Introduction to the Semantic Paradoxes

Bryan Frances

Introduction

Among people who love language, one of the most famous sentences in the English language is one that allegedly doesn't even make any sense: 'Colorless green ideas sleep furiously'. The philosopher-linguist-political activist Noam Chomsky discussed that sentence in a 1957 work of his in order to disprove certain theories in linguistics. Since the sentence is so charming it became famous independently of its initial purpose.

But is the Chomsky sentence really meaningless? On the contrary, it strikes the clear majority of philosophers as quite simple to understand. In fact, most of them hold that the sentence is not just meaningful but straightforwardly *true*. Let me explain.

What the Chomsky sentence says is this: every idea that is both colorless and green does a certain thing: sleep in a furious manner. That is, any and every colorless green idea sleeps furiously. Yet another way to put it: *if* something is a colorless green idea, *then* it sleeps furiously. But notice that the Chomsky sentence doesn't say that there actually *are* any colorless green ideas! It makes no such claim. All it says is that *if there are any*, then they all sleep in a furious manner. A sentence of the form 'If there is an X, then it has characteristic C', or 'All Xs have characteristic C', doesn't say that there actually are any such Xs! All it says is that *if* there really are any Xs, then something follows (each of them has characteristic C).

Let 'CGI' mean *colorless green idea*, and let 'SF' mean *sleep furiously*. So, the Chomsky sentence says: if there is a CGI, then it SF. Hence, in order for the sentence to be false there has to be a CGI that fails to SF. That is the key to seeing the true meaning of the Chomsky sentence. Hence, the following is correct:

> The Chomsky sentence is false = there is a colorless green idea that does *not* sleep furiously.

Well, can you come up with any ideas that fit the bill, that make the right side the equality true? In other words, can you find a colorless green idea that doesn't sleep furiously? Of course not! There aren't any colorless green ideas at all, let alone ones that don't sleep furiously. The Chomsky sentence is odd because it talks about a category—that of colorless green ideas—that has nothing in it. In that respect it is like the sentence 'White unicorns have four legs', as there are no white unicorns.

So the right half of the equation indented above is false. Since the left half equals the right half, the left half is false as well. So, it's false that the Chomsky sentence is false. But if it's false that a sentence is false, then the sentence is true. That is, the Chomsky sentence is *true*!

Now, I'm not endorsing the above argument for the truth of the Chomsky sentence. It's a good argument, but not completely airtight, for reasons that need not detain us. All I'm trying to prove here is a quite modest thesis: *in order to see if a sentence is true, we have to pay very careful attention to its meaning*. When it comes to the Chomsky sentence we know perfectly well that it's a ridiculous, absurd, silly sentence. All you have to know in order to see this is that there *can't* be colorless green things and ideas don't sleep at all, either furiously or not. The surprise is that the Chomsky sentence is also true, or so the above argument reasonably suggested.

The Chomsky sentence illustrates how tricky truth and meaning can be. There are many good questions to ask about truth. Here is a small sample of them:

- What does it mean to say that truth is objective? The claim 'Neptune is colder than Earth' is true, but in what sense is its truth independent of us?
- Does truth come in degrees—can one claim be more true than another? If so, what is that scale of truth?
- How often do we arrive at the truth? That is, how often do our beliefs about the world manage to get things right?
- How often do we know we have got the truth? It could happen that we get lucky and just fall on top of it; but how often do we *know* we have got it?

Those are great questions, no doubt about it. But there is a much more fundamental question, one that philosophers like me worry about even if we don't examine it in classes for undergraduates or go over it in popular philosophy books: are there *any truths at all*, whether they be known or unknown, objective or subjective, total or partial?

The paradoxes I introduce in this essay—the Liar, Grelling's, and the No-No—seem to show that the notion of truth is highly mysterious, *perhaps even contradictory*. They seem to show that the concept of truth is akin to the concept of a naked woman with a blue dress on—it's just incoherent. The idea that truth might be contradictory strikes most people as completely nuts. After all, isn't the most obvious thing in the world that some sentences are true? Surely, sentences such as 'Bill Clinton was the US President in 1996', 'Lady Gaga plays the piano', 'The Earth is larger than any gorilla', and '2 + 2 = 4' are true!

The view that there are no truths whatsoever is *Alethic Nihilism*, since 'alethic' means 'relating to truth'.

There is a remarkable thing about Alethic Nihilism: we know, with perfect certainty, that it's not true. As I will prove below, there is simply *no way in hell* Alethic Nihilism can be true. So why on earth are we wasting time investigating it?

