34 Conceptual Engineering and Replacements for Truth

Kevin Scharp

1 Conceptual Engineering

Conceptual engineering is an exciting new movement in analytic philosophy that focuses on how to evaluate our concepts and how to improve them. Philosophers have been doing this sort of thing for a long time, but the recent excitement is over thinking *explicitly* about how best to improve our concepts. Conceptual improvement has itself become a major topic in philosophy.

The idea that Western philosophers should be answering "what is" questions, like "What is knowledge?" and "What is virtue?" goes back to the ancient Greeks. Over 2,000 years later, *analytic* philosophy was founded on the idea that the best way to answer these kinds of questions is by providing conceptual analyses. A conceptual analysis of knowledge would be something like a definition of "knowledge" and similarly for "virtue." In addition to this dominant *descriptive* theme in Western philosophy, there have been those emphasizing the *prescriptive*—calling for changes in how we think or talk. However, it has only been in the last couple of decades that significant numbers of Western philosophers have turned their attention to conceptual engineering as an explicit philosophical methodology that is worthy of study in its own right.

Many conceptual engineers focus on meanings in addition to concepts, and there is considerable disagreement about the distinction between concepts and meanings. Some treat them more or less the same, while others are careful to distinguish them.[1] Usually, the term "conceptual engineering" is taken to cover both the projects focusing on concepts and those focusing on meanings. In what follows, I will take meanings to be semantic values studied by those doing natural language semantics, whereas concepts are discussed at length below. Because conceptual engineers are divided on whether it is acceptable to phrase their philosophical projects in terms of concepts, I intend to use the phrase "representational device" as a general term to cover concepts, meanings, words, or any other way of representing the world that might be affected by one of these proposals from conceptual engineers.

There are two major groups within contemporary philosophy that provide contrast with conceptual engineering: (i) the conservatives, who take proper philosophy to be purely descriptive, and (ii) those philosophers who think we ought to focus on finding better beliefs and theories. Examples of conservative stances are the ones Ludwig Wittgenstein and David Lewis take in the following passages:

> Philosophy may in no way interfere with the actual use of language, it can in the end only describe it. For it cannot give it any foundation either. It leaves everything as it is.[2]

> One comes to philosophy already endowed with a stock of opinions. It is not the business of philosophy either to undermine or to justify these preexisting opinions, to any great extent, but only to try to discover ways of expanding them into an orderly system.[3]

Philosophers have made few explicit statements to the effect that we ought to aim to find better beliefs about philosophical topics, but this view seems to be widespread. For example, it is natural to interpret philosophical debates as comprised of factions each arguing that their theory is right and the others are wrong. For example, epistemological internalists argue that everyone should adopt their beliefs about the conditions on knowledge, while epistemological externalists argue that everyone should adopt *their* beliefs about knowledge. Neither side in this dispute focuses on whether the concept of knowledge is a good concept, whether it has any internal defects, whether it ought to be replaced with one or more concepts that do its jobs better, and so on. It is these latter issues that conceptual engineers emphasize. Since concepts are usually assumed to be the constituents of beliefs and other thoughts, anyone who thinks our concepts need to be changed also thinks that at least some of the beliefs involving those concepts need to be changed as well. So conceptual engineers agree that we need new beliefs, but say that we first need better representational devices out of which better beliefs can be constructed.

In general, we have three major camps—(i) those who see proper philosophy as purely or mostly descriptive, (ii) those who see proper philosophy as primarily engaged in a search for the right beliefs (theories), and (iii) those who see proper philosophy as primarily focused on a search for the right representational devices. These three groups can be thought of as increasingly radical—the first camp (the conservatives) think our concepts and beliefs are mostly in order, the second camp (the moderates) think our concepts are in order, although we need better beliefs (i.e., theories), and the third camp (the conceptual engineers) think we need better concepts and better beliefs. Of course, these categories are intended for heuristic purposes—not all philosophers fit neatly into one camp, and there are varying degrees of overlap between the camps. Still, they are helpful for understanding the thrust of the conceptual engineering movement and what its proponents see as their opposition.

Conceptual engineers think that certain philosophical problems can be addressed by: (i) thinking about how to evaluate our concepts or meanings and (ii) if the evaluation is negative, thinking about how to make improvements. The potential improvements

are a diverse lot. They include: no longer using some concept or meaning for some purpose, introducing a new concept or meaning for some purpose, altering an existing concept or meaning, and changing which concept is expressed or meaning is had by a certain word. In some of these cases, we can think of an individual concept or meaning being improved. This kind of change requires thinking of concepts or meanings as things that can keep their identities through changes. If meanings are just intensions or extensions, then this isn't a plausible assumption because intensions are just mathematical functions and extensions are mathematical sets. If one has a more exotic view of meanings, then perhaps they can retain their identities through changes. Since concepts are mental representations, they might be able to persist through changes as well, but on some views they cannot. Whatever one's views on these matters, we can call *revision* any process of changing an existent concept or meaning. In particular, *conceptual revision* is changing a concept and *semantic revision* is changing a meaning.

In contrast to revision, the process of introducing a new concept or meaning to do something we previously tasked with some old concept or meaning is called *replacement*. Again, we can distinguish conceptual replacement from semantic replacement, but it is important to recognize that neither one involves changing an existing meaning or concept. Instead, replacement is a change to our conceptual scheme or meaning repertoire as a whole.

I defend a replacement strategy for the concept of truth. We ought to replace our concept of truth, for certain purposes, with a team of two concepts as a way of addressing the problems caused by the liar and other paradoxes.[4] This might seem like a lot of work, but it turns out that by "we," I mean only those theorists engaged in doing semantics for expressively rich languages (like English) that have the resources to formulate the paradoxes. I am definitely not suggesting that anyone else do anything at all with respect to the concept of truth (unless perhaps they find themselves trying to solve the liar paradox or its relatives).

This replacement project involves multiple parts. There is an *evaluation of our concept of truth* as defective—the concept itself is the source of the paradoxes. There is a *characterization of the defect*—the concept of truth has certain constitutive principles and these are inconsistent. By following these inconsistent constitutive principles, one can reason to a contradiction in the liar paradox. There is also a *suggestion for replacement*—when doing semantics for natural language, one ought to use the two replacement concepts instead of the concept of truth.

