Minimalism, Deflationism, and Paradoxes[*]

Michael Glanzberg
University of Toronto

September 22, 2009

This paper argues against a broad category of deflationist theories of truth. It does so by asking two seemingly unrelated questions. The first is about the well-known logical and semantic paradoxes: Why is there no strengthened version of Russell's paradox, as there is a strengthened version of the Liar paradox? Oddly, this question is rarely asked. It does have a fairly standard answer, which I shall not dispute for purposes of this paper. But I shall argue that asking it ultimately leads to a fundamental challenge to some popular versions of deflationism.

The challenge comes about by pairing this question with a second question: What is the theory of truth about? For many theorists, there is an obvious answer to this question: the theory of truth is about truth bearers and what makes them true. But this answer appears to bring with it a commitment to a substantial notion of truth, which deflationists cannot bear. Deflationists might prefer a very different answer: the theory of truth is not really about anything. There is no substantial property of truth, so there is no domain which the theory of truth properly describes. Not all positions under the name 'deflationism' subscribe to this view, but I shall argue that the important class of so-called minimalist views do.

I shall argue that this sort of deflationist answer is untenable, and thus argue in broad strokes against minimalism. I shall argue by way of a comparison of the theory of truth with the theory of sets, and consideration of where paradoxes, especially strengthened versions of the paradoxes, may arise in each. This will bring the two seemingly unrelated questions together to form an anti-deflationist argument. I shall show that deflationist positions that accept the idea that truth is

[*] This is a revised and expanded version of my "Minimalism and Paradoxes" Synthese 135 (2003): 13–36. Thanks to Brad Armour-Garb and J. C. Beall, and to Otavio Bueno, Alex Byrne, Ned Hall, Richard Heck, Jim Pryor, Susanna Siegel, Judith Thomson, Ralph Wedgwood, and Steve Yablo. Some of this material was presented at the University of Southern California. Thanks to my audience there for valuable comments and discussion.

not a real or substantial property are too much like naive set theory. Like naive set theory, they are unable to make any progress in resolving the paradoxes, and must be replaced by a drastically different sort of theory. Such a theory, I shall show, must be fundamentally non-minimalist. I shall then turn to the question of how close to minimalism one can come and avoid the problem I shall raise. I shall suggest, though more tentatively, that a much wider class of deflationist views of truth are undermined by the argument I shall present.

My argument proceeds in six sections. In the first, Section (I), I make the comparison between naive set theory and the minimalist version of deflationism, and explain the sense in which both theories can be said not to be about anything. In Section (II), I point out how both theories suffer from nearly identical problems, as Russell's paradox and the Liar paradox may be seen to be extremely similar in important respects. In Section (III), I show where these parallels break down. The sort of response to the Liar which might be offered by minimalism proves to be unstable, in that it is vulnerable to the Strengthened Liar paradox. The standard response to Russell's paradox in set theory is not so unstable. In Section (IV), I investigate the source of this difference. I argue there that any theory of truth able to evade the Strengthened Liar must at least be about some domain in the way that standard set theory is about the domain of sets. Then in Section (V), I show that the usual ways of avoiding the Strengthened Liar meet this condition by abandoning minimalism for a more correspondence-like notion of truth. Finally, I show in Section (VI) that no minimalist position can meet the condition, and so none is tenable. I conclude this section by considering whether any other version of deflationism might fare better. I shall tentatively suggest none can.

I. Minimalism and Naive Set Theory

The term 'deflationism' covers a wide range of philosophical views. My primary concern here is with the species of deflationism known as 'minimalism'. This is again really a class of views. As a class, it is distinguished by three distinctive marks. The first is the idea that in some appropriate sense there is no substantial or genuine property of truth. For instance, Paul Horwich writes:

> Unlike most other predicates, 'is true' is not used to attribute to certain entities (i.e.

> statements, beliefs, etc.) an ordinary sort of property—a characteristic whose underlying nature will account for its relation to other ingredients of reality. Therefore, unlike most other predicates, 'is true' should not be expected to participate in some deep theory of that to which it refers . . . (Horwich, 1990, p. 2.)

Horwich does not quite claim that there is no such property as truth, but the sense in which there is a property seems to amount to little more than there being a predicate of truth in our language. There is no genuine phenomenon of being true which this predicate describes.

The remaining two marks of the class of views I shall consider center around the T-schema:

> 's' is true iff s.

Those who maintain that there is no substantial property of truth cannot maintain that instances of this schema hold because of the nature of the property; it cannot hold in virtue of the nature of truth. Instead, they must say that the schema holds analytically, or perhaps by definition, or perhaps by stipulation. This is the second mark of the class of minimalist views. There are, of course, important differences between the specific ideas of analyticity, definition, and stipulation; but they will not matter for our purposes here. What is important here is that any of these options provides the T-schema with a status that ensures its truth without looking to the nature of the property of truth ('underlying nature' as Horwich puts it). In what follows, I shall compare this status to that of logical truth.

The final mark of the class of minimalist views is the idea that rather than describing a feature of truth, the T-schema provides us with a device of disquotation. This device is useful, for instance, as it allows us to make infinitary generalizations. Putting these marks together, a minimalist holds that the stipulative or analytic T-schema provides us with a useful linguistic device, rather than describing a genuine property.

The class of views on which I shall focus are thus marked by the claims that there is no substantial property of truth, the T-schema is analyticity, and that truth is a device of disquotation. It is more or less standard to call any view within this class minimalism, though as I mentioned, a number of distinct positions within the class can be discerned. It will thus be useful to identify a particularly straightforward version of minimalism, which I shall call pure minimalism. Pure minimalism is distinguished by taking the instances of the T-schema to hold for any well-formed declarative sentence. Beyond that, it insists that the marks of minimalism comprise all there is to say about truth.[1]

Pure minimalism factors into two components. One is a theory in the logician's sense. The core of the theory is the T-schema, taken now as an axiom schema:

$$Tr(\ulcorner\phi\urcorner) \leftrightarrow \phi. \tag{T}$$

We need to construe this as added to a theory strong enough to do some elementary syntax. It must have a name $\ulcorner\ulcorner\phi\urcorner\urcorner$ for each sentence $\ulcorner\phi\urcorner$, and I shall assume the theory is strong enough to allow the Diagonal Lemma to apply. Let us call this theory M.

The other component consists of the philosophical commitments of pure minimalism. In many cases, we think of the philosophical commitments going with a formal theory as helping to describe the intended interpretation of the theory. For a minimalist, this is on odd way to put the idea (as I shall discuss more in a moment), but at the very least, the philosophical commitments do help explain how M is to be understood.

Pure minimalism includes the feature of minimalism about truth bearers. Truth bearers are appropriate candidates for truth, and have a truth status. They are true or false.[2] Truth bearers need not be true, but they must be truth apt. According to pure minimalism, to be a truth bearer is nothing but to figure into predications of $\ulcorner Tr\urcorner$ or $\ulcorner\neg Tr\urcorner$. In the presence of the T-schema or the axiom schema (T), this occurs for every well-formed declarative sentence. Pure minimalism is

[1] Many current minimalist positions depart from pure minimalism in some ways. I shall return to other versions of minimalism in Section (VI). It is not entirely clear whether anyone has actually held pure minimalism, but regardless, I believe it encapsulates an important idea, which is reflected in the positions of a number of authors. Pure minimalism is often attributed to Ayer (1946). Ayer defines truth for propositions rather than sentences, which technically to make him something other than a pure minimalist. However, as I shall discuss more in Section (VI), his definition of proposition is sufficiently closely tied to sentences that this difference may be insignificant. Horwich (1990) likewise holds most of the theses of pure minimalism, but construes truth as applying to propositions. Some of his remarks, especially in Horwich (1994), suggest his departure from pure minimalism may be as minimal as Ayer's. The third mark of minimalism—truth as a device of disquotation—is closely associated with Quine, and Quine (1986) comes quite close to pure minimalism. Of course, Quine would never stand for an inconsistent theory like M.

[2] Views which rely on many-valued logics will rather say that that truth bearers are true, false, or any of the other truth values.

thus minimalist about truth bearers in that it says no more about what makes something a truth bearer than it does about truth. Of course, the class of well-formed declarative sentences must be delineated by the syntax component of the theory, but this tells us nothing about their status as truth bearers. When it comes to this status, all that the theory tells us derives from the analytic (or whatever other appropriate status) schema (T). There is thus no underlying property that makes it the case that declarative sentences are all truth bearers, as there is no underlying property that makes the instances of (T) hold.[3]

I shall compare pure minimalism to naive set theory. This will help frame the question of what, according to the minimalist, the theory of truth is about. Like pure minimalism, naive set theory factors into two components. One is again a formal theory. Again it is captured primarily by a single axiom schema, the naive comprehension schema:

$$\text{(COMP)} \qquad y \in \{x \mid \phi(x)\} \leftrightarrow \phi(y).$$

As with (T), we must think of this as added to an appropriate base theory, which is able to construct a name $\ulcorner \{x \mid \phi(x)\} \urcorner$ for the set determined by $\ulcorner \phi \urcorner$. We might also assume a principle of extensionality, but it will not matter for the discussion to follow. Let us call this theory N.