Think about it: Alethic Nihilism says that no statement is true. But ‘There are no truths’ is a statement. If it were true, then what it says would have to be true. It says there are no truths. Thus, if Alethic Nihilism is a truth, then there are no truths. Well, that’s just crazy: we just proved that if Alethic Nihilism is true then it’s not true. And that shows, pretty conclusively, that Alethic Nihilism just can’t be true, since its truth would entail that it’s both true and not true.

That’s somewhat paradoxical in itself. It seems as though Alethic Nihilism *shouldn’t* be contradictory. It seems as though it should be possible for every single statement to fail to be true, because the notion of truth is incoherent. But if *there are no truths*, then why isn’t ‘There are no truths’ true?

I can see a way that might avoid the paradox: if all our sentences are just noise, just babble that is good enough for coordinating our lives but not good enough for truth, then it makes sense that none of them are true. Whenever it seems as though someone is actually saying something coherent—like I am right now, or like how the alethic nihilist does when she articulates her view—appearances are deceptive.

In any case, there are no generally accepted solutions to these paradoxes about truth. They have defied solution for *millennia*. It isn’t hard to find some philosopher or logician X who is confident that he or she knows the right solution. Even today you can find these folks; some of them are my friends. But it’s also very, very easy to find *lots* of experts who are just as accomplished as X and who firmly believe that X’s proposed solution is just plain wrong. Even X will admit this uncomfortable fact about how her theory is judged by other experts.

At first glance, the paradoxes of truth presented below might strike you as being nothing more than amusing brainteasers. That’s a common reaction, one that I encounter when I put the following at the very end of a long true-false test in my introductory philosophy class:

T F	178.	The correct answer to 179 is ‘true’.
T F	179.	The correct answer to 178 is ‘false’.

I tell my students that these two problems don’t count in their test score, but many students are charmed by them and will write smiley faces in the margin of their exam. What they don’t know is that they just got a taste of a problem that for many centuries has produced nightmares in the best mathematicians, logicians, and philosophers of all time.

One is initially inclined to think that they are brainteasers that are so complex that we can’t figure out the solution. One could design a math problem so complex that we don’t have the brainpower to solve it, but this wouldn’t suggest that we are mistaken in what we think about simple mathematical concepts like *number*, *addition*, and *multiplication*. The importance of the paradoxes of truth is that the individual ideas that make up the paradoxes seem to be *truisms*, things that are completely obviously true and couldn’t possibly be false. But if some of those alleged truisms are really false, which is what the paradoxes prove, then that means that *there is something profoundly wrong with our simplest, most* *basic ideas about truth and meaning*. People find that very unsettling. That's probably the main reason why they continue to study them after all these centuries. The mathematical analogy would be finding a knockdown proof that if 2 + 2 = 4, then all penguins can fly. It would make us think that maybe we were wrong that 2 + 2 = 4. And of course that would make us skeptical that we have much of a grasp of what addition and equality even mean.

In order to *understand* the three paradoxes of truth mentioned above—the Liar, Grelling's, and the No-No—before one even tries to get a handle on various proposed solutions, one needs to be comfortable thinking about how lots of interesting sentences talk about not dogs or cats or elections or baseball but sentences. That is, we need to get familiar with analyzing *sentences that talk about sentences*.

Sentences that Talk about Sentences

We can talk about sentences. For instance, we can talk about the following three sentences:

> Snow is white.
> Grass is green.
> The Chicago Cubs will never win the World Series again.

We can say, truthfully, that each of these sentences has no more than ten words. We can say, truthfully, that at least two of them are true. We can say, truthfully, that exactly two of them are about color. That gives us three more sentences we can talk about.

> Each of the above three indented sentences has no more than ten words.
> At least two of the above indented sentences are true.
> Exactly two of the above indented sentences are about color.

Notice that those last three sentences are *sentences about sentences.* They are sentences that talk about not people or dogs or monuments or laptops but sentences. Here are two more:

> All English sentences over five words in length contain verbs.
> All the English sentences in this book are less than twenty words in length.

Whether or not these sentences are true doesn't matter to our purposes. All I want now is for you to see that there are perfectly good, morally upstanding sentences that talk about sentences.