In the rest of this chapter, I place this project in a more general context. The next two sections lay out a comprehensive case for the claim that the *concept* of truth is inconsistent and that there is no *property* of being true. I then look at some properties that are similar to what we thought the property of truth should have been like and say a bit about what these properties might be like. Finally, I close with a discussion of various strategies for replacing the concept of truth.

2 Two Traditions on Truth

Over the last decade or so, there has been a fruitful interaction between those working on the nature of truth and those working on the paradoxes that affect truth, like the liar paradox. The project of replacing truth, described above, is an application of conceptual engineering to the literature on the paradoxes affecting the concept of truth. The liar paradox is the most famous of these, and it is that by reflecting on sentences like "This very sentence is not true," it is easy and quick to derive a contradiction: the sentence is both true and not true.[5] There is a school of thought on the liar paradox that diagnoses all the paradoxes as defects in our very concept of truth. We can call this the *inconsistency approach*.

Inconsistency theorists claim that the principles essential to the concept of truth permit a competent reasoner to derive a contradiction (e.g., liar sentences are both true and not true). The inconsistency approach goes back to Alfred Tarski's pioneering work in the 1930s, but it has really taken off since 2002, when Matti Eklund published "On Inconsistent Languages."[6] Since then, much has been done to explore various options and provide details for what were once just suggestions. Because my own work falls in this tradition, I am especially interested in these exploratory endeavors. Much has been written about how concepts can be inconsistent, how inconsistent concepts can be possessed and used, and what sorts of logics and semantics go best with an inconsistency approach. However, there is little on the property of being true from this perspective.

On the other hand, there has been a tremendous amount of work within the "nature of truth" tradition on the *metaphysics of truth*. The central issue in the metaphysics of truth discussion is whether the property of being true is deflationary or substantive. Is being true more like a logical property (e.g., being necessary) or more like a scientific property (e.g., being a mammal)? Is the property of being true cut out to explain anything?[7] Almost all the work on the metaphysics of truth assumes without question that the concept of truth is just fine and that there is a unique property of being true.

Our unifying topic is: *What does the metaphysical landscape look like for the inconsistency theorist?* This particular bridge over the two broad traditions has yet to be investigated in much detail. Is there a property of being true? If so, what is its nature? Might there be more than one equally good candidate for this property? If so, what are these candidates? How exactly can there be a property of being true if the concept of truth is defective? And, most importantly, if there is no such property, then are there any properties that are somewhat similar to what we thought the property of truth would be like? For this last question, think about it in this way. We have lots of beliefs about the nature of the property of being true.[8] If there are no properties that satisfy these beliefs, then there still might be properties that come close enough to count as the property of being true. And even if there are no properties that come close enough, there will no doubt be properties that are somewhat similar in various ways to what we thought the property of truth would have to be like.

In order to have a term for talking about all these properties at once without prejudging whether there is a unique property of being true, we can use "aletheic property" ("Aletheic" is just an adjective synonymous with "pertaining to truth.") Hence, an *aletheic property* is any property that is similar to what we think the property of truth would have to be like. With this terminology in hand, we can say: we are investigating the class of aletheic properties, and in particular we are investigating the aletheic properties from the point of view of an inconsistency approach.

According to the usage here, *the property of being true* is the property designated by *the English word "true,"* which expresses *the concept of truth*. When something is true, it has the property of being true, whether it is a sentence, story, song, proposition, theory, utterance, prediction, or whatever. Likewise, anything that has the property of being true is true. The property (being true), the word ("true"), and the concept (truth) are not to be confused.[9] Inconsistency theorists focus on the concept of truth and the word "true," while the metaphysics of truth is about the property of being true. We shall question whether there is a property of being true but *not* whether there is a word "true" or whether there is a concept of truth.

3 Truth Platitudes

Roughly, a *platitude* is a claim that seems obviously true, self-evident, or commonsensical. The platitudes for a given concept might be analytic (i.e., true in virtue of their meaning alone), but for our purposes, we do not need such a strong assumption. All the platitudes we care about are intuitively uncontroversial. We can say that all the platitudes we consider would be accepted by the vast majority of competent users of "true" as obvious. Moreover, they all count as *constitutive* of the concept of truth in a certain sense. Namely, if a person rejects one of these principles in a conversation, that is a pro tanto reason to think that the person's word "true" does not express the concept of truth. I have developed this notion of constitutive principles elsewhere, but it will not play a role in what follows.[10]

The point of assembling the platitudes is to find something in the world that does the best job of satisfying the platitudes. If the world cooperates, then the platitudes will be satisfied by a unique thing in the world. However, even if nothing perfectly fits the platitudes, something might fit them relatively well—well enough to say that it is what the term in question is ultimately about. In either of these cases, one might continue the investigation by considering whether the thing that fits the platitudes is fundamental or derivative and, if it is derivative, how it relates to the fundamental level of reality.

What are the platitudes for truth? There are so many to choose from, but we are going to focus on a very specific set of logical platitudes. *Logical platitudes* are those that involve truth and other logical expressions; for example, a conjunction is true if and only if both conjuncts are true. We focus on 12 specific platitudes because they are so central and they were the source of a very thorough investigation. My aim in discussing

truth platitudes is to argue that there is no property of being true because no real property even comes close to satisfying most of the central logical platitudes for truth.

Harvey Friedman and Michael Sheard focused on 12 logical platitudes for truth. We call them the *Friedman-Sheard Criteria*:[11]

(T-In) If p, then ⟨p⟩ is true. [E.g., if snow is white, then "snow is white" is true.]

(T-Out) If ⟨p⟩ is true, then p. [E.g., if "snow is white" is true, then snow is white.]

(T-Enter) If it is provable that p, then it is provable that ⟨p⟩ is true. [E.g., any theory that has "snow is white" as a theorem also has "'snow is white' is true" as a theorem.]

(T-Exit) If it is provable that ⟨p⟩ is true, then it is provable that p. [E.g., any theory that has "'snow is white' is true" as a theorem also has "snow is white" as a theorem.]

(¬T-Enter) If ⟨p⟩'s negation is provable, then it is provable that ⟨p⟩ is not true. [E.g., any theory that has "snow is not white" as a theorem also has "'snow is white' is not true" as a theorem.]