Like pure minimalism, naive set theory comes with a philosophical component as well. I have in mind naive set theory as it would have been understood by someone who really held it: Frege, or in some form perhaps a traditional logician.[4] Such a theorist would hold that (COMP) is in some way a logical principle. Now, there have been a great many ideas about what makes something a logical principle. But common to them is one of two thoughts. Either logical principles are schematic, and so not about anything in particular, or logical principles are about absolutely everything. Both of these lines of thought reach the same conclusion: logical principles do not hold because of the nature of a specific range of objects or properties or phenomena.

[3] A well-know argument of Jackson et al. (1994) attempts to show that minimalism about truth does not lead to minimalism about truth bearers. I do think that the points I am making here reveal a genuine commitment of pure minimalism, as I shall discuss more in Section (VI), when discussing departures from pure minimalism.

[4] Both Frege (at some moments) and, say, the Port-Royal logicians, would have preferred 'extension' to 'set'.

In this regard, naive set theory turns out to be remarkably similar to pure minimalism. Both assign remarkably similar status to their fundamental principles (COMP) and (T). They agree that there is no underlying nature of anything in particular which makes these principles true. Both agree that there are no specific objects or properties that the fundamental principles of their respective theories describe. Rather, these principles are in the general class of the logical, or the analytic, or the definitional. I do not want to go so far as to assimilate logical truth to analytic or definitional truth. I only need to note that the principles in these categories share the important feature of there being no underlying natures of anything in particular to which they owe their truth.

In this way, both pure minimalism and naive set theory may be described as not being about anything in particular. Now, it may be noted that the truth predicate occurs in (T) and set abstracts in (COMP), so it could be said that one is about truth and the other sets. But at best, this is so in an entirely minimal way. As we observed, for neither schema is there any special domain of objects, properties, events, or any other phenomena that makes its instances true. As they are logical or analytic, these schemas are not about anything in particular. In the case of naive set theory, this is reflected both by the philosophical gloss on the theory, and by the unrestricted nature of set abstraction. As a matter of logic, any objects of any kind may be collected into a set. There is thus no special domain of the theory. In the case of pure minimalism, we can likewise observe that the theory cannot reveal a basic feature of the property of truth, nor can it provide any more substantial an account of what makes something a truth bearer. There is no more a special domain of this theory than there is of the naive theory of sets. As I mentioned above, the truth bearers are the syntactically well-formed declarative sentences. It is thus tempting to say that the theory is about these sentences. But it is so only in a trivial way. The theory appropriates some syntax, but this tells us nothing about truth. The principles that are supposed to tell us something about truth fall into a different category, and these hold of well-formed sentences only because this is the way the stipulations themselves are syntactically well-formed. Truth is thus predicated as widely as makes syntactic sense, not on the basis of the nature of any particular domain. Though when we write the theory down we rely on some syntax to do so, a far as the basic commitments of minimalism go, there is in no substantial sense a special domain of the theory of truth.

Pure minimalism and naive set theory do differ in some ways. Pure minimalism does not quite claim to be a matter of logic, and naive set theory offers nothing like semantic ascent. But we have now seen an important similarity between them. Both theories rely on principles which hold in some other way than by accurately describing a domain, and as a result both theories are in similar ways not genuinely about anything.

II. Paradoxes

So far, we have identified pure minimalism as a representative of the class of minimalist views. We then saw that pure minimalism is in one important respect like naive set theory, as both theories can be described as not being about anything in particular.

Pure minimalism has a formal component M, and naive set theory has a formal component N. Both M and N are inconsistent, as is well-known. Russell's paradox shows N to be inconsistent, and the Liar paradox does the same for M. But the response to its paradox has been quite different for each. Naive set theory is usually taken to be a disaster, while minimalism is often taken to be in need of modification but still viable. Given the similarities between the two theories we have seen, this may appear odd. This section will show how odd it is, by showing just how similar the paradoxes are in some important respects. In the following sections, this will lead us to consider a crucial difference between responses to the paradoxes, which will in turn show us something about the viability of minimalism.

Let us first consider the familiar Liar paradox, which shows M to be inconsistent. Using the Diagonal Lemma, we can find a sentence $\ulcorner\lambda\urcorner$ such that:

$$M \vdash \lambda \leftrightarrow \neg Tr(\ulcorner\lambda\urcorner).$$

Combining this with (T) gives the contradiction:

$$M \vdash Tr(\ulcorner\lambda\urcorner) \leftrightarrow \lambda \leftrightarrow \neg Tr(\ulcorner\lambda\urcorner).$$

The Diagonal Lemma hides the procedure for producing $\ulcorner\lambda\urcorner$, but it is clear that $\ulcorner\lambda\urcorner$ 'says of itself' that it is not true. We then ask about the truth of this sentence, and see that it is true just in case it is not true.[5]

We do virtually the same thing to produce Russell's paradox, which shows N to be inconsistent. With the Liar, we found a sentence that says of itself that it is not true. Here we need a predicate that says something is not in itself, i.e. $\ulcorner \neg x \in x \urcorner$. With the Liar, we asked about the truth of that very sentence. Here we ask about this predicate applying to its own extension. Let its extension be $R = \{x \mid \neg x \in x\}$. From (COMP) we have:

$$N \vdash R \in R \leftrightarrow \neg R \in R.$$

As with the Liar, we have a contradiction.

The two paradoxes differ in that Russell's paradox involves class abstracts and membership, while the Liar paradox truth involves truth, but otherwise, we do basically the same thing in both. The similarity between the two may be brought out even more explicitly by replacing (T) and (COMP) with a single principle. Consider a family of predicates $\ulcorner Sat_n(x, y_1, \ldots, y_n) \urcorner$, and corresponding axioms:

$$\text{(SAT)} \qquad Sat_n(\ulcorner \phi \urcorner, y_1, \ldots, y_n) \leftrightarrow \phi(y_1, \ldots, y_n).$$

If we replace $\ulcorner Sat_0(\ulcorner \phi \urcorner) \urcorner$ by $\ulcorner Tr(\ulcorner \phi \urcorner) \urcorner$, we have (T). If we replace $\ulcorner Sat_1(\ulcorner \phi \urcorner, y) \urcorner$ by $\ulcorner y \in \{x \mid \phi(x)\} \urcorner$ we have (COMP).

A more general diagonal construction yields the inconsistency of (SAT). We need only be able to prove for any predicate $\ulcorner F(x, y_1, \ldots, y_n) \urcorner$ there is a $\ulcorner Q(y_1, \ldots, y_n) \urcorner$ such that:

$$Q(y_1, \ldots, y_n) \leftrightarrow F(\ulcorner Q \urcorner, y_1, \ldots, y_n).$$

This is a straightforward modification of the more familiar Diagonal Lemma (see Boolos, 1993). Let S be a theory that contains (SAT) and can prove this generalized Diagonal Lemma.

[5] Many minimalists add a clause saying something like 'only non-problematic instances of (T)'. The success of this has been discussed by McGee (1992) and Simmons (1999).

The proof that S is inconsistent is a generalization of both the Liar and Russell arguments. Consider the predicate $\ulcorner \neg Sat_n \urcorner$ for any n. Using the generalized Diagonal Lemma, we may find a predicate $\ulcorner \sigma_n(y_1, \ldots, y_n) \urcorner$ such that:

$$S \vdash \sigma_n(y_1, \ldots, y_n) \leftrightarrow \neg Sat_n(\ulcorner \sigma_n \urcorner, y_1, \ldots, y_n).$$

Combining this with (SAT), we have:

$$S \vdash Sat_n(\ulcorner \sigma_n \urcorner, y_1, \ldots, y_n) \leftrightarrow \sigma_n(y_1, \ldots, y_n) \leftrightarrow \neg Sat_n(\ulcorner \sigma_n \urcorner, y_1, \ldots, y_n).$$

For each n, the schema (SAT) produces inconsistency.

The argument here is the same as that used in both the Liar and Russell's paradoxes. For the case of $n = 0$, we have the Liar. The $\ulcorner Sat_0 \urcorner$ instances of (SAT) are just (T), and the generalized Diagonal Lemma yields $\ulcorner \lambda \urcorner$. For the case of $n = 1$, we have a version of Russell's paradox. The $\ulcorner Sat_1 \urcorner$ instances of (SAT) provide a version of (COMP). The generalized Diagonal Lemma give us a formula $\ulcorner \rho(y) \urcorner$ such that:

$$\rho(y) \leftrightarrow \neg Sat_1(\ulcorner \rho \urcorner, y).$$

This is essentially the Russell predicate. As with the original Russell predicate, applying it to itself we see:

$$\rho(\ulcorner \rho \urcorner) \leftrightarrow \neg Sat_1(\ulcorner \rho \urcorner, \ulcorner \rho \urcorner) \leftrightarrow \neg \rho(\ulcorner \rho \urcorner).$$

The use of (SAT) makes all the more clear that the formal differences between the Liar and Russell's paradox are incidental, amounting to no more than the presence of a parameter. This has virtually no effect on the way the paradoxes are generated.[6]

We now have seen two theories, pure minimalism and naive set theory, that are strikingly similar in an important respect. Both can be described as not being about anything in particular. We

[6] My presentation of M and N, and of the paradoxes, draws heavily on Feferman (1984). For further discussion of $\ulcorner Sat_n \urcorner$, and its relation to set theory, see Parsons (1974b).