Now notice that we can give *names* to sentences. After all, we can give names to just about anything (people, dogs, monuments, laptops). Maybe your favorite sentence is one from the philosophical theory Existentialism: 'Existence precedes essence'. Suppose you think that that's the deepest, most wonderful sentence ever. You're in love with that sentence. In order to help express your love for it you give it a name. Names are convenient. If I want to talk about my favorite dog, I could either use a phrase like ‘My favorite dog’ or I could just use a name of that dog. The same holds for your favorite sentence, the one from Existentialism. Instead of having to use long phrases like ‘My favorite philosophy sentence’ you can just use the name of the sentence. You decide to call it ‘George’. You could have used ‘Fred’ or ‘Constance’ or ‘Bubble’, but you decided on ‘George’. So, as you use it in some linguistic contexts, ‘George’ is the name of a sentence, and not a person, dog, monument, or laptop. When Laura Bush uses ‘George is simply wonderful and deep’ to talk about her husband the former president of the USA, she is using ‘George’ to talk about a person, and not a sentence, dog, monument, or laptop. When you use ‘George is simply wonderful and deep’ to talk about Existentialism, you are using ‘George’ to talk about a sentence, and not a person, dog, monument, or laptop.

Let’s give a name to the sentence ‘All English sentences over five words in length contain verbs’. Let’s call it ‘Alan’. So, Alan is a sentence, and not a person, dog, monument, or laptop. Now we can note some interesting facts about Alan. One such fact is the fact that Alan is a sentence that is about lots of sentences (naturally, he is a sentence that is about every English sentence over five words in length, and that’s a lot of sentences). Another interesting fact is that Alan is the type of sentence that *makes a claim.* The sentence ‘Please close the door’ is about a door but does not make any claim; instead, it makes a request. The sentence ‘Close the freaking door now you worthless maggot’, as used by some large, nasty, dangerous person who has absolute power over you, does not make a claim or a request; instead, it states a command (well, maybe it makes some claims implicitly: the door might not be so good, as it’s a “freaking” door, and you are a worthless maggot, or you are at least somehow *akin to* a worthless maggot, and in a way that reflects poorly on you). Alan doesn’t make a request or a command; it makes a claim, something that’s either true or false. For our purposes it doesn’t matter whether the Alan claim is true or not; all we care about is the fact that it makes a claim.

Of course, Alan makes a *particular* claim, not all claims. If the dictionaries I used are reliable, it makes the same claim as these Spanish and French sentences:

> Todas las oraciones inglesas sobre cinco palabras en longitud contienen verbos.
> Toutes les phrases anglaises plus de cinq mots de longueur contiennent des verbes.

We can name those two indented sentences, respectively, ‘Carlos’ and ‘Pierre’. Carlos and Pierre are sentences, and not people, dogs, monuments, or laptops. Alan, Carlos, and Pierre are three declarative sentences that all “say the same thing” or “make the same claim” even though they belong to different languages.

The sentence ‘All dogs over fifty pounds will scare Joey’ *applies to*, or “talks” about, many dogs. It applies to all dogs over fifty pounds, and it says of such dogs that they will scare Joey. The sentence ‘Any butler who hates his master will be conflicted’ applies to all butlers who hate their masters. Analogously, Alan, Carlos, and Pierre “apply to” many sentences. They apply to all English sentences over five words in length, and each of Alan, Carlos, and Pierre says that such sentences contain verbs.

Now, finally, please notice that Alan, Carlos, and Pierre all apply *to Alan* as well as to many other English sentences. That is, since Alan, Carlos, and Pierre each apply to *all* English sentences that are over five words in length, and Alan happens to be an English sentence over five words in length, it follows that each of Alan, Carlos, and Pierre apply to (or "talk about") Alan. Notice further that neither Carlos nor Pierre applies to itself, as each says something about *English* sentences only, and neither Carlos nor Pierre is an English sentence.

Thus, we have shown that a perfectly ordinary, morally upstanding sentence (Alan) applies to itself. There is nothing peculiar with a sentence that says something about itself! Of course, Alan says something about sentences other than himself as well. Remember that he says something about *all* English sentences over five words in length, and he is just one of many sentences that satisfy that condition.

Now you might disagree here. You might think that it's okay for a sentence to talk about or apply to *other* sentences, but there is some semantic or other philosophical incoherence in a sentence talking about or applying to itself. Carlos and Pierre are okay, as they apply to all English sentences over five words in length, and neither Carlos nor Pierre is an English sentence. When it comes to Alan, however, you might say that he applies to all English sentences over five words in length *except himself*, as there is something problematic with a sentence that applies to itself. Or, you might say that *if* Alan were meaningful at all, taken as a whole sentence, it would have to apply to absolutely all English sentences over five words in length, including himself; since that's impossible (or so you say), you conclude that Alan isn't really meaningful (although he has meaningful parts put together in a grammatically sound way). I don't agree with that position, but it won't matter in the least for my arguments below. We'll be able to derive a contradiction from your view, as you'll soon see.