(¬T-Exit) If it is provable that ⟨p⟩ is not true, then ⟨p⟩'s negation is provable. [E.g., any theory that has "'snow is white' is not true" as a theorem also has "snow is not white" as a theorem.]

(T-Rep) If ⟨p⟩ is true, then ⟨⟨p⟩ is true⟩ is true. [E.g., if "snow is white" is true, then "'snow is white' is true" is true.]

(T-Del) If ⟨⟨p⟩ is true⟩ is true, then ⟨p⟩ is true. [E.g., if "'snow is white' is true" is true, then "snow is white" is true.]

(T-Comp) Either ⟨p⟩ or ⟨p⟩'s negation is true. [E.g., either "snow is white" is true or "snow is not white" is true.]

(T-Cons) It is not the case that both ⟨p⟩ and ⟨p⟩'s negation are true. [E.g., it is not the case that both "snow is white" is true and "snow is not white" is true.]

(U-Imb) If every instance of a universal generalization is true, then the universal generalization is true. [E.g., if all the instances of "everything is white" are true, then "everything is white" is true.]

(E-Exc) If an existential generalization is true, then some instance of the existential generalization is true. [E.g., if "something is white" is true, then some instance of "something is white" is true.]

Friedman and Sheard worked meticulously to find every inconsistent subset of these principles, and they gave consistency proofs for each consistent subset. That means they found every single paradox hiding in these principles, and they figured out which groupings of principles are paradox-free. It turns out that a grand total of *27* (!) distinct paradoxes are lurking in these 12 logical platitudes for truth.

We have to be careful because Friedman and Sheard make several background assumptions in their reasoning:

(MP) If a conditional is true and its antecedent is true, then its consequent is true. [E.g., if "if snow is white, then something is white" is true and "snow is white" is true, then "something is white" is true.]

(Taut) All tautologies are true. [E.g., "if snow is white, then snow is white" is true.]

(PRE) All axioms of PRE are true, where PRE is a certain mathematical theory that contains specific mathematical equations. [E.g., "0=0" is true.]

Friedman and Sheard also assume Peano arithmetic, which is a popular axiomatic theory of arithmetic (e.g., a theory of the natural numbers, addition, multiplication, etc.).

In assuming (PRE) and Peano Arithmetic, Friedman and Sheard are following standard protocol in the literature on axiomatic theories of truth. They use numerals in their object language to refer to symbols, expressions, and sentences of the object language via a famous method called Gödel numbering. The truth predicate of their object language has only numbers in its extension, and these are intended to be interpreted as the Gödel numbers of the true sentences of the object language. Because Friedman and Sheard use Gödel numbers and arithmetic to prove various things about the sentences of their object language, they need some kind of mathematical theory for these proofs; Peano Arithmetic and PRE are the mathematical theories they use.[12]

Because we care about natural language, we are not concerned with any of these details beyond noting that we achieve the same sorts of aims by assuming various things about sentences of English (e.g., that liar sentences exist and that the liar is identical to the sentence "the liar is not true"). Given that (MP), (Taut), and (PRE) are assumed throughout their discussion, it makes sense for us to add them to the list of Friedman-Sheard Criteria. There is no reason to think that they are somehow more sacrosanct than any of the other platitudes. However, giving up (PRE) would not help much with the aletheic paradoxes in natural language because reliance on (PRE) is largely an artifact of the object language in Friedman and Sheard's treatment. When reasoning about natural language paradoxes, we use various assumptions about which words refer to which sentences instead.

Friedman and Sheard showed that there are exactly nine maximal consistent subsets of the Friedman-Sheard Criteria. These subsets are:

A. T-In, T-Enter, ¬T-Exit, T-Del, T-Rep, T-Comp, U-Imb, E-Exc.

B. T-Rep, T-Cons, T-Comp, U-Imb, E-Exc.

C. T-Del, T-Cons, T-Comp, U-Imb, E-Exc.

D. T-Enter, T-Exit, ¬T-Enter, ¬T-Exit, T-Cons, T-Comp, U-Imb, E-Exc.

E. T-Enter, T-Exit, ¬T-Enter, T-Del, T-Cons, U-Imb.

F. T-Enter, T-Exit, ¬T-Exit, T-Del, U-Imb.

G. T-Enter, T-Exit, ¬T-Exit, T-Rep, U-Imb.

H. T-Out, T-Exit, ¬T-Enter, T-Rep, T-Del, T-Cons, U-Imb.

I. T-Exit, ¬T-Exit, T-Rep, T-Del, U-Imb.

Every other subset of these principles that is not listed explicitly here or is not a subset of one of these listed is inconsistent. In figure 34.1, there is a table of the Friedman-Sheard Criteria on the left and the consistent subsets on the top, with "✓" indicating inclusion of the platitude on the left of that row in the subset listed at the top of that column.

Remember that the three additional platitudes that Friedman and Sheard used as background reasoning should be appended to each of the nine consistent subsets of Friedman-Sheard Criteria.

Some observations about the table deserve comment. First, Friedman and Sheard label the maximally consistent subsets in the order in which they prove them consistent, and they begin with the easiest cases first. Other than that there is nothing significant about the order.

Next, Universal Imbibe (U-Imb) is a member of every single consistent subset, which means that it is not involved in any of the paradoxes Friedman and Sheard find; none of the other Friedman-Sheard Criteria is like this. I shall use the term "innocuous" for this feature, and "destructive" for its converse. With this terminology, we can say that (U-Imb) is *the most innocuous* of the Friedman-Sheard Criteria, given their background assumptions.

Note how *destructive* (T-In) and (T-Out) are individually; that is, each one is inconsistent with lots of other subsets of the Friedman-Sheard Criteria. Each one shows up only one time among the consistent subsets of Friedman-Sheard Criteria; (T-In) is in subset A, and (T-Out) is in subset H. None of the other criteria are even close to being that destructive. Remember that (T-In) and (T-Out) are the two conditionals conjoined in the famous biconditional:

(Schema T) ⟨p⟩ is true if and only if p.[13]

So *each direction* of (Schema T), *all by itself*, is ridiculously destructive. No wonder they cause so much trouble together!