I should mention that in pointing out the similarities between the Liar paradox and Russell's paradox, I am not particularly taking issue with the original distinction between semantic and logical paradoxes of Ramsey (1926). The issues he raised are somewhat different than those that bear here.

have also seen two paradoxes, or rather two versions of basically the same paradox, which show the two theories to have inconsistent formal components. From here on, however, the situations with truth and sets diverge rather drastically, as we shall see in the next section.

III. Strengthened Paradoxes

Responses to these paradoxes are well-known. In this section, I shall consider representative responses to each. I shall show that even taking a response to the Liar into account leaves pure minimalism vulnerable to an additional 'strengthened' paradox, while a standard way of of responding to Russell's paradox is not so vulnerable. In the following sections, I shall use this difference to argue that not being about anything is a fatal flaw in minimalism.

Let us first consider how the pure minimalist might respond to the Liar. The pure minimalist may well want to maintain that the paradox is simply a technical 'glitch', and does not present a deep problem for the philosophical position. According to this stance, the right response is to hold on to the philosophical account of truth, as much as is possible, but find a way to modify the formal theory M to avoid what is seen as a merely technical failure. The usual approach is to say that though the basic idea behind (T) is right, it is technically misstated, and needs to be revised.

One leading idea for revising (T) is to make the truth predicate somehow partial, so that problematic sentences like $\ulcorner\lambda\urcorner$ come out neither true nor false. There are many different ways to implement this idea. For discussion purposes, I shall sketch one that that is relatively simple, and remains in some ways close in spirit to M. The idea is to replace the axiom schema (T) with the following collection of inference rules:

(INF)
$$\frac{P \vdash Tr(\ulcorner\phi\urcorner)}{P \vdash \phi} \qquad \frac{P \vdash \neg Tr(\ulcorner\phi\urcorner)}{P \vdash \neg\phi}$$
$$\frac{P \vdash \phi}{P \vdash Tr(\ulcorner\phi\urcorner)} \qquad \frac{P \vdash \neg\phi}{P \vdash \neg Tr(\ulcorner\phi\urcorner)}.$$

Call the resulting theory P. A theory like P can be modified or extended in many ways, but it will suffice to illustrate the point as it stands.[7] P makes truth partial in the following sense. For

[7] Many theories implement partiality in a more model-theoretic way, along the lines of Kripke (1975). I have taken as my example for discussion a proof-theoretic approach, which modifies (T) explicitly. I have chosen this route

some sentences $\ulcorner\phi\urcorner$, we have $P \vdash Tr(\ulcorner\phi\urcorner)$, so according to P, $\ulcorner\phi\urcorner$ is true. For some sentences $\ulcorner\phi\urcorner$ we have $P \vdash \neg Tr(\ulcorner\phi\urcorner)$, so according to P, $\ulcorner\phi\urcorner$ is false. (Observe if $P \vdash \neg Tr(\ulcorner\phi\urcorner)$, then $P \vdash Tr(\ulcorner\neg\phi\urcorner)$.) But for some sentences, like $\ulcorner\lambda\urcorner$, we have neither, so P assigns such sentences neither the value true nor the value false. (In many cases, we expect to get results like $\ulcorner Tr(\ulcorner\phi\urcorner)\urcorner$ from P together with some other theory, which tells us the facts of some special science. It is the theory of truth together with the theories of physics, chemistry, etc. which tells us what is true. But the role of theories of special sciences does not matter for a sentence like $\ulcorner\lambda\urcorner$, which contains no terms from any special science not already incorporated into P. All that appear in $\ulcorner\lambda\urcorner$ are a sentence name, $\ulcorner Tr\urcorner$, and the negation operator. Hence, I shall ignore this role in what follows, and just speak of what P tells us is or is not true.)

Philosophically, it appears that the move from M to P does not change the commitments of pure minimalism. The rules in (INF) might be glossed as having the same analytic or definitional status as (T), and they do substantially the same job of introducing a device of disquotation. There are some complications, of course. Inference rules are not schemas whose instances can be analytically or definitionally true. But we can say that the transition from premise to conclusion is in some way analytically correct—perhaps correct in virtue of the meaning of 'true'—or is correct as a matter of stipulation, etc. Hence, the pure minimalist can give much the same gloss to these rules as was given to (T). Though there are a number of issues raised by the step from M to P, I think we can fairly grant P to the pure minimalist for argument's sake.

P is consistent; yet the Liar paradox makes trouble for it nonetheless. This is because of what is known as the Strengthened Liar paradox. We reason as follows. The partiality of P ensures that $\ulcorner\lambda\urcorner$ does not come out true, in that $P \nvdash Tr(\ulcorner\lambda\urcorner)$. So it seems, using P as a guide, we have come to conclude $\ulcorner\lambda\urcorner$ is not true, i.e. $\ulcorner\neg Tr(\ulcorner\lambda\urcorner)\urcorner$. But $\ulcorner\lambda\urcorner$ just 'says' $\ulcorner\neg Tr(\ulcorner\lambda\urcorner)\urcorner$. P itself tells us this, as $P \vdash \lambda \leftrightarrow \neg Tr(\ulcorner\lambda\urcorner)$. Thus, it appears that just relying on P, we have come to conclude $\ulcorner\lambda\urcorner$. We are now back in paradox.

mostly because it is simple to present, and eases the comparison with naive set theory. Most of what I say applies equally to other approaches to partiality.

The rules of (INF) appear in McGee (1991), though McGee has much more to say about the issue. For proof-theoretic investigation of similar systems, see Friedman and Sheard (1987). Much stronger systems invoking partiality are developed in Feferman (1991). I have discussed some further issues surrounding partiality in my (forthcomingb).

Now, this inference cannot be carried out in P, so P remains consistent. But it still poses a problem. The conclusion we draw seems to be entirely correct, whether it can be carried out in P or not. P is designed precisely to make sure $\ulcorner\lambda\urcorner$ does not come out true. That is how consistency is achieved. So, we simply rely on P to come to the conclusion that $\ulcorner\lambda\urcorner$ is not true. Insofar as P is supposed to capture the notion of truth, it appears this is just the conclusion $\ulcorner\neg Tr(\ulcorner\lambda\urcorner)\urcorner$. Opinions differ on just how serious a problem this is, and how it may be solved.[8] For our purposes here, all I need to insist upon is that the inference is intuitively compelling, and poses a problem that requires a solution one way or another.

The partiality response to the Liar, embodied in P, is vulnerable to the Strengthened Liar paradox. So, our modified pure minimalism based on P is likewise vulnerable. Continuing our comparison between the theory of truth and the theory of sets, we should consider a solution to Russell's paradox, and see if there is a strengthened version of this paradox to which the solution remains vulnerable.

There is, I believe, a standard response to Russell's paradox. It has two components, corresponding to the two components of naive set theory. First, the inconsistent formal theory N is replaced. There are a number of plausible candidates to replace it, but for illustration, let us take the Bernays-Gödel theory of sets and classes BGC. Second, the philosophical gloss on the naive theory is replaced by an account of the domains of sets and classes. Again there are a few competitors, but for argument's sake, let us assume some version of the iterative conception of set, together with the idea that (proper) classes are the extensions of predicates of sets.[9] Let us call the combination of these the standard theory.

[8] I have discussed this further in my (2001).

[9] BGC is a two-sorted theory, with variables $\ulcorner x\urcorner$ for sets and $\ulcorner X\urcorner$ for classes. The crucial axioms are restricted class comprehension $\ulcorner\forall X_1,\ldots,\forall X_n\exists Y(Y=\{x\mid\phi(x,X_1,\ldots,X_n)\})\urcorner$ where only set variables are quantified in $\ulcorner\phi\urcorner$, and axioms that say every set is a class and if $X\in Y$, then X is set. BGC also has the usual pairing, infinity, union, powerset, replacement, foundation, and choice axioms, as well as a class form of extensionality (see Jech, 1978).

For those unfamiliar with the iterative conception of set, it is roughly the idea that the sets are build up in stages. The process starts with the empty set $\varnothing$, then forms all the sets that can be formed out of those, i.e. $\varnothing$ and $\{\varnothing\}$, then all sets that can be formed out of those, i.e. $\varnothing$, $\{\varnothing\}$, and $\{\varnothing,\{\varnothing\}\}$, and so on. (For more thorough discussion, see Boolos (1971, 1989).)

BGC is convenient for this discussion because it talks about classes explicitly, which will be useful when we return to Russell's paradox. However, everything I say could be expressed perfectly well if we chose a formal theory like ZFC that describes only the domain of sets. We can always, on the informal side, introduce (predicative) classes as the extensions of predicates of sets.

In calling this theory standard, I by no means want to suggest that either of its components is beyond controversy. The iterative conception of set is still a matter of philosophical investigation and debate. Whether or not it justifies all the axioms of BGC is a matter of dispute. The continuum problem looms large as a difficulty of both components, and it is a commonplace idea that the formal theory BGC itself may not strong enough for some purposes. Nonetheless, both components are standard in that they are to be found in introductory set theory texts, and they enjoy reasonably wide, if often qualified, endorsement. Let us take the standard theory for granted, for purposes of this discussion.

The standard theory provides the standard solution to Russell's paradox. (COMP) has been dropped, and the revised theory BGC is presumably consistent. In the Liar paradox case, we were able to re-run the paradox to create a problem for our proposed solution via partiality. The question that needs to be asked, given remarkable similarity between the Liar paradox and Russell's paradox, is if we can do the same for the standard solution to Russell's paradox. Can we re-run Russell's paradox to get a strengthened version of it that poses a problem for this revised theory of sets?