Temporarily setting aside the 'sentences can't apply to themselves' view just discussed in the previous paragraph, we can come up with a perfectly ordinary, morally upstanding sentence that says something about *just* itself, and no other sentences. (Recall that Alan talked about himself and lots of other sentences.) As a warm-up, consider this sentence and call it 'Marsha':

> The 10th indented sentence in this essay is Spanish.

Marsha is pretty clearly about Carlos! After all, Marsha talks about the 10th indented sentence in this essay, and that sentence is none other than Carlos. Marsha is about, or applies to, just one sentence, the Spanish sentence Carlos. Furthermore, Marsha is *true*, as she is saying, correctly, that Carlos is Spanish. If Marsha had been 'The 10th indented sentence in this essay is French', then she would have been *false*.

Now using Martha as a guide, we can come up with a sentence that is about, or applies to, just itself. Consider this sentence and call it 'Fred':

> The 13th indented sentence in this essay is over five words in length.

As it turns out, Fred talks about…itself only! Obviously, it talks about the 13th indented sentence in this chapter. And as a matter of odd coincidence, Fred is that very sentence (the sentence below, ‘This very sentence contains more than five words’, is the 14th indented sentence). Furthermore, it seems pretty clear that Fred is *true*. After all, he is over five words in length. If Fred had been ‘The 13th indented sentence in this essay is *less* than five words in length’, then Fred would have been *false*.

Thus, we have apparently shown that a perfectly ordinary, morally upstanding sentence can apply to *just* itself and be true as well. We could have tried to make this point much quicker, with a sentence like this:

> This very sentence contains more than five words.

It seems that that sentence says of itself, and nothing else, that it contains more than five words. It refers to itself by containing the phrase ‘This very sentence’. But I went through the long proof of the existence of self-referential and true sentences (such as Fred) in part so you could get familiar with the phenomenon of sentences talking about sentences, including themselves. Now we move on to a fictional story about the philosopher Plato.

Mystery: The Liar Paradox

Late in life, Plato grows to think that his mentor Socrates was actually pretty dense. Plato is teaching in a classroom. He knows that at that very second Socrates is teaching in the room next door (pretend that Socrates didn’t bravely take the hemlock and die for his convictions). Plato also knows through long experience that Socrates has the following teaching style: he writes exactly one sentence on the whiteboard, one that seems to him to be a very profound truth, and then he spends the rest of class time talking about it. Plato has witnessed this teaching style many times.

Plato wants to convey to his students his low opinion of Socrates. He also wants to mock Socrates’ teaching style. Plato and his students think that Socrates is now teaching in room 101. So Plato does his best impression of Socrates’ voice and mannerisms and then writes on the whiteboard

> The sentence written on the whiteboard in room 101 isn’t true.

(Here we pretend that Plato writes in English and there are whiteboards in ancient Athens.) The students laugh nervously at Plato’s cleverness (‘nervously’ because they can see that Plato is being a jerk). We can call this sentence ‘Plato’s mean sentence about Socrates’. Or, we can call it ‘Plato’s mean sentence’. Or, we can give it a simple name, such as ‘Bubba’. Let’s call it ‘Tom’ instead. So, Tom is a sentence, and not a person, dog, monument, or laptop. Tom is the sentence Plato wrote on the whiteboard. As I said, Plato thinks that Socrates is in room 101. He thinks Socrates is in room 101, he thinks that whatever room Socrates is in will have just one sentence on the whiteboard, and he thinks that that sentence isn’t true; that’s why Plato wrote ‘The sentence written on the whiteboard in room 101 isn’t true’.

Plato and his students think that Plato is teaching in room 100 and Socrates is teaching in room 101. But they’re wrong about that! As it turns out, Plato is in room 101 and Socrates is in room 102. Thus, Tom, the sentence Plato wrote on the whiteboard in room 101, is actually about…Tom! After all, Tom is the sentence ‘The sentence written on the whiteboard in room 101 isn’t true’; so Tom is about whatever happens to be the sentence in room 101; but Tom *is* the sentence in room 101; thus, Tom is about Tom. Plato would have claimed that his sentence, Tom, is about whatever sentence *Socrates* had happened to write on his whiteboard. Maybe Socrates’ sentence was ‘Existence precedes essence’ or ‘Virtue sucks’ or ‘Plato wasn’t as good a student as everyone thinks’. But Plato would have been wrong about that. As it turned out, Tom is a sentence that is about Plato’s sentence, not Socrates’ sentence.