Only one subset contains (T-Enter), (T-Exit), (¬T-Enter), and (¬T-Exit) together. If one is looking for a more innocuous version of (Schema T) that is consistent, then this combination is one option.[14] Together these four rules say something like: a theory proves a sentence if and only if the theory proves the truth attribution to that sentence, and a theory proves a negated sentence if and only if the theory proves the negation of the truth attribution to that sentence. Again, these rules cannot be used in hypothetical reasoning—for example, when one argues using a reductio or conditional proof. Even

	A	B	C	D	E	F	G	H	I
(T-In) $P \rightarrow T\langle p \rangle$	✔								
(T-Out) $T\langle p \rangle \rightarrow P$								✔	
(T-Enter) $\vdash P \rightarrow \vdash T\langle p \rangle$	✔			✔	✔	✔	✔		
(T-Exit) $\vdash T\langle p \rangle \rightarrow \vdash p$				✔	✔	✔	✔	✔	✔
($\neg$T-Enter) $\vdash \neg p \rightarrow \vdash \neg T\langle p \rangle$				✔	✔			✔	
($\neg$T-Exit) $\vdash \neg T\langle p \rangle \rightarrow \vdash \neg p$	✔			✔		✔	✔		✔
(T-Comp) $T\langle p \rangle \ V \ T\langle \neg p \rangle$	✔	✔	✔	✔					
(T-Cons) $\neg(T\langle p \rangle \ \Lambda \ T\langle \neg p \rangle$		✔	✔	✔	✔			✔	
(T-Del) $T\langle T\langle p \rangle \rangle \rightarrow T\langle p \rangle$	✔		✔		✔	✔		✔	✔
(T-Rep) $T\langle p \rangle \rightarrow T\langle T\langle p \rangle \rangle$	✔	✔					✔	✔	✔
(U-Imb) $(\forall x)T\langle \phi(x) \rangle \rightarrow T\langle (\forall x)\phi(x) \rangle$	✔	✔	✔	✔	✔	✔	✔	✔	✔
(E-Exc) $T\langle (\exists x)\phi(x) \rangle \rightarrow (\exists x)T\langle \phi(x) \rangle$	✔	✔	✔	✔					

Figure 34.1
Maximal consistent subsets of Friedman-Sheard Criteria.

though this combination of the four derivation rules is considerably more innocuous than (Schema T), they are still *very* destructive together—they occur together only in subset D.

The most important thing to note is how *empty* the table is. Only 54 of the 108 squares are occupied. *Half!* Even if the table were only 75% occupied, that would be really bad—there would be lots of contradictions within those 12 Friedman-Sheard Criteria.

The Friedman-Sheard Criteria seem to give us *nine* aletheic properties—one property for each consistent subset of the Friedman-Sheard Criteria. We can call any property that satisfies the platitudes in subset A *the property of being FSAtrue*. We can call any property that satisfies the platitudes in subset B *the property of being FSBtrue*. And so on.

At last we have something to work with when thinking about which property is the property of being true.[15] And the results are grim. We are assuming for now that there is a property of being FSAtrue and a property of being FSBtrue and so on for each of our nine families. But which one of these is the property of being true? Of course, none of them satisfy all the Friedman-Sheard Criteria. But more damning is that none of them *even come close* to satisfying all the Friedman-Sheard Criteria. The best meet two-thirds of the criteria (subset A and subset D). Maybe you think that is close enough to count as the property of being true, but that view runs into another problem—there are two properties that satisfy 8 out of the 12 criteria. So even if the Friedman-Sheard Criteria were the only truth platitudes, no property comes close to satisfying all of them, and multiple distinct properties seem to do equally good jobs of satisfying some of them.

4 The Nightmare

There are 27 distinct paradoxes among the 12 Friedman-Sheard Criteria, and there are nine maximal consistent subsets of these criteria. The largest two maximal subsets have only eight members, but the average size of the maximal subsets is a mere six members. There are 495 distinct subsets of the Friedman-Sheard Criteria with eight members, but only *two* of these subsets are consistent! Let that sink in for a moment.

Moreover, if we were to weight the Friedman-Sheard Criteria, (T-In) and (T-Out) would easily get the highest weights because of how central they are to the functioning of "true." Recall that these are each direction of the celebrated (Schema T). But each of these central criteria shows up in only one of the maximal subsets.

However, the situation is actually much worse. For complex technical reasons, no aletheic properties satisfy three of these maximal subsets of platitudes—A, D, E. Hence, only Subsets B, C, F, G, H, and I are satisfied by properties at all.

In addition, Subset G and Subset H violate obvious conditions on theories of any kind—that is, that theories should be consistent and not self-refuting.[16] When we eliminate subsets A, D, E, G, and H, we are left with only *four* subsets of Friedman-Sheard critieria, and each of these subsets contains only *five* members. Five! Out of 12 criteria.

And none of these four subsets—B, C, F, and I—contain either of the most important criteria: (T-In) and (T-Out).

The number of ways in which one can derive contradictions from the logical platitudes about truth is simply mind-boggling. This is exactly the nightmare scenario that many theorists writing about the nature of truth pray never happens. All the candidate properties are so far from satisfying even minimal subsets of truth platitudes that there is no decent way to judge which one of the aletheic properties is *the* property of being true. They are so far from what we think of when we think of the concept of truth that it does not make sense to call any of them the property of being true. The result of this exercise is that there is no property of being true.[17]

If we place these results in the literature on the inconsistency approaches to the paradoxes, then the familiar perspective is changed dramatically. The *usual* point at which those of us who are inconsistency theorists about truth conclude our case is that (Schema T) is both central to the concept of truth *and* inconsistent in reasonable logics. Hence, the concept of truth is defective.[18] That is, the typical inconsistency theorist puts *all* the weight on a *single* logical platitude, (Schema T), and a *single* paradox, the version of the liar that shows (Schema T) is inconsistent. I want to be clear that I agree with this conclusion, but this standard argument for the inconsistency of truth, by itself, is not enough. It does not give us anything like a comprehensive picture of the defect in our concept of truth. *Inconsistency theorists have, so far, only seen the tip of the iceberg*. There are 27 distinct paradoxes in just the 12 most basic logical platitudes for truth. The concept of truth is far, far worse off than we have noticed.