The answer is that we cannot. As we have already observed, the formal theory BGC is presumably consistent (though for the usual Gödelian reasons, a proof of this bound to be less than satisfying.) But unlike the case of P, which is also consistent, the formal theory BGC together with the philosophical component of the standard theory, give us a way to avoid the strengthened paradox.

To see why, let us first recall how the standard solution resolves Russell's paradox. According to the standard theory, the Russell class R is simply not a set. There is, according to the Bernays-Gödel theory, a proper class R, which we may think of as the extension of the predicate $\ulcorner \neg x \in x \urcorner$ where $\ulcorner x \urcorner$ ranges over sets. From the axiom of foundation, in fact, we know that R is coextensive with the class V of all sets. But R is a proper class; it is not a set. (Both the formal and informal sides of the theory confirm this.) Only sets are members of classes. Indeed, only set terms can occur on the left of the membership sign $\ulcorner \in \urcorner$, so we cannot even ask if $R \in R$ or $\neg R \in R$.

Pursuing the parallel between the paradoxes, we might attempt to reinstate a strengthened version of Russell's paradox by an analogous argument to the Strengthened Liar. We get the Strengthened Liar by noting that we still have the Liar sentence $\ulcorner\lambda\urcorner$, and asking what the theory in question tells us about its truth status. Of course, we still have the Russell predicate $\ulcorner\neg x \in x\urcorner$. In parallel with asking about the truth status of $\ulcorner\lambda\urcorner$, we might ask what falls in the extension of this predicate. In particular, we might ask if the object R falls in its extension. In the Liar case, we got the answer that $\ulcorner\lambda\urcorner$ is not true. Likewise, here we get the answer that R does not fall within the extension of the predicate $\ulcorner\neg x \in x\urcorner$. With the Liar, this led back to paradox, as we seemed to have reached exactly the conclusion $\ulcorner\neg Tr(\ulcorner\lambda\urcorner)\urcorner$.

But here the parallel ends. There is no such problem with R. There would have been, if the predicate $\ulcorner\neg x \in x\urcorner$ said that x does not fall in the extension of x. If so, we would be forced to conclude $R \in R$, leading to paradox. But that is not what the predicate says at all. Rather, it says that the set x is not a member of the set x. We know that R is not a set, whereas the extension of $\ulcorner\neg x \in x\urcorner$ is a collection of sets, so R is not among them. This does not produce any paradox. The invitation to conclude that as R does not fall within the extension, then it does after all, is simply a confusion of sets with classes, and of set membership with falling within a class. The extension of the Russell predicate is determined only by the facts about set membership.

There is thus no strengthened Russell's paradox for the standard theory. It would be an overstatement to say that there are no problems for the standard theory presented by this sort of reasoning. For instance, if we ask why the set/class boundary falls were it does, and so why the universal class is not a set, we run into some well-know questions. Opinions differ on how pressing these questions are.[10] But regardless, it is striking that they do not present us with a paradox at all, and certainly do not reinstate a version of Russell's paradox. We see that both components of the standard theory—formal and philosophical—work together to ensure that the standard solution to Russell's paradox is invulnerable to a strengthened version.

We have now seen a crucial difference between the responses to the Liar and Russell's paradox. The response to the Liar via partiality is vulnerable to the Strengthened Liar, while the standard solution to Russell's paradox is not vulnerable to a strengthened version. This difference emerges

[10] This sort of problem was originally pressed by Parsons (1974b). For a response, see Boolos (1998b).

in spite of the two paradoxes themselves being formally very similar, as we saw in Section (II). As the difference is not in the paradoxes, it must lie in the theories of sets and of truth we build in response to the paradoxes. This shows us something important about the theory of truth. The theory based given by pure minimalism, even modified to use the partial theory P, is instable in the face of the paradox; while the standard set theory is not. This shows us both that this theory of truth is not adequate, and that it is lacking something which the standard set theory has. Our task now is to find out what standard set theory has and pure minimalism lacks, and see if a minimalist approach to truth can provide it.

IV. Stability and Divisiveness

What is it about pure minimalism that makes it unstable in the face of the paradox, and what must a better theory of truth look like? We have seen that in spite of the paradoxes for sets and truth being remarkably similar in formal respects, the standard set theory is not so unstable. So, to see what form a viable theory of truth must take, we should begin by looking at what makes the standard set theory stable.

Recall the point from Section (I) that both naive set theory and the pure minimalist theory of truth are in an important sense not about anything. The standard set theory is entirely different. It is genuinely about something: sets and classes. This is so in two ways. First of all, the formal theory BGC itself makes claims about the extent and nature of the domain of sets: claims that are true specifically of that domain, and do not hold of other domains. It provides principles of nature, like extensionality and foundation, set existence principles like infinity and restricted comprehension, and some generation principles like powerset that show how sets are generated from other sets. The existence and generation principles work together to describe the extent of the domain of sets.

The iterative conception of set works with the formal theory, to help make clear what the intended interpretation of the theory is. This helps us to further understand the extent and nature of the sets. Together, the formal and informal components of the standard theory go some way towards describing the domains of sets and classes. Of course, they have some well-know failings. They do not by any means complete the task as they stand. But anyone who understands the two components of the standard theory can reasonably claim to understand something of what sets there are, and something of how they behave; understand well enough, at least, to understand something about the difference between sets and classes.

This is crucial to the stability of the standard solution to Russell's paradox. The formal and philosophical components of the theory come together to allow us to conclude that the Russell class R is not a set. Even if some aspects of the extent of the domain of sets remain unclear, both components clearly support this conclusion. Once the distinction between sets and classes is in place, and it is established that R is not a set, we can rely on this to decline the invitation to draw paradoxical conclusions. Once we see the difference between set membership and falling within a class, and understand the Russell predicate as a predicate of sets, the invitation may be seen clearly to be a gross mistake. We would like to make the same sort of reply to the Strengthened Liar. We would like to say that in coming to conclude the Liar sentence is not true after all, we make a similarly gross mistake. The question is what we need from the theory of truth to be able to do so.

In describing the domains of sets and classes, the standard theory of sets behaves as we expect of most theories. Most theories some way or another divide off their subject-matter from the rest of the world. It is the correctness of the description of the subject-matter provided by the theory that makes the theory true. The specification of the subject-matter can be done in part by the formal components of the theory, and in part by the informal or philosophical account of its intended interpretation. The correctness of both components are then determined by the subject-matter so specified. Let us call this feature feature divisiveness.[11]

Theories may be divisive in different ways and to different degrees. Perhaps the most striking case is the second-order theory of arithmetic. In this case, the formal theory itself fully determines its domain of application, by being categorical. Few theories live up to this rather demanding standard. Our standard set theory certainly does not. But together, the formal component of the standard theory—BGC—and the philosophical explanation of its intended interpretation—the iterative conception—do provide a substantial account of the domains of sets and classes, as we

[11] A number of people have pointed out to me that 'divisive' may carry connotations which make it an unfortunate choice of terms. An anonymous referee suggested 'discriminate' instead. However, to keep my terminology the same as that of "Minimalism and Paradoxes," I am leaving it unchanged.

have observed. This is enough to at least partially specify the domain the theory is about, and to which it is responsible for its correctness. Perhaps most importantly for our purposes, the standard theory is divisive enough to draw some basic conclusions about what does not fall within the domain of sets. In particular, it makes a clear distinction between sets and classes, which enables us to conclude that the Russell class is not a set. This makes the theory divisive enough to be stable in the face of the paradox.

We expect empirical theories to be divisive to roughly this degree as well. An example much like the case of set theory is to be had from quantum mechanics. My friends in physics assure me that the domain of application of this theory is phenomena of very small scale. Just what is small scale is explained in part by the more informal gloss given to the theory, but also in part by the value of Planck's constant. More generally, it is no surprise that any decent theory should describe whatever it is about well enough to give some indication of what that domain is. Such an indication had better enable us to conclude, at least in some of the most basic cases, that something is not in the domain. For the most part, any good theory should be divisive.[12]

Both naive set theory and pure minimalism are notable for being as non-divisive as can be. Both are so by design: it is a reflection of philosophical commitments of both. This is a consequence of the point of Section (I) that neither theory is properly about anything. As we saw there, neither theory describes any particular domain, to which it would be responsible for its truth. As I noted in Section (I), the range of instances of (T) or (INF) is limited by syntax; but not because that is the limit of the domain these principles describe, but rather only because that is the limit of what can be written down. Pure minimalism still provides no real divisive content. It has no principles that reveal the nature of truth or the things to which truth applies. It thus has no principles that

[12] In describing a theory as divisive if it divides off its subject-matter from the rest of the world, I do not mean to require that a divisive theory must only apply to a proper subdomain of objects, or in the case of a physical theory, a proper subdomain of physical objects. We should construe the theory's subject-matter broadly, to include not only the objects to which the theory applies, but also the properties of them it describes. (In Quinean jargon, we should consider both ontology and ideology.) Some of the important examples for this paper, including standard set theory, arithmetic, and the divisive theories of truth I shall consider below, are divisive in part by applying to particular proper subdomains of objects. I do not know if this is so for quantum mechanics, but it need not be for the theory to be divisive. It would be enough for the theory to apply to all physical objects, but to describe only their small-scale properties. To take a less difficult example, suppose that Newtonian mechanics had applied to all physical objects. My friends in philosophy of science tell me that even if it had, it would still be a theory that describes the motions of physical objects (when speeds are not too great). It would thus still be reasonably divisive.

explain the nature of truth bearers and demarcate their domain. It cannot, for pure minimalism holds there is no such thing as a nature of truth or truth bearers!