Now Plato clearly *intended* to talk about Socrates’ sentence. If Socrates’ sentence was ‘Virtue sucks’, then that was the sentence Plato was *trying* to pick out when he used the description ‘The sentence on the whiteboard in room 101’. We can even say that in some important sense Plato was talking about or referring to Socrates’ sentence ‘Virtue sucks’, and in that sense he was not really talking about or referring to his own sentence. All of that seems right, but consider just the *literal meaning* of Plato’s sentence. Regardless of his ultimate intentions or purposes, his sentence is about itself and no other sentence. In order to see what the Liar Paradox is, just focus on the literal meaning and set aside the (no doubt important) non-literal meaning(s) of Plato’s sentence.

Earlier we saw that it certainly appears as though the sentence Fred is about Fred and Fred alone. Now we see that it certainly appears that Plato’s sentence Tom is (literally) about Tom and Tom alone. We also saw earlier that Fred is true. Well, is Tom true? Let’s figure it out.

We begin with some elementary observations about language and truth. For instance, if the sentence ‘All dogs are cute’ is true, then all dogs are cute. If the sentence ‘Existence precedes essence’ is true, then existence precedes essence. If the sentence ‘John killed Howard in the drawing room with a bazooka’ is true, then John killed Howard in the drawing room with a bazooka. That’s pretty obvious, right? In general, if you take an English declarative sentence and plug it in for the dots in the following, you’ll end up with a true sentence:

> If the sentence ‘…’ is true, then ….

So, we can fill in the dots with ‘All dogs are cute’, ‘Existence precedes essence’, ‘John killed Howard in the drawing room with a bazooka’, or ‘The sentence written on the whiteboard in room 101 isn’t true’, as each of those four is a declarative English sentence (again, if you think the latter sentence (Tom) is really meaningless or otherwise defective, your view will be adequately discussed below).[1] And when we fill in the dots that way we must get true sentences; here they are:

1. If the sentence ‘All dogs are cute’ is true, then all dogs are cute.
2. If the sentence ‘Existence precedes essence’ is true, then existence precedes essence.
3. If the sentence ‘John killed Howard in the drawing room with a bazooka’ is true, then John killed Howard in the drawing room with a bazooka.
4. If the sentence ‘The sentence written on the whiteboard in room 101 isn’t true’ is true, then the sentence written on the whiteboard in room 101 isn’t true.

But when we look carefully at what happens in case (4), the one involving Tom, we get strange results, as we’re about to see.

First, a crucial but short digression. Recall that George is the sentence ‘Existence precedes essence’. That being so, if the first sentence below is true then of course the second is true as well:

> If the sentence ‘Existence precedes essence’ is true, then existence precedes essence.
>
> If George is true, then existence precedes essence.

Similarly, if Lana is the sentence ‘John killed Howard in the drawing room with a bazooka’, then if the first sentence below is true then the second is true too:

> If the sentence ‘John killed Howard in the drawing room with a bazooka’ is true, then John killed Howard in the drawing room with a bazooka.
>
> If Lana is true, then John killed Howard in the drawing room with a bazooka.

Now the same thing must apply to Tom. That is, if the first indented sentence below (which is just (4) from above) is true—and recall that above we showed that it (i.e., (4) above) *did* have to be true—then the second must be true as well:

> If the sentence ‘The sentence written on the whiteboard in room 101 isn’t true’ is true, then the sentence written on the whiteboard in room 101 isn’t true.

[1] Well, not always. Suppose you’re happy and I’m unhappy. You say to us ‘I am happy’. Now suppose I want to talk about the sentence you just said out loud. I say ‘If ‘I am happy’ is true, then I am happy’. Arguably, this ‘if-then’ sentence of mine is false. The ‘if’ part of my sentence is true, as it was referring to *your* sentence and saying it’s true (which is correct, since you are indeed happy), but the ‘then’ part of my sentence is false. And the reason why the ‘then’ part of my sentence is false is that when I used the word ‘I’ in *that* sentence, it referred to *me*, not you. When I use the word ‘I’ it refers to me; when you use it it refers to you. Words like ‘I’, ‘her’, ‘now’, and ‘here’, which change their referents depending on the context in which they are used, even though their meanings remain constant, are called *indexicals*. Unfortunately, there are no indexicals in Tom that are changing their referents. So the ‘if-then’ sentence involving Tom must be true just like the three others about dogs, essence, and John’s bazooka.

> If Tom is true, then the sentence written on the whiteboard in room 101 isn't true.

Call the immediately above indented 'if-then' sentence T. So T isn't Tom! Those are two distinct sentences. Since we already showed that (4) is true, and we also showed that if (4) is true then T is true too, we know that *T is true*. Don't forget this result.