5 If There Is No Property of Being True, Then What Is There?

Now that we have some sense of which aletheic properties exist and how they differ from what we intuitively take the property of being true to be like, we can turn to the question of what these aletheic properties are like. In particular, we can think about the question of whether deflationism is right about any of these aletheic properties.

Deflationism about truth has been influential in analytic philosophy for decades now, but there has been recent interest in what deflationists should say about the property of being true. First-generation deflationists tended to deny that there was any property denoted by "true," but the subsequent debate made it clear that most of them meant instead that there is no *substantive* property denoted by "true."[19] Even so, most of the attention was on what form a deflationary theory of truth should take and on what a deflationist should say about the role of truth predicates in our linguistic practice. However, the debate has now shifted to deflationary views on the property of being true, with the result being that earlier characterizations have been seen as inadequate. There is a new sense that the debate over deflationism about truth might well be adjudicated best by evaluating what deflationists say about the property of being true.

Jeremy Wyatt's 2016 paper, "The Many (Yet Few) Faces of Deflationism," catalogs five deflationary theses about the property of being true:

(Transparency) Being true is a metaphysically transparent property,

(Non-explanatory) Being true is a non-explanatory property,

(Unconstituted) Being true is not constituted by any other property,[20]

(Abundant) Being true is an abundant property, and

(Logical) Being true is a logical property.

These need not be accepted by every deflationist, but each has considerable support. A few clarificatory comments on these deflationary theses are in order.

First, being true is a *metaphysically transparent* property if and only if anyone who possesses the concept of truth is in a position to know all the essential facts about the property of being true. In other words, the property of being true does not have some hidden essence to be discovered by some investigation. Simply having the concept of truth is enough to be in a position to know everything important about the property of being true (but not, obviously, which things have that property).

Second, being true is a *non-explanatory* property iff there are no facts that are explained by facts about the property of being true.[21] It is important to stress that "true" might occur in a theory only in its expressive role as a device of generalization despite the fact that the theory in question is genuinely explanatory. The word "true" serves as a device of generalization when it changes a sentence position into a singular term position in a sentence. (Non-explanatory) entails that the word "true" does not (or should not) play any role in any explanatory theory other than this generalizing role.

Third, the property of being true is *not constituted* by any other property iff it is not the case that there is some property, being F, such that all and only things that have the property of being true have the property of being F, *and* anything that has the property of being true has it *because* it has the property of being F. For example, if truth is properly analyzed as correspondence with the facts, then true things have the property of being true *because* they have the property of corresponding with the facts. Deflationists deny that truth is constituted by correspondence or by anything else. We could say that, together, (Non-explanatory) and (Unconstituted) imply that the property of being true doesn't explain anything and nothing explains it.

A property is *abundant* iff it is to some extent unnatural, in Lewis's sense. Naturalness, for Lewis, is objective, and it explains objective similarities among things.[22] Natural properties "cut nature at its joints," in Plato's phrase. Most metaphysicians accept that there are degrees of naturalness, but it is not clear whether deflationists endorsing (Abundant) mean that being true is merely not perfectly natural or they mean that it is highly unnatural.[23]

A property is *logical* iff it is invariant under certain one-one transformations of the world onto itself.[24] There are other ways to define "logical," but Wyatt follows Tarski,

who was inspired by Klein, and this invariance tradition is one of the most respected when it comes to defining what is logical. A one-one transformation of the world maps everything in the world to something in the world. For example, the identity transformation maps everything to itself. If the distribution of a property is unaffected by any such transformation, then the property is logical.

What are we to make of these deflationary theses? We already know that there is no property of being true, so deflationist theses about truth in particular are false. Nevertheless, there are aletheic properties, and these are somewhat like what we expected the property of being true to be like. It makes sense to investigate whether any of the aletheic properties count as deflationary in any of these ways. That is our focus in the rest of this section.

Do the aletheic properties satisfy (Transparency)?[25] For example, is the property of being FSAtrue metaphysically transparent? Is anyone who possesses the concept of FSAtruth in a position to know all the essential facts about the property of being FSAtrue? This depends on what one takes to be *essential* to the property of being FSAtrue. If we assume that possessing the concept requires some kind of acceptance of the axioms in subset A of the Friedman-Sheard Criteria, then the answer is clearly *no*.[26] One might understand and accept these axioms without being able to figure out that there is only one model for this subset whose domain is the natural numbers and whose arithmetic vocabulary have their standard interpretations. And such a person might also not be in a position to figure out that everything is in the extension of "true" in this model. The fact that every sentence of the object language has the property of being FSAtrue seems essential to me, but people differ on what is essential even in obvious cases, much less on esoteric subjects like this one. This result is even *necessary* in some sense because it holds in all the relevant models. The same sort of thing can be said for the other aletheic concepts and properties. I doubt that any of them is transparent.

(Non-explanatory) should be treated as highly dubious. The concept of truth shows up in truth-conditional semantic theories, which are among the most widely accepted theories for doing natural language semantics in the science of linguistics.[27] Denying that these theories have explanatory power would be like denying that Newton's theory of mechanics or Maxwell's theory of electrodynamics has explanatory power. It is an open question as to whether any of the aletheic properties (or their related concepts) have this explanatory power, but they might work in the same sorts of semantic theories just as well as we thought truth would.[28]

Although we now know that there is no property of being true, the verdict on the explanatory power of the genuine aletheic properties is complex. Truth-conditional semantic theories come in a dizzying variety, and they are often tailored to the specific linguistic expressions under consideration. Are any of the aletheic properties we canvassed up to the task? In the vast majority of cases, the clauses of a truth-conditional semantic theory could be added consistently to any of the axiomatic theories we have studied. The reason is that these semantic theories are not designed to apply to

language fragments that contain truth predicates. Instead, semantic theories for epistemic modals, for example, apply to language fragments that contain epistemic modals, and semantic theories for conditionals apply to language fragments that contain conditionals. The problems in which we are interested crop up only for a truth-conditional semantic theory when they are interpreted as applying to sentences in which truth predicates occur. Hence, in this sense, any of our aletheic properties would be up to the task of satisfying the principles of most truth-conditional semantic theories.