Revising the formal theory by replacing M with P still leaves pure minimalism entirely non-divisive. Consider what P tells us about the domain of truths or truth bearers. It does prove some facts about truth, for instance, $P \vdash Tr(\ulcorner 1+1=2 \urcorner)$ (assuming that the theory is based on arithmetic). It even makes some existence claims, as we know $P \vdash \exists x Tr(x)$. But the theory is still as minimally divisive as can be. When the theory does prove something about the extent or nature of truth, it is only because something else having nothing to do with truth—nothing to do with its subject-matter—does most of the work. Once something else about the theory proves, for instance, $\lceil 1+1=2 \rceil$, then the theory is able to deduce $\lceil Tr(\ulcorner 1+1=2 \urcorner) \rceil$. From there, it can perform an existential generalization to get $\lceil \exists x Tr(x) \rceil$. But the principles governing truth that are the heart of the theory only play a role in deducing these facts in the step from $\lceil 1+1=2 \rceil$ to $\lceil Tr(\ulcorner 1+1=2 \urcorner) \rceil$. The rest is a completely independent matter of arithmetic or logic. The theory can determine that something is the case, and then add that it is a truth, and then extract some logical consequences from this fact. But it cannot say anything about truths, the purported objects the theory is describing, more directly. None of the principles of truth in P by themselves make any substantial claims about truths in general, but only relate truth to specific sentences whose correctness has been independently decided. P states no general principles which can help us to understand the extent of the domain of objects to which truth applies. The theory P is thus only divisive where something unrelated to truth makes it so. This is just as the philosophical principles of pure minimalism would have it. Both the formal and informal components of pure minimalism make it as non-divisive as can be, and modifying the formal theory to incorporate partiality does nothing to change this.

To further our comparison with set theory, imagine a theory of sets more like our partial theory of truth P. It would have some principles which, once you concluded something else, could be used to conclude that some set exists, or has certain members. As a result, we could use the theory to generate a list of specific statements that say that something is a member of something else ($\lceil a \in b \rceil$), or not a member of something else ($\lceil a \notin b \rceil$). But the theory could make only trivial generalizations about membership or set existence, such as those that followed from elements of the list by logic (by analogy with $P \vdash \exists x Tr(x)$). Unlike BGC, it could tell us nothing more substantial about the extent and nature of the list. And unlike what the standard theory says about the Russell class, it could not tell us anything about why certain sentences could not be on the list.

In contrast to the standard theory, this makes the theory we are now imagining not divisive enough to enable us to reply to the attempt at a strengthened Russell's paradox. If we found a pair of objects a and b such that we determined somehow that the theory could not have on the list $\ulcorner a \in b \urcorner$, we would be able to conclude only that, as far as the theory tell us, $a \notin b$. We would not be able to draw any more subtle conclusions. Crucially, we would not be able to draw the conclusion that $\ulcorner a \in b \urcorner$ is not on the list because b is outside of the range of objects the theory is attempting to describe, or because a is the kind of object that cannot be a member of anything. Hence, if were were to observe that $\ulcorner R \in R \urcorner$ cannot be on the list, we would indeed fall into a strengthened Russell's paradox. We would have to conclude that as far as the theory tells us, $R \notin R$, from which the paradox follows.

The theory we are now imagining lacks the resources to make the reply of the standard solution to Russell's paradox. It cannot say that we do not have $\ulcorner R \in R \urcorner$ (it is not 'on the list') because R is a proper class, and so cannot enter into membership relations with any set or class. It cannot make this reply because it fails to draw a stable distinction between sets and classes. This failure in turn derives from its failure to be sufficiently divisive. A stable distinction between sets and classes could only follow from a sufficiently divisive specification of the theory's subject-matter, as we have with the standard theory. Without this much divisiveness, as we have seen, we are simply left with the paradoxical conclusion.

A theory of sets that fails to be divisive in this way cannot avoid the strengthened paradox. Pure minimalism likewise fails to be divisive, and so cannot avoid the Strengthened Liar. When we encounter $\ulcorner \lambda \urcorner$, pure minimalism, even modified by P, allows us nothing to say except that according to P, $\ulcorner \lambda \urcorner$ is not true. As pure minimalism fails to be divisive, we cannot go on to make any substantial claim about why this is so. We can cannot observe that it is so because $\ulcorner \lambda \urcorner$ falls outside the domain of the theory, or falls under a special category within the theory, as we say about the Russell class on the standard set theory. Thus, the non-divisiveness of pure minimalism leaves it with a paradoxical conclusion as well.

In Section (I), I pointed out that pure minimalism, like naive set theory, is in a sense not about anything. We have seen in this section that this makes pure minimalism highly non-divisive. Following up on the discussion of paradoxes in Sections (II) and (III), we have seen that failing to be reasonably divisive renders pure minimalism unable to respond to the Strengthened Liar, even when modified to use a partial theory like $\boldsymbol{P}$. We have also seen that a sufficiently divisive theory, like the standard set theory, easily dismisses the attempt at a strengthened Russell's paradox, even though the two paradoxes are themselves virtually alike. The moral is that to have any prayer of avoiding the Strengthened Liar, we must look for a more divisive theory of truth. In the next section, we will consider what such a theory of truth might look like.

V. Divisive Theories of Truth

A viable theory of truth must be more divisive than pure minimalism. It must be more like the standard theory of sets in describing some specific domain. It must be about something! What would such a more divisive theory of truth look like? In this section, I shall offer some reasons to think that any more divisive theory of truth should be expected to be highly non-minimalist, or more generally non-deflationist. It will look much more like a correspondence-based theory.

As is often pointed out, it is too much to ask of a theory to characterize the domain of all truths. This would be impossibly demanding, as it would make the theory the complete theory of absolutely everything; containing all sorts of facts about all sorts of subjects.[13] But it is reasonable to ask the theory to be divisive about truth bearers. This could make the theory divisive about truth in the right way, as delineating a domain of truth bearers appropriately delineates the range of application of the truth predicate. If we can delineate its range of application properly, we might be able to offer a stable reason for declining to apply the truth predicate to the Liar sentence, which

[13] There is a significant question of whether there is really a coherent notion of absolutely all truths. Some reasons to be skeptical may be found in Grim (1991). Related issues are discussed in Parsons (1974a) and my (forthcominga). However, my worry here is much more pedestrian. It is already too much to ask of the theory of truth to contain all our current knowledge, whether or not a single complete theory of absolutely everything makes sense.

might lead to a stable solution to the Liar.

To describe what such a divisive theory might look like, we should return once more to the debate over deflationism. Deflationism is often contrasted with the idea of a correspondence theory of truth. Outside of its classical form, such as in works of Russell, it is notoriously difficult to state clearly what the correspondence theory of truth is. Nonetheless, there is a core idea behind talk of correspondence which points towards more divisive theories.

The core idea is that truth bearers are representational. They describe the world as being some way, and are true if the worlds is that way. There is some leeway in just how we characterize truth bearers along these lines. We might say that truth bearers are propositions, where propositions are objects that encapsulate collections of truth conditions (to borrow a phrase from Hartry Field). Truth then obtains when the actual circumstance is among a proposition's truth conditions.

Propositions themselves are not really crucial to this idea. We could just as well say that truth bearers are interpreted sentences, where the interpretations provide truth values for sentences (in contexts, where appropriate). An extensional approach might stop there, while a more intensional approach might provide richer semantic contents for sentences. A similar theory could be given based on utterances rather than sentences.

Whether we opt for propositions or not, some idea of correspondence is usually built into these sorts of representational approaches. It comes up naturally in explaining how a sentence expresses a proposition, is given an interpretation, or is otherwise representational. In many cases, such explanations proceed by identifying the extensions—the referents—of components of a sentence (or intensions, built up from the referents). This accounts for representation in terms of 'word-to-world' relations, and how the worldly referents themselves interact. This is in essence a correspondence idea (though one that can avoid explicit commitments to facts).

More importantly, it is a highly non-minimalist idea. So long as being representational is itself construed as a substantial property, as it is on the correspondence-based approach, it implicates substantial properties of being a truth bearer and of truth. On this sort of approach, being a truth bearer is determined by whether a sentence expresses a proposition or is otherwise a representation, which is determined by substantial facts about reference. Likewise, whether or not a truth bearer is true is fixed by these facts. Such a representational account of truth and truth bearers embodies enough of the correspondence idea to stand in opposition not only to minimalism, but to most any form of deflationism.[14]

A theory along these lines could wind up being sufficiently divisive. An account of how truth bearers are representational could provide a suitably divisive picture of the domain of truth bearers, so long as it is able to make a principled distinction between sentences that genuinely provide truth bearers—express propositions or are otherwise representational—and sentences that may look like they provide truth bearers but do not. From this, we could distinguish two ways of failing to be true. One is to be a truth bearer that is not true, and the other is to fail to be a truth bearer at all. The partial theory P attempted to implement a distinction like this, by making the truth predicate partial. But in lacking a divisive theory of truth bearers, it failed to do so in way that is sufficiently robust to resist the Strengthened Liar.