Sentence T is an 'if-then' sentence. Its 'if' part reads: 'Tom is true'; its 'then' part reads as follows:

> the sentence written on the whiteboard in room 101 isn't true

As you can now plainly see, the 'then' part of T, immediately above, talks about a particular sentence, the one written on the whiteboard in room 101. Of course, that's the sentence Plato wrote, Tom, as Tom is the very sentence in room 101. Thus, the 'then' part of T is equivalent to this:

> the sentence written on the whiteboard in room 101—that is, Tom—isn't true

Or, for short,

> Tom isn't true

So, in sum, the ***true*** sentence T comes to this:

> If Tom is true, then Tom isn't true.

Now that's a pretty weird sentence (a bit like how the Chomsky sentence is weird, but in a different way). It says that *if* a certain sentence is true, then it would have to be both true *and* not true. So sentence T is saying that if you *assume* that Tom is a true sentence (that's the 'if' part of T), then you reach a crazy conclusion: Tom would have to be true and not true. So the assumption leads to a contradiction.

Well, that seems to show pretty conclusively that the assumption is false, so Tom is *not* true. After all, *if he were true,* then he would be both true and not true. That's nuts. So, he must not be true.

No paradox yet. In fact, it's totally unsurprising that Tom isn't true! Tom is a screwy sentence! Surely he won't be *true*! He barely even seems coherent!

Be careful to not confuse T and Tom: these are two entirely distinct sentences, one true (that's T) and the other … well, we are in the process of figuring that out. Here's Tom:

> The sentence written on the whiteboard in room 101 isn't true.

And here's T:

> If Tom is true, then Tom isn't true.

We proved that T is true. *There's nothing wrong with T*. What T says, however, is that if Tom is true then he's not true as well. That doesn't show that *T* is false; it shows that *Tom* is screwy. T is a perfectly fine, morally upstanding, friendly, attractive, sensual, and true sentence that shows that Tom isn't true.

So, we've proven that *Tom is not true*. Tom is a sentence that talks about just Tom and Tom is not true. Fred is a sentence that talks about just Fred and Fred is true. Thus, Tom is not as fortunate as Fred: only one of them is true. You might even say that Tom is really meaningless, in some strong sense of 'meaningless', and so he doesn't really talk about anything at all, himself or anything else. Or maybe he doesn't "make sense". Or maybe even this: he isn't really a sentence at all, even though he's made of perfectly good English words ordered in an apparently grammatically sound way. Some people react that way to Tom; they say the things I just wrote. That's fine; I won't argue for or against any of those ideas. All I want you to admit now is the simple result that *Tom isn't true*, independently of whether he has various kinds of meaning deficits or other linguistic problems. Naturally, if you think Tom isn't even meaningful, you'll agree that he isn't *true*. You'll be happy with the argument thus far. No paradox yet.

But if Tom isn't true, as we just confidently concluded, what does that mean? Well, Tom is the sentence Plato wrote on the whiteboard in room 101. After all, you'll recall that Tom is by definition the sentence Plato wrote, and the sentence Plato wrote was on the whiteboard in the room Plato was in, viz. room 101. And we just said that Tom isn't true. Thus, since Tom is that whiteboard sentence, and Tom isn't true, it follows that the whiteboard sentence, the one in room 101, isn't true. That is, we have proven that *the sentence on the whiteboard in room 101 isn't true*.

Now please look at the italicized sentence, call it 'X', from the last part of the previous paragraph that we just *proved* to be true: X is true. Hmm…haven't we seen that sentence X before? Yes, of course we have! It's just Plato's mean sentence about Socrates! X is the sentence he wrote on the whiteboard! X appears on Plato's whiteboard *as well as* at the end of the previous paragraph of this chapter: it's the very same sentence written in two places. So, Plato's mean sentence about Socrates, X, is true, as we just proved it in the previous paragraph. Obviously, Plato's mean sentence is none other than Tom. That is, Plato's mean sentence, Tom, or X, is true. X = Tom. That is, Tom is true.

Oops. Tom is *true*? Earlier we confidently concluded that Tom is *not* true. So he's true and not true—which is a contradiction (a contradiction is a sentence of the form 'P and not-P')? That's the ***Liar Paradox***: we just gave what seemed to be a rigorous proof of a contradiction. Our job now is to figure out what went wrong in the seemingly perfect argument that led to the contradiction. Or, alternatively, we need to explain how some contradictions (e.g., "Sentence Z is both true and not true") can be true.

If you are a stickler for having every step of the argument laid out explicitly, then you'll appreciate the fact that we can set out the paradox in step-by-step fashion. Here is one way to spell out the first part of the argument:

1. Suppose for the sake of argument that Tom is true.

Recall that Tom = 'The sentence written on the whiteboard in room 101 isn't true'. So it follows from (1) that the sentence 'The sentence written on the whiteboard in room 101 isn't true' is true. On to the next premise:

2. If you take an English declarative sentence and plug it in for the dots in the following, you'll end up with a *true* sentence: If the sentence '…' is true, then ….