But there is a catch. If we ask ourselves whether any of the aletheic properties studied so far could satisfy *all* the principles of a truth-conditional semantic theory when it is interpreted as applying to a language with a truth predicate and the resources to construct liar sentences, then the answer is *no*. There are plenty of paradoxes hiding among these principles, and so no aletheic property is going to satisfy all of them (for a detailed argument, see Scharp [2013]). Therefore, any of the aletheic properties could serve an explanatory role in almost any truth-conditional semantic theory, but when it comes to a truth-conditional semantic theory *for a truth predicate*, none of them are up to the task. Never fear, because there is a way to fix this problem, which is a topic of the next section.

Are any of the aletheic properties *unconstituted*? This is hard to say because it is not obvious how to individuate properties, and it isn't clear what kind of dependence is invoked with the claim that the property of being true is *constituted* by some other property. If something like a reductive explanation is given as the reading of "constituted," then the answer is probably going to be *no* because uncontroversial reductive explanations (think thermodynamics and statistical mechanics) are rare in philosophy. If the standard for successful reduction is somewhat relaxed, say, to a weak supervenience claim, then perhaps the odds are a bit better, but probably not by much. What kind of explanatorily powerful property is going to explain any one of these aletheic properties? They are each hopelessly gerrymandered—their extensions zigzagging around to avoid the plethora of impossibility results. Hence, it looks like all of the aletheic properties are unconstituted.

One might protest this conclusion: but we already have such properties! Correspondence to the facts or coherence, or superwarrant maybe.[29] My reply: none of the classic analyses of truth are remotely plausible in light of the results in sections 3 and 4. First of all, there is no property of being true. That result is inconsistent with every purported constitution theory for the property of being true. Second, even if one denies this conclusion, it is hard to see that the property of corresponding to the facts could constitute any one of the aletheic properties. Consider the property of being FSFtrue (i.e., satisfying subset F of the Friedman-Sheard Criteria). This property fails to obey (¬T-Enter). Hence, even if one can prove a theorem from subset F of form ¬p, one cannot conclude that ¬Tp. For example, it would be like proving "snow is not white" but failing to prove "'snow is white' is not true." Does the property of corresponding to the facts behave just like this? Highly doubtful. And the same goes for the other familiar analyses of truth.

What about *abundance*? All of the aletheic properties are highly unnatural in one sense. Naturalness is supposed to explain the objective similarities between things. If we intuitively think of all the truths as having some kind of objective similarity, then every aletheic property is going to violate this in myriad ways. And it is hard to believe that anyone thinks that all the FSFtruths or all the FSCtruths have objective similarity. Hence, all the aletheic properties are deflationary in this sense.

Finally, what about the logicality of the aletheic properties? Wyatt argues that no deflationist should accept this thesis because it is easily refuted.[30] His argument is that some transformations will map some true proposition onto a false proposition, so truth is not preserved under all transformations. But Wyatt assumes that propositions are things in a particular world, whereas it is much more common to assume that propositions are sets of possible worlds, and so not members of any particular possible world. If that is right, then propositions are not among the things in a world, and so Wyatt's argument fails.

Certainly, some aletheic properties count as logical: when the predicates that denote them get their own clauses in the semantic theory for languages in which they occur, just like negation and the rest of the logical vocabulary. If that is the case, then the interpretation of the predicate is by definition invariant across all models.[31] So at least some aletheic properties count as logical properties.

We can summarize the results in this section:

(i) No aletheic properties are transparent.

(ii) No aletheic properties are non-explanatory.

(iii) All aletheic properties are unconstituted.

(iv) All aletheic properties are abundant.

(v) At least some aletheic properties are logical (depending on how one understands logical expressions and how one formulates a semantic theory for a language with expressions that denote these properties).

It deserves to be emphasized that deflationist theories of truth have multiple aspects or parts, some of which pertain to the property of being true, some to the concept of truth, some to the word "true," and some to the structure of any acceptable theory of truth. Moreover, there is no property of being true, so any theory, deflationist or not, that entails that there is such a property is false. At present, we are evaluating only *deflationist views of aletheic properties*, not deflationist views as a whole. As such, even those versions of deflationism that come out as acceptable on the present inquiry might be false for some other reason (e.g., because deflationist theories are typically taken to consist of all and only the nonparadoxical instances of (Schema T), but it is difficult or impossible to specify in advance which instances these will be).

6 New Aletheic Concepts

There is a meaningful English word, "true," and there is a concept of truth expressed by that word, but the concept of truth is defective in the sense that its platitudes are inconsistent. Moreover, they are *seriously* inconsistent—the average size of the nine maximally consistent subsets of 12 Friedman-Sheard Criteria is only six, and the average size of the somewhat reasonable ones is only five. As such, there is no property of being true. Still, there are plenty of properties that are somewhat like what we think the property of being true should have been like. These are the *aletheic properties*, and there is one for each consistent subset of platitudes for the concept of truth.

Instead of ending our inquiry in this bleak place, we have decided to see whether we can improve our situation by adding new aletheic *concepts* to our conceptual scheme. Our guide to the world of new aletheic concepts is the realm of aletheic properties. In other words, for each aletheic property, there is a new aletheic concept, distinct from the concept of truth. There are, of course, lots of other concepts that are similar to but distinct from the concept of truth, but the ones based on aletheic properties come with a guarantee: *I am not inconsistent*.[32] Each new aletheic concept has, as its constitutive principles, all the platitudes in the theory it satisfies—for example, the concept of FSFtruth has as its platitudes all the axioms of Subset F of the Friedman-Sheard Criteria, which is satisfied by the property of being FSFtrue.

One significant issue for choosing between new aletheic concepts is whether the aletheic properties they denote are deflationary or substantive, and in what sense. We saw five ways of drawing the deflationary/substantive distinction for the aletheic properties from Wyatt and conducted a preliminary investigation into which aletheic properties have these features. Against the background of a conceptual engineering project, the question becomes: what do we want our replacement concepts to be?[33] What are the considerations for and against various features—explanatory power, expressive power, naturalness? And which ones do we want, given what we are interested in doing with our replacement concepts?