If we had a sufficiently robust, sufficiently divisive theory of truth bearers, we could genuinely begin to resist the Strengthened Liar. We could begin to do what pure minimalism could not. This claim needs to be made with some care. It is well-known that appealing to propositions or truth-value gaps does not by itself suffice to solve the Strengthened Liar. My point is rather that a divisive theory of truth bearers is required to even begin to make progress towards addressing it. If we had such a theory, we could go this far in responding to the Strengthened Liar: when we conclude that $\ulcorner\lambda\urcorner$ is not true, we could come to two different conclusions. Either $\ulcorner\lambda\urcorner$ is (or expresses) a truth bearer, but is not true, or $\ulcorner\lambda\urcorner$ is not (or does not express) a truth bearer. The latter is very much analogous to concluding that the Russell class R is not a set. If we could reach this conclusion, we could resist the invitation to infer that we have concluded λ, just as when we say that R is not in the extension of the Russell predicate, we have not concluded $R \in R$. It was in not providing a principled way to draw this sort of distinction that P and pure minimalism left us no where to turn to avoid the Strengthened Liar.

Drawing a stable distinction between different ways of failing to be true is the first step in

[14] I should note that in asking for a theory which is divisive about truth bearers, I am not raising the classic question of what the primary bearers of truth are. Rather, the issue here is one of distinguishing truth bearers, be they primary or otherwise, from non-truth bearers in a sufficiently divisive way.

solving the Strengthened Liar. It is not the last. Crucially, we still need to explain what sense is to be made of the conclusion that $\ulcorner \lambda \urcorner$ is not true, even if this is because it is not a truth bearer. Many have taken this to require that we invoke a hierarchy of truth predicates. My own preference is for an approach relying more heavily on ideas about context dependence. I shall not advocate for any particular approach here.[15] I only claim that we need our theory of truth to draw the distinction between different ways of not being true to proceed at all. Drawing it requires being divisive about the domain of truth bearers, in just the way the standard theory is about sets, and in just the way that pure minimalism is not. Just as having some set/class distinction is not by itself enough to solve Russell's paradox, having some characterization of the domain of truth bearers is not by itself enough to solve the Strengthened Liar; but it is a necessary precondition.

I have argued that the required divisiveness might be found in the idea of truth bearers being representational, which forms the core of a correspondence-based theory of truth. Unlike the minimalist approach, this view maintains that there is some underlying nature to the property of truth, to be found in the ideas of representation and the world fitting a representation. If this yields substantial principles about what makes a well-formed sentence or utterance a genuine truth bearer, it could provide just the divisiveness we need to begin building a stable response to the Liar. Of course, much more needs to be said to articulate the ideas of representation or correspondence. There have been a great many attempts to do so over the years. Rather than rehearse them here, I shall restrict myself to a much more modest goal. I shall conclude this section by pointing out that many of the leading ways of dealing with the Strengthened Liar rely on these ideas.

Most approaches to the Strengthened Liar ultimately rely on some sort of hierarchy. To see how this relates to correspondence and representation, we should start with the idea of grounding developed by Kripke (1975).[16] Sentences are divided into two classes: grounded and ungrounded. Grounded sentences are naturally assigned the values true or false, while ungrounded ones are not. An account of grounding can thus be used to provides a divisive account of truth bearers. The

[15] I have discussed my own preferred approach to the Strengthened Liar in my (2001; forthcomingb; MS).

[16] The term first appears in Herzberger (1970), which also makes some comparisons with paradoxes in set theory. Kripke's work provides an extensive development of the idea. Technically, in Kripke's framework, grounded sentences are those that are true or false in the least fixed point, and the informal sketch I give below echoes this idea. However, it would not change the basic points I make about grounding if one were to offer some reason for starting with a more extensive assignment of truth values and generating a larger fixed point.

account Kripke gives of grounding is, informally speaking, one of starting by describing the world in non-semantic terms, and then building up successively more complex descriptions involving semantic expressions. This process reaches a fixed point, which circumscribes the domain of grounded sentences. This notion of grounding is naturally taken to provide just the sort of account of representation to which I alluded a moment ago. We start with the idea of reference for non-semantic terms, and on the basis of it assign truth values to sentences involving only these terms. We then progressively build up truth assignments for sentences containing semantic terms (namely the truth predicate). As is well-known, the presence of self-reference and other semantic complexity makes this process transfinite, and it assigns truth values to some but not all sentences. Those that are assigned a truth value—either true or false—at the end of the process are the grounded ones. The picture that emerges is one of truth bearers being those sentences that ultimately describe the world through this iterative process. We may well say that these are the sentences that express propositions, or are representational. They are the truth bearers.

It is well-known that in the face of the Strengthened Liar, Kripke ultimately appeals to "the ghost of the Tarski hierarchy" (1975, p. 80). This is a hierarchy based on grounding rather than on syntax. At the first level is the truth predicate produced by the process of generating grounded sentences just described. From the first level, we can ascend to the next level in the hierarchy by reflecting on the entire process, and noting that on the basis of it certain sentences are not true. On this view, the basis for the hierarchy is the idea of grounding. Grounding generates the first-level truths and falsehoods, and then provides the material for the construction of the next level through some sort of reflection on the process, and so on.[17]

With this in mind, we should consider the more traditional hierarchical approach stemming from Tarski (1935), which imposes a hierarchy of indexed truth predicates and syntactic restrictions on how they may appear in a sentence. We can see the syntactic requirements Tarski imposes as requiring an explicit syntactic representation of much the same kind of grounding process as Kripke describes (though in Tarski's work, lacking the transfinite and level-merging aspects). Hence, we can

[17] The idea of such reflection is fundamental to the approach of Parsons (1974a). Both Parsons and Burge (1979) pursue these sorts of ideas with much more explicit attention to the role of context. I have discussed the nature of this sort of reflection in my (forthcomingb).

see Tarski's approach as based on the same kinds of ideas about representation or correspondence as Kripke's. It is hotly debated just how much of a deflationist Tarski is, and I do not want to take a stand on Tarski interpretation here so much as to point out how a natural understanding of his theory draws on correspondence ideas. Moreover, it is important to note that irrespective of this, his theory is divisive in one vital sense. Surface well-formedness is not sufficient to make a sentence a truth bearer. A sentence must also be, we might say, well-indexed: its truth predicates must be indexed so as to make it a genuine sentence of the Tarski hierarchy of languages. Insofar as a characterization of this property must be part of the full theory of truth, we have a divisive principle. I have merely suggested that one source of such a principle is in correspondence-based ideas.

One other important example is the theory of Barwise and Etchemendy (1987). Their work is based on a situation theory which is explicitly modeled on ideas about the correspondence theory of truth as seen by Austin (1950).

I am not here advocating a particular solution to the Strengthened Liar; rather, I am claiming that a theory which can provide one must be divisive about truth bearers. Moreover, I claim, the general outlook of a representational approach to truth—embodying some aspects of the idea of correspondence—provides a basis for the development of such a divisive theory. We have seen how some of the leading theories that have been developed to solve the Strengthened Liar rely on this outlook.

We have now seen that pure minimalism is untenable, as it is unstable in the face of the Strengthened Liar. We have also seen why. In attempting to not be about any subject-matter, it is not sufficiently divisive to provide an adequate response to the paradox. We have also seen that theories which avoid strengthened paradoxes are divisive, both in the case of sets and of truth. Finally, we have seen that theories of truth based on ideas of representation can be sufficiently divisive to avoid the Strengthened Liar. These, we have seen, are fundamentally correspondence-based theories, and so are diametrically opposed to minimalism.

VI. Minimalism and Deflationism

I have stressed the parallel between pure minimalism and naive set theory on the one hand, and the divisiveness of standard set theory and a theory of truth able to address the Strengthened Liar on the other. In the face the paradoxes, a viable theory of truth, like a viable theory of sets, must be sufficiently divisive. I have noted that the general idea of the correspondence approach to truth provides a basis for such divisiveness. This gives us good reason to expect a viable theory of truth to be non-minimalist, and more generally non-deflationist. But I have yet to consider explicitly whether any significant departures from pure minimalism could be sufficiently divisive, without giving up on their deflationist ideals.

In this section, I shall argue they cannot, in two steps. First, I shall argue that minimalist approaches cannot be sufficiently divisive, even if they depart from pure minimalism in some respects. The second step will be more tentative. In it, I shall turn my attention to whether any other sort of deflationism could be divisive enough. I shall there express some skepticism about how any philosophically rich deflationism could achieve the needed divisiveness.

Let us quickly review the problem for pure minimalism. It is clear from the arguments above that pure minimalism cannot yield a theory which is divisive on truth bearers. Philosophically, the position is intended to be as non-divisive as can be. Pure minimalism is, by its nature, not a theory about any domain, so it cannot be divisive. Formally, we have seen that moving from $\boldsymbol{M}$ to $\boldsymbol{P}$ does not help matters. This is striking, as $\boldsymbol{P}$ was designed to avoid the Liar. Its failure shows that just changing the formal theory does not make the view divisive. We can safely conclude that no change to the formal component compatible with the philosophical side of pure minimalism will make it divisive.