We saw this principle before, as it generated true sentences such as 'If the sentence 'All dogs are cute' is true, then all dogs are cute'. This time, however, we don't plug in 'All dogs are cute' but Tom, the sentence 'The sentence written on the whiteboard in room 101 isn't true'. When you do so, you get the following sentence, which according to (2) is our next true premise:

3. By the principle in (2), if the sentence 'The sentence written on the whiteboard in room 101 isn't true' is true, then the sentence written on the whiteboard in room 101 isn't true.

Summing up the step-by-step argument thus far, by (1) we have it that this is true:

> 'The sentence written on the whiteboard in room 101 isn't true' is true.

And by (3) we have it that this is true:

> If 'The sentence written on the whiteboard in room 101 isn't true' is true, then the sentence written on the whiteboard in room 101 isn't true.

If you put both of those results together (via the inference rule known as "modus ponens") you get the result that this is *true*:

> The sentence written on the whiteboard in room 101 isn't true.

But the sentence on that whiteboard is just Tom. So, Tom isn't true.

Hence, what we have done thus far is prove that if we assume that Tom is true (which is how we started, in (1)), then we are forced to conclude, by the reasoning above, that Tom is not true as well. So the assumption in (1) leads to a contradiction: the contradiction that Tom is both true and not true. So that assumption in (1) must have been false. Since the assumption was 'Tom is true', and we just proved that that assumption is false, we can conclude that Tom is not true.

Now we can proceed to the next stage of the argument, which starts off exactly where the first stage ended:

4. Tom isn’t true.

Recall that Tom is the sentence written on the whiteboard in room 101. Hence, by (4), the sentence written on the whiteboard in room 101 isn’t true. On to the next premise:

5. If you take an English declarative sentence and plug it in for the dots in the following, then you’ll end up with a true sentence: If …, then ‘…’ is true.

For instance, if you plug in ‘All dogs are cute’ for the dots, you end up with the if-then sentence ‘If all dogs are cute, then ‘All dogs are cute’ is true’—which of course is true. The principle in (5) is the reverse of the principle in (2) above. In the next premise we plug in not ‘All dogs are cute’ but ‘The sentence written on the whiteboard in room 101 isn’t true’ into (5). Here is what we get:

6. By the principle in (5), if the sentence written on the whiteboard in room 101 isn’t true, then ‘The sentence written on the whiteboard in room 101 isn’t true’ is true.

Summing up, by (4) we know that this is true:

> The sentence written on the whiteboard in room 101 isn’t true.

And by (6) we know that this is true:

> If the sentence written on the whiteboard in room 101 isn’t true, then ‘The sentence written on the whiteboard in room 101 isn’t true’ is true.

Thus, if you put both of those results together (via modus ponens) you get the result that this is true:

> ‘The sentence written on the whiteboard in room 101 isn’t true’ is true.

But ‘The sentence written on the whiteboard in room 101 isn’t true’, which we just said was true, is none other than Tom. Thus we just showed that if we start out with the assumption that Tom isn’t true—that was premise (4)—we end up with the result that Tom is true. That is, we have proven that if Tom isn’t true, then he is true.

So that’s the disaster: if Tom is true, then he isn’t true; and if he isn’t true, then he is true. No matter what you say about Tom—he’s true, he isn’t true—you end up contradicting yourself. That’s a paradox.

In order to have a rational and informative response to the paradox, one must either (a) show *exactly where* in the above argument mistakes are made, or (b) explain how some contradictions can be true. And even that isn't enough: you have to prove that the alleged mistake really is a mistake. Let's face it: no one cares what *you* think about the liar paradox; all we care about is what you can *prove* about it.

Let me emphasize this last point, as many people miss it. ***In order to have a "response" to the paradox a person must take a stand on each of (1)-(6), saying whether each is true; and whenever he or she says that one of (1)-(6) is false, he or she must justify that opinion***. Whatever you find yourself inclined to say about the paradox, when you're finished articulating it ask yourself this: which premise is false and why? If you can't answer that question in detail, then you don't really have a response to the paradox at all. This holds for ***all*** paradoxes.

Mystery: Grelling's Paradox

In scientific work one tries to get a lot of data, so that one's theory is thoroughly tested. We do the same thing in philosophy: when looking for a theory of X (where X can be just about anything), we gather as many test cases as possible related to X. Grelling's Paradox is similar to the Liar Paradox. In order to reveal the paradox we start with some simple reflections on language.