There is good reason to think that no single new aletheic concept is up to the whole job because no single aletheic concept can do everything we expect the concept of truth to do. For example, it is widely accepted that we use "true" to endorse propositions that we cannot assert directly. We can capture this role by saying that a truth predicate functions as a *device of endorsement* (i.e., an attribution of truth to a sentence entails that sentence). The flip side of this role is *a device of rejection*. In order to serve as a device of endorsement, the truth predicate must obey (T-Out), and in order to serve as a device of rejection, the truth predicate must obey (T-In). Of course, we already know that in a classical setting no single concept obeys these two principles; thus, no concept can serve as both a device of endorsement and rejection given classical logic and the expressive resources to construct liar sentences. However, if we replace truth with two

concepts, we can split the workload, allowing one to serve as a device of endorsement and the other to serve as a device of rejection.[34]

Another aspect of the same problem—that no single aletheic concept can perform truth's explanatory role—pertains to truth-conditional semantics in both philosophy and linguistics. Recall that *any* of the aletheic properties could serve the explanatory role required by the vast majority of semantic theories in the truth-conditional tradition. The reason is that few of these theories are intended to apply to fragments of natural language that include a truth predicate. However, none of them can be used to formulate a successful truth-conditional semantic theory that applies to the fragments of a natural language containing a truth predicate and minimal expressive resources (e.g., names of sentences). This should be obvious: the aletheic properties differ substantially from what we thought the property of being true should be like, so the aletheic properties deliver nothing like what we think of as truth conditions.

However, if we adopt the concepts I suggest as replacements for truth—ascending truth and descending truth—into our conceptual scheme, then we can formulate a successful semantic theory using them. Ascending truth obeys something like (T-In), and descending truth obeys something like (T-Out). The semantic theory that uses these replacement concepts specifies *ascending* truth conditions and *descending* truth conditions for all the sentences in a fragment of natural language that contains a truth predicate, an ascending truth predicate, and a descending truth predicate. Moreover, for safe sentences—roughly, the nonparadoxical ones—the ascending truth conditions are identical to the descending truth conditions, which are identical to the truth conditions. Thus, *ascending and descending truth-conditional semantics* reduces to *truth-conditional semantics* in all the familiar cases where the distinction between ascending truth and descending truth is negligible. That is similar to the situation in physics—relativistic mechanics reduces to Newtonian mechanics in all the familiar cases where the distinction between relativistic mass and proper mass is negligible. All this with classical logic and no revenge paradoxes; for details, see Scharp (2013).

Whether my suggested replacements for the concept of truth can actually do all the things we think truth should be able to do is still an open question. Truth's expressive role is complex and truth's explanatory role is vast. Here I have focused only on the expressive roles of acceptance, rejection, and generalization and the explanatory role in contemporary natural-language semantics. As such, we have just scratched the surface of the conceptual engineering project for truth.

What we desperately need is a catalogue of all the inconsistencies among all the central logical platitudes for truth listed in section 2 and a catalogue of all the maximal consistent subsets of these criteria. John Burgess has already suggested doing something like this, so we can reiterate his call to arms.[35] We can call it the *aletheic platitudes project* (APP). When complete, APP would be a gold mine for the study of aletheic properties. Moreover, APP would finally allow us to complete the conceptual engineering project

of finding the best team of new aletheic concepts to replace our defective concept of truth. I think the community of mathematically minded philosophers and logicians can work together to make APP a real success.

Notes

1. See Scharp (2013) and Cappelen (2018) for those that distinguish them.

2. Wittgenstein (1953, sec. 124).

3. Lewis (1973, 88).

4. See Scharp (2013).

5. See Künne (2003) for an overview of the *nature of truth tradition*, and see Field (2008) for an overview of the *paradox tradition*.

6. Tarski (1933); Eklund (2002). See also Burgess (2006); Scharp (2013).

7. See Beall and Armour-Garb (2005) for a helpful summary and classic papers on deflationism. See Wyatt (2016) for an overview of the discussion about whether the property of being true is deflationary.

8. I am not going to be precise about this matter yet. I use phrases like "beliefs about what truth should be like" and "beliefs about what truth must be like" and "beliefs about what is essential to truth" interchangeably. See section 3 for further discussion of this topic and the nature of the platitudes for truth.

9. See Bar-On and Simmons (2007) and Asay (2013, chap. 1) for good examples of clarity on this.

10. See Scharp (2013); see also Lynch (2009, 13) for a similar view.

11. Friedman and Sheard (1987). See also Friedman and Sheard (1988); Leigh and Rathjen (2010); Leigh (2015). In the remainder, "p" is a sentential variable, which means the instances of the schema have a sentence in place of "p," and the angle brackets are a naming device associated with sentential variables. That is, in the instances of the schema, the name of the sentence replaces the angle bracket expression. Quotation marks are used to form names of the expressions occurring inside them.

12. Friedman and Sheard also use (PRE) in results about the proof-theoretic strength of various theories, but this topic is independent of ours.

13. In each instance of Schema (T), the letter "p" is filled in with a sentence, and "<p>" is filled in with a name of that sentence.

14. See Greenough (2001) for a philosophical motivation for this combination.

15. An alternative approach would investigate properties that satisfy *many but not all* instances of certain truth platitudes, rather than properties that satisfy *all* the truth platitudes in some maximal consistent set of platitudes. For example, Horwich (1998) makes a suggestion like this

for just (Schema T). Evaluating this alternative is beyond the scope of this chapter. See McGee (1992); Horwich (1998); and Schindler (2015) for discussion.

16. See Scharp (2020) for more details.

17. If I *had* to pick, I would follow Jeremy Wyatt who (personal communication) suggested Subset F as the closest to what the property of being true would be like because it alone among the survivors has both directions of some kind of disquotational platitude: (T-Enter) and (T-Exit)—as well as (¬T-Exit). Also, note that some deflationists (e.g., expressivist views like those described in Schroeder [2010]) might welcome the conclusion that there is no property of being true. Still, disquotationalists and minimalists, which make up the overwhelming majority of deflationists, do accept that there is a property of being true, just not a substantive property. See more on this in section 5.

18. Eklund (2002); A. Burgess (2006); and my own presentation in Scharp (2013).

19. See Armour-Garb and Beall (2005) for a survey.

20. Wyatt claims that (Unconstituted) should not count because Horwich rejects it, but I include it because it is prominent in the literature and adds to the discussion.

21. Wyatt formulates this condition in terms of facts about *the essence of* the property of being true. However, I am not clear on what this rules out. Regardless of how one formulates the condition, deflationists have long held that the property of being true can play no role in explaining meaning. As long as this result follows, nothing turns on the differences in formulation for (Non-explanatory).

