163.
- Tappenden, Jamie. (1999). "Negation, Denial, and Language Change in Philosophical Logic," in *What is Negation?*, D. Gabbay and H. Wansing (eds.), Dordrecht: Kluwer.
- Tennant, Neil. (1995). "On Paradox without Self-Reference," *Analysis* 55: 199-207.
- Truncellito, David. (2000). "Which Type is Tokened by a Token of a Word-Type?" *Philosophical Studies* 97: 251-266.

- Wetzel, Linda. (1993). "What are Occurrences of Expressions?" *Journal of Philosophical Logic* 22: 215-220.
- Williamson, Timothy. (1997). "Imagination, Stipulation and Vagueness," in *Philosophical Issues* 8: *Truth*, E. Villanueva (ed.), Atascadero, CA: Ridgeview.
- . (2000). "Semantic Paradox and Semantic Change," in *Proceedings of the Twentieth World Congress of Philosophy, vol. 6*, A. Kanamori (ed.), Bowling Green: Philosophy Documentation Center.
- . (2003). "Understanding and Inference," *The Aristotelian Society Supplement* 77: 249-293.
- . (Forthcoming). "Reference, Inference and the Semantics of Pejoratives", in a *festschrift* for David Kaplan, J. Almog and P. Leonardi, (eds.), Oxford: Oxford University Press.
- van Benthem, J. F. A. K. (1978). "Four Paradoxes," *Journal of Philosophical Logic* 7: 49-72.
- Yablo, Stephen. (1985). "Truth and Reflection," *Journal of Philosophical Logic* 14: 297-349.
- . (1993a). "Definitions, Consistent and Inconsistent," *Philosophical Studies* 72: 147-175.
- . (1993b). "Hop, Skip and Jump: The Agnostic Conception of Truth," *Philosophical Perspectives* 7: 371-396.
- . (1993c). "Paradox Without Self-Reference," *Analysis* 53: 251-252.