In Section (I) I introduced pure minimalism as a simplified version of minimalism, useful for discussion purposes. We thus must consider whether any other other minimalist positions might achieve the needed divisiveness, while staying faithful to the core ideals of minimalism. I shall address this question in two parts, and after them turn to the remaining task of discussing deflationism more generally.

First, there are a number of positions that depart from pure minimalism in ways that do not affect the issue of divisiveness about truth bearers. Some views that hold to the basic principles of minimalism take the truth predicate to apply to propositions. However, these views remain minimalist by saying that for each well-formed declarative sentences, there is a proposition expressed by it. There is no more an underlying fact about what makes a sentence express a proposition than there is an underlying nature of truth. Hence, these theories are no more divisive than pure minimalism on truth bearers.[18]

Second, we need to consider views that depart from pure minimalism in more significant ways, yet remain in some sense minimalist. An important example is a view derived from Wright (1992). Wright proposes a criterion for the truth aptness of sentences based on "surface constraints of syntax and discipline" (Wright, 1992, p. 35). Truth bearers are then the truth-apt sentences. The basic idea of "discipline" is that of norms of use which are typical of assertion, including the appropriateness of embedding in antecedents of conditionals, for instance. This is a significant departure from pure minimalism. As it is a departure precisely on the issue of truth aptness, it might allow for a minimalist theory to be a little more divisive than pure minimalism.[19]

However, I do not see how anything along the lines of surface constraints of syntax and norms of use can provide a criterion that will help in the face of the Liar. The problem there is precisely that we do seem to have a sentence that meets the constrains of syntax and of discipline as well, but one that cannot be a truth bearer. Compare this once again with the set theory case. Proper classes bear all the surface marks of sets. They have members, are extensional, and so on. By these lights, the Russell class is as good a set as any other, and hence the paradox. What avoids the paradox is a much more substantial ontological distinction, between set-like objects that are sets, and those that are not. In the truth case, a criterion like Wright's still seeks to make anything that looks like a truth bearer be a truth bearer. This is the force of the constraints of syntax and discipline being surface constraints. The addition of discipline refines the notion of looking like a truth bearer, but

[18] This is quite close to the view held by Ayer (1946). It may be the view held by Horwich (1990). There, Horwich provides a use-based account of proposition-individuation, but says little about the conditions for a sentence to express a proposition. As I mentioned, remarks in Horwich (1994) suggest a minimalist approach to this issue, which would make his view fall within the scope of the objection pressed here. (See Field (1992) for some further discussion.)

[19] Wright's use of the term 'minimalism' is slightly different from mine, making the question of whether to attribute a form of minimalism in my sense to him somewhat delicate. However, the idea of discipline is a natural one for a minimalist in my sense to appropriate, whether the result is Wright's own position or not.

the problem with the Liar is that we have something that does look like a truth bearer by any of these standards, and we need a theory that provides a stable answer that nonetheless it is not one. We need a more divisive theory than the combination of syntax and discipline can give.

Thus, a departure from pure minimalism along Wright's lines fails to be sufficiently divisive. But more importantly, the reason it fails shows us a much more general reason why anything that counts as minimalism will fail to be sufficiently divisive. A minimalist approach, however refined, is committed to there being no 'underlying nature' or 'substantial property' of truth. As a result, any minimalist approach is committed to there being nothing but overt surface properties that determine whether something is a truth bearer or not. Anything else would eo ipso provide an 'underlying nature' of truth, as it would provide some underlying facts about what counted as a truth bearer and what did not. This would make truth 'substantial' in just the way the minimalist says it is not. The Liar is such a problem in part because the Liar sentence really does appear on its face to be a perfectly good truth bearer. It meets all the overt or surface criteria for being one. Because of this, minimalist solutions to the Liar tend to be unstable in the face of the Strengthened Liar. When we try to make the Liar sentence not a truth bearer, we then find reasons to reinstate it as a truth bearer after all. Overt or surface criteria give us no reason to which we might appeal to reject this conclusion. Hence, the kind of divisiveness we need to begin to address the Strengthened Liar is precisely one that is not based on overt surface features. This is not something any brand of minimalism can provide.

Before leaving minimalism, we should pause to discuss briefly the argument of Jackson et al. (1994) that minimalism about truth does not by itself lead to minimalism about truth aptness. First of all, we are already in a position to see that simply avoiding commitments about truth bearers or truth aptness cannot save minimalism. As we saw in Section (III), a theory like P is already formally silent on truth bearers, in that it has sentences for which it proves neither $\ulcorner Tr(\ulcorner\phi\urcorner)\urcorner$ nor $\ulcorner\neg Tr(\ulcorner\phi\urcorner)\urcorner$, yet it is vulnerable to the Strengthened Liar. It is vulnerable because it shows how to predicate truth of some sentences, and then shows that the Liar sentence is not among them. It lacks an explanation of why this is not simply the conclusion $\ulcorner\neg Tr(\ulcorner\lambda\urcorner)\urcorner$, and this in turn leads to paradox. When it comes to the puzzle of the Liar, having a theory that is simply silent on truth bearers does not give a sufficiently divisive theory. To the contrary, what we need is a theory that gives us a principled distinction between truth bearers and things that look on surface like truth bearers but are not, to which we can appeal in responding to the Strengthened Liar. If we somehow excise commitments on truth aptness from minimalism, we get no such thing. Instead, we get a theory which makes no substantial claims about truth bearers, but still allows the drawing of conclusions about applications of the truth predicate. This allows for the Strengthened Liar, and does not give us anything like the resources needed to resolve it.

Thus, the arguments of Jackson et al. (1994) cannot provide a safe haven for minimalism. They note that an instance of the T-schema, such as ' 'torture is wrong' is true iff torture is wrong', only confers truth conditions if the right-hand side of the biconditional is itself truth apt. This claim appears to be entirely correct, but it does not help with the issue at hand. At best, it provides a way for a minimalist to be silent about issues of truth aptness. As we have seen, this will not suffice.[20]

I conclude that no minimalism can be divisive enough to respond to the Strengthened Liar, and so no minimalism is tenable. This now brings us to the second issue of this section: the question of whether some other sort of deflationism about truth, somehow different from minimalism in any of its forms, could evade the problem I have raised. The vague nature of the category of deflationism makes this question rather hard to answer, but I shall offer some reasons why the problem is very difficult for deflationist positions to overcome.

As far as I can tell, the only deflationist positions which might find the needed divisiveness are those that, unlike minimalism, do include a substantial and divisive theory of propositions or of content, but still remain committed to a strong form of (T) and hold it to be analytic or necessary. The commitment to the necessity of (T) is a mark of deflationism, but nonetheless it is not easy to

[20] Jackson et al. (1994) go on to offer a more substantial account of truth aptness. As I shall discuss below in Footnote (22), I do not see how this winds up leaving room for minimalism at all.

It should be stressed that the notion of truth aptness at issue here is the fairly weak one of application of the truth predicate making sense. So long as application of the truth predicate is governed by enough discipline to make overt contradictions repugnant, we have to deal with paradoxes that may arise for it. This is the case whether or not the assertions in question wind up being truth apt in the sense relevant to, say, non-cognitivism. Thus, I believe that the issues primarily under consideration by Jackson et al., as well as those debated by Boghossian (1990), Wright (1992), and Soames (1999), are somewhat different than those being investigated here. (Surely no one ever really thought non-cognitivism could solve the Liar!)

formulate such a position in a way that remains deflationist. Ramsey (1927) is sometimes offered as an example, but I am inclined to side with Field (1986) in saying that Ramsey is not a genuine deflationist. Ramsey interpretation aside, the point is that any view that has a substantial account of propositions as encapsulating truth conditions can certainly have a strong version of (T), but is no more or less deflationary than its account of truth conditions. Non-deflationism about truth conditions is non-deflationism.[21] The only way to pursue this line as a deflationist seems to be to offer a non-truth-conditional account of propositions or content, and yet hold that propositions (or contentful sentences) are truth bearers.[22]

This is close to the view held by Field (1986; 1994). Field certainly would not opt for propositions, but he employs a robust notion of content based on conceptual role. This is coupled with a deflationary account of truth, in which the truth predicate applies to sentences that the speaker understands, and thus which have content. The relation between the two sides of (T) is one of 'cognitive equivalence', which is, presumably, at least as strong as the necessity or analyticity minimalism proposes.[23] As the notion of content to which Field appeals is itself not at all minimal, there is room for a view like this to be divisive about truth bearers. It will be, so long as it yields a substantial criterion for whether a sentence is contentful, and so can stand as a truth bearer.

Though a theory like Field's can be divisive about truth bearers, it is not at all clear to me how it can be sufficiently divisive to respond to the Strengthened Liar. It remains unclear how the theory can really explain why the Liar sentence is not one we understand, and so is not a truth bearer. The same kinds of points as I raised above about syntax and discipline apply here. The Liar looks like a well-behaved sentence. We understand each of the terms in it, and we can use

[21] As was pointed out in Parsons (1974a), any view that contains a theory of propositions as truth conditions leads to a strong version (T). On such a view, the truth predicate is just the operation of evaluating an intension on the actual world. In the notation of intensional logic, we have something like $\ulcorner Tr(p) \leftrightarrow {}^{\vee}p \urcorner$. Coupled with the principle of intensional logic $\ulcorner {}^{\vee\wedge}\phi \leftrightarrow \phi \urcorner$, we get the T-schema in the form $\ulcorner Tr({}^{\vee\wedge}\phi) \leftrightarrow \phi \urcorner$. This has the status of a truth of mathematics, if not logic. The truth predicate is basically the operation of function evaluation. Even so, it is not clear that we should conclude that it is entirely trivial. Having self-applicative truth puts us in an untyped world, and experience with, say, the untyped λ-calculus shows that function application in such a setting is far from trivial.