22. See Lewis (1986); see also Sider (2011).

23. See Asay (2013) and Edwards (2013).

24. This is very rough; see MacFarlane (2000) for a detailed treatment.

25. Thank you to a referee who emphasized this question.

26. And I would say that the antecedent of this conditional is far too demanding for concept possession.

27. See Chierchia and McConnell-Ginet (1990).

28. Many deflationists have defended the immodest view that deflationism *is* incompatible with truth-conditional semantics, but there are attempts to reconcile deflationism with a reasonable modesty toward the sciences. For example, Michael Williams argues that the concept of truth plays only a generalizing role in truth-conditional semantic theories; see Williams (1999). See also A. Burgess (2011) and McGee (2016).

29. See Künne (2003) for detailed commentary on each of these attempts to analyze truth.

30. Wyatt (2016, 16–17).

31. See Scharp (2013), where this sort of formulation is given for ascending truth and descending truth.

32. And an even further guarantee against consistent but omega-inconsistent concepts. As far as I know, no one has considered omega-inconsistent concepts before.

33. See Haslanger (2000) for inspiration here.

34. See Scharp (2013) for a detailed argument.

35. J. Burgess (2011).

References

Armour-Garb, B., and Beall, Jc, eds. 2005. *Deflationary Truth*. New York: Open Court.

Asay, J. 2013. *The Primitivist Theory of Truth*. Cambridge: Cambridge University Press.

Bar-On, D., and Simmons, K. 2007. "The Use of Force against Deflationism." In D. Greimann and G. Siegwart, eds., *Truth and Speech Acts*, 61–89. New York: Routledge.

Beall, Jc. 2009. *Spandrels of Truth*. Oxford: Oxford University Press.

Burgess, A. 2006. "Identifying Fact and Fiction." PhD diss., Princeton University.

Burgess, J. 2011a. "Friedman and the Axiomatization of Kripke's Theory of Truth." In N. Tennant, ed., *Foundational Adventures: Essays in Honor of Harvey M. Friedman*, 125–148. London: Templeton Press/College.

Burgess, A. 2011b. "Mainstream Semantics+Deflationary Truth." *Linguistics and Philosophy* 34: 397–410.

Cappelen, H. 2018. *Fixing Language: Conceptual Engineering and the Limits of Revision*. Oxford: Oxford University Press.

Chierchia, G., and McConnell-Ginet, S. 1990. *Meaning and Grammar: An Introduction to Semantics*. Cambridge, MA: MIT Press.

Edwards, D. 2013. "Truth as a Substantive Property." *Australasian Journal of Philosophy* 91: 271–294.

Eklund, M. 2002. "Inconsistent Languages." *Philosophy and Phenomenological Research* 64: 251–275.

Field, H. 2008. *Saving Truth from Paradox*. Oxford: Oxford University Press.

Friedman, H., and Sheard, M. 1987. "An Axiomatic Approach to Self-referential Truth." *Annals of Pure and Applied Logic* 33: 1–21.

Friedman, H., and Sheard, M. 1988. "The Disjunction and Existence Properties for Axiomatic Systems of Truth." *Annals of Pure and Applied Logic* 40: 1–10.

Greenough, P. 2001. "Free Assumptions and the Liar Paradox." *American Philosophical Quarterly* 38: 115–135.

Haslanger, S. 2000. "Race and Gender: (What) Are They? (What) Do We Want Them to Be?" *Noûs* 34: 31–55.

Horwich, P. 1998. *Truth*. 2nd ed. Oxford: Clarendon Press.

Künne, W. 2003. *Conceptions of Truth*. Oxford: Clarendon Press.

Leigh, G., and Rathjen, M. 2010. "An Ordinal Analysis for Theories of Self-Referential Truth." *Archive for Mathematical Logic* 49: 213–247.

Leigh, G. E. 2015. "Some Weak Theories of Truth." In T. Achourioti, H. Galinon, J. M. Fernández, and K. Fujimoto, eds., *Unifying the Philosophy of Truth*, 281–292. New York: Springer.

Lewis, D. 1973. *Counterfactuals*. Oxford: Blackwell.

Lewis, D. 1986. "New Work for a Theory of Universals." *Australasian Journal of Philosophy* 61: 343–377.

Lynch, M. 2009. *Truth as One and Many*. Oxford: Oxford University Press.

MacFarlane, J. 2000. "What Does It Mean to Say That Logic Is Formal?" PhD diss., University of Pittsburgh.

McGee, V. 1992. "Maximal Consistent Sets of Instances of Tarski's Schema (T)." *Journal of Philosophical Logic* 21: 235–241.

McGee, V. 2016. "Thought, Thoughts, and Deflationism." *Philosophical Studies* 173: 3153–3168.

Priest, G. 2006. *In Contradiction*. 2nd ed. Oxford: Oxford University Press.

Scharp, K. 2013. *Replacing Truth*. Oxford: Oxford University Press.

Scharp, K. 2020. "Conceptual Engineering for Truth: Aletheic Properties and New Aletheic Concepts." *Synthese*. https://doi.org/10.1007/s11229-019-02491-4.

Schindler, T. 2015. "A Disquotational Theory of Truth as Strong as Z_2^-." *Journal of Philosophical Logic* 44: 395–410.

Schroeder, M. 2010. "How to Be an Expressivist about Truth." In C. D. Wright and N. J. L. L. Pedersen, eds., *New Waves in Truth*, 282–298. New York: Palgrave Macmillan.

Sider, T. 2011. *Writing the Book of the World*. Oxford: Oxford University Press.

Tarski, A. 1933. "The Concept of Truth in Formalized Languages." In J. Corcoran, ed., and J. H. Woodger, trans., *Logic, Semantics, Meta-Mathematics*, 152–278. Indianapolis: Hackett, 1983.

Williams, M. 1999. "Meaning and Deflationary Truth." *Journal of Philosophy* 96: 545–564.

Wittgenstein, L. 1953. *Philosophical Investigations*. Translated by G. E. M. Anscombe. Oxford: Basil Blackwell.

Wyatt, J. 2016. "The Many (Yet Few) Faces of Deflationism." *Philosophical Quarterly* 263: 362–382.