[22] I suspect that Jackson et al. (1994) miss this point. They consider a view of truth aptness based in part on whether a sentence expresses the content of a belief, while belief states are said to be "designed to fit the way things are" (p. 297). This appears already to brings with it a correspondence-like notion of truth for what sentences express, and so is not anything a deflationist can accept.

[23] Field also considers some departures from the requirement that the truth predicate only apply to sentences the speaker understands, especially in the postscript to Field (1994) provided in Field (2001).

it in apparently well-structured arguments. Hence, it may well appear that we should count the Liar sentence as one we understand. (As it might be put, which part of $\ulcorner \neg Tr(\ulcorner \lambda \urcorner) \urcorner$ don't you understand?) Moreover, when we consider the Strengthened Liar, we are confronted with perfectly good cases in which we do seem to understand it. Yet if we have good reason to count the Liar sentence as one we do understand, it counts as a contentful sentence—as a truth bearer. If so, then the theory is not divisive enough after all.

This is not the end of the matter. Clearly Field or others may have more to say.[24] So let me simply express skepticism about how a theory along these lines can be adequately divisive. Finally, let me note that the challenge I have raised to this sort of theory is a general one to deflationism. The most likely way to provide a divisive yet deflationist theory, I have suggested is to provide a theory which invokes a non-truth-conditional notion of proposition or of content. As the response to Field shows, being sufficiently divisive on truth bearers is not simply a matter of having something to say about content or propositions. It must be enough to explain what is problematic about sentences like the Liar. It is hard to see how this could be done without returning to the correspondence-based ideas I canvassed in Section (V). At the very least, it is hard to see how this could be done without establishing a link between propositions, divisively and non-truth-conditionally described, and representational or otherwise non-deflationist aspects of truth bearers. Can this be done without converting the apparently non-truth-conditional account of propositions into a substantial account of truth and truth conditions? More generally, can it be done while retaining deflationism? I am inclined to doubt it can.[25]

[24] Field has in fact said more. At some points, he has asked if the cognitive equivalence of $\ulcorner Tr(\ulcorner \phi \urcorner) \urcorner$ and $\ulcorner \phi \urcorner$ really leads to (T) (cf. Field, 2001). More recently (2002), he has considered the possibility of changing the logic of $\ulcorner \rightarrow \urcorner$ to allow for the truth of all instances (T), including the paradoxical ones, while retaining consistency. This might keep the link between cognitive equivalence and (T). I remain skeptical, for several reasons. First of all, I remain somewhat skeptical of this sort of (fairly radical) revision of logic, especially if what we are offered is supposed to be metaphysically deflationary. Moreover, I remain uncertain if the kind of strengthened paradoxes that have been my focus here may be reintroduced into Field's theory, especially if they make use of meta-level reasoning that bypasses the special logic Field imposes on $\ulcorner \rightarrow \urcorner$. This raises a great many issues, too many to be discussed here. So let me just cautiously register my skepticism.

[25] One view that is probably not open to the objection I have raised per se is that sketched in Soames (1984). The idea there is to construe truth as a property of abstract interpreted languages. Though he does not pursue the matter, much of what I have said about how a theory could be made sufficiently divisive could easily be carried over to this setting. As I understand it, the view is offered as deflationist in that truth is construed as not a metaphysically important notion, but more a piece of mathematics. (Soames (1999) takes a similar stance towards deflationism, and discusses the Liar explicitly. In response to the Strengthened Liar, he there appeals to a "Tarski-like hierarchy" (p. 181).) As Soames is well aware, this is very far indeed from minimalism, and even from most other positions that

Deflationism is a wide and vaguely defined area, so such sweeping conclusions should be drawn with care. At least, I think it is a fair challenge to deflationists, especially minimalists or those of Field's variety, to ask how sufficient divisiveness can be achieved.

References

Austin, J. L., 1950. Truth. Aristotelian Society Supplementary Volume 24:111–129. Reprinted in Austin (1961).

———, 1961. Philosophical Papers. Oxford: Oxford University Press. Edited by J. O. Urmson and G. J. Warnock.

Ayer, A. J., 1946. Language, Truth and Logic. 2nd edn. London: Victor Gollancz.

Barwise, J. and J. Etchemendy, 1987. The Liar. Oxford: Oxford University Press.

Boghossian, P. A., 1990. The status of content. Philosophical Review 99:157–184.

Boolos, G., 1971. The iterative conception of set. Journal of Philosophy 68:215–232. Reprinted in Boolos (1998a).

———, 1989. Iteration again. Philosophical Topics 42:5–21. Reprinted in Boolos (1998a).

———, 1993. The Logic of Provability. Cambridge: Cambridge University Press.

———, 1998a. Logic, Logic, and Logic. Cambridge: Harvard University Press.

———, 1998b. Reply to Charles Parsons' "Sets and Classes". In Logic, Logic, and Logic, pp. 30–36. Cambridge: Harvard University Press.

Burge, T., 1979. Semantical paradox. Journal of Philosophy 76:169–198. Reprinted in Martin (1984).

Feferman, S., 1984. Toward useful type-free theories, I. Journal of Symbolic Logic 49:75–111. Reprinted in Martin (1984).

offer themselves as deflationist. I see no reason to argue over who gets the term 'deflationist', but I believe Soames' position is significantly different from the kinds under discussion here.

———, 1991. Reflecting on incompleteness. Journal of Symbolic Logic 56:1–49.

Field, H., 1986. The deflationary conception of truth. In G. MacDonald and C. Wright, eds., Fact, Science and Morality, pp. 55–117. Oxford: Blackwell.

———, 1992. Critical notice: Paul Horwich's Truth. Philosophy of Science 59:321–330.

———, 1994. Deflationist views of meaning and content. Mind 103:249–284. Reprinted in Field (2001).

———, 2001. Truth and the Absence of Fact. Oxford: Oxford University Press.

———, 2002. Saving the truth schema from paradox. Journal of Philosophical Logic 31:1–27.

Friedman, H. and M. Sheard, 1987. An axiomatic approach to self-referential truth. Annals of Pure and Applied Logic 33:1–21.

Glanzberg, M., 2001. The Liar in context. Philosophical Studies 103:217–251.

———, forthcoming a. Quantification and realism. Philosophy and Phenomenological Research.

———, forthcoming b. Truth, reflection, and hierarchies. Synthese.

———, MS. A contextual-hierarchical approach to truth and the Liar paradox. Manuscript.

Grim, P., 1991. The Incomplete Universe. Cambridge: MIT Press.

Herzberger, H. G., 1970. Paradoxes of grounding in semantics. Journal of Philosophy 67:145–167.

Horwich, P., 1990. Truth. Oxford: Basil Blackwell.

———, 1994. The essence of expressivism. Analysis 54:19–20.

Jackson, F., G. Oppy, and M. Smith, 1994. Minimalism and truth aptness. Mind 103:287–302.

Jech, T., 1978. Set Theory. New York: Academic Press.

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

Martin, R. L., ed., 1984. Recent Essays on Truth and the Liar Paradox. Oxford: Oxford University Press.

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

———, 1992. Maximal consistent sets of instances of Tarski's schema (T). Journal of Philosophical Logic 21:235–241.

Parsons, C., 1974a. The Liar paradox. Journal of Philosophical Logic 3:381–412. Reprinted in Parsons (1983).

———, 1974b. Sets and classes. Noûs 8:1–12. Reprinted in Parsons (1983).

———, 1983. Mathematics in Philosophy. Ithaca: Cornell University Press.

Quine, W. V., 1986. Philosophy of Logic. 2nd edn. Cambridge: Harvard University Press.

Ramsey, F. P., 1926. The foundations of mathematics. Proceedings of the London Mathematical Society, Second Series 25:338–384. Reprinted in Ramsey (1931).

———, 1927. Facts and propositions. Aristotelian Society Supplementary Volume 7:153–170. Reprinted in Ramsey (1931).

———, 1931. The Foundations of Mathematics and Other Logical Essays. London: Routledge and Kegan Paul.

Simmons, K., 1999. Deflationary truth and the Liar. Journal of Philosophical Logic 28:455–488.

Soames, S., 1984. What is a theory of truth? Journal of Philosophy 81:411–429.

———, 1999. Understanding Truth. Oxford: Oxford University Press.

Tarski, A., 1935. Der Wahrheitsbegriff in den formalizierten Sprachen. Studia Philosophica 1:261–405. Translation by J. H. Woodger as "The Concept of Truth in Formalized Languages" in Tarski (1983).

———, 1983. Logic, Semantics, Metamathematics. 2nd edn. Indianapolis: Hackett. Edited by J. Corcoran with translations by J. H. Woodger.

Wright, C., 1992. Truth and Objectivity. Cambridge: Harvard University press.