We start with a linguistic stipulation. In the rest of this chapter I'm going to use the word 'phrase' to pick out *any* string of *any* linguistic symbols whatsoever. The string can be just one symbol long or a billion symbols long. It need not make any sense at all! So each of the following counts as a distinct phrase because they are not perfectly identical:

> Go to bed
> Go to bed..
> Go to bed…….
> Bed go go to
> 48593j d9wf 845792
> blah halb 87 *&(*&(___________ufhushs

I could have used 'string of symbols' instead of 'phrase', but the former is long and unwieldy.

Now consider the phrase 'is a giraffe'. This phrase is true of, or applies to (I'm using 'true of' and 'applies to' as synonyms here) lots of things. Those things are, unsurprisingly, giraffes. For instance, if Janice is a giraffe, then 'is a giraffe' is true of Janice; 'is a giraffe' doesn't apply to you because you aren't a giraffe. The phrase 'is a giraffe' applies to all and only giraffes. The phrase 'is over 500 kilograms in mass' applies to my car, like most cars, but it doesn't apply to me, or my kids, or my laptop: my car is over 500 kilograms in mass but none of those other things are.

We can do something similar with other phrases, ones that apply not to animals or heavy objects but bits of language. Consider ‘is a noun’. This phrase is true of the word ‘dog’ because the word ‘dog’ is a noun. The phrase ‘is a noun’ isn’t true of Fido the dog, as no *dog* is a noun, but it is true of ‘dog’. The phrase ‘is a noun’ is true of lots of words. Naturally, ‘is a noun’ is true of, well, every single noun in every single language. Another example: ‘is a verb’ is true of ‘run’ but not ‘Homer Simpson’, as the latter isn’t a verb but the former is a verb.

We don’t want to confuse words with what they stand for. For instance, Einstein was a scientist, but ‘Einstein’ is a *name* of a scientist. Einstein was a person; ‘Einstein’ is not a person but a word, in this case a name. ‘Giraffe’ is not a giraffe; it’s a word (in this case, it’s a noun).

Now consider ‘is a phrase three words long’. This phrase is true of ‘agonize threat cookie’, ‘Roberta is happy’, and ‘muggle wand Voldemort’ since each of those phrases is three words long. It is not true of ‘Roberta is happy and hungry’ or ‘agonize’ because they aren’t three words long.

Now consider ‘is a phrase six words long’. This phrase is true of ‘Roberta is happy, hungry, and tall’ and ‘muggle wand Voldemort is akin sleep’. Interestingly, it is also true of ‘is a phrase six words long’, or so it certainly appears. That is, ‘is a phrase six words long’ is true of ‘is a phrase six words long’. That is, ‘is a phrase six words long’ is *true of itself*. Similarly, ‘is a phrase’ is true of itself. So is ‘is an English predicate’ and ‘contains a verb and a noun’. On the other hand, ‘is 500 words long’ isn’t true of itself because it’s much less than 500 words long. Similarly, ‘is a giraffe’ isn’t true of itself because, well, ‘is a giraffe’ is a bit of English and certainly not an animal of any kind.

Call the phrases that are true of themselves *autological.* So, ‘is a phrase six words long’, ‘is a phrase’, ‘is an English predicate’, and ‘contains a verb and a noun’ are all autological, at least apparently (you might think that no phrase can really be true of itself; that’s fine, the paradox will apply to you anyway). ‘is a giraffe’ isn’t autological. See the pattern:

> ‘is a phrase six words long’ is a phrase six words long
> ‘is a phrase’ is a phrase
> ‘is an English predicate’ is an English predicate
> ‘contains a verb and a noun’ contains a verb and a noun

Each of those four indented sentences is *true*; that’s what it *means* to be autological. In general, the rule is this: a phrase P (e.g., P might be ‘is a phrase six words long’) is autological exactly when the following is a true sentence when we plug P in for the dots (as we did for the four indented sentences immediately above):

> ‘…’ …

That’s the *defining test* for autologicality. Plugging in ‘is a phrase six words long’ for the dots gives us the true sentence

> ‘is a phrase six words long’ is a phrase six words long

That’s why ‘is a phrase six words long’ is autological.

Now consider another idea: call a phrase *heterological* just in case it is *not* true of itself. So a phrase is heterological exactly when it’s not autological; the term ‘heterological’ is perfectly synonymous with ‘not autological’. So, ‘is ten words long’ is heterological. So is ‘is a giraffe’. See the pattern:

> ‘is a giraffe’ is a giraffe
> ‘is ten words long’ is ten words long
> ‘is a good friend’ is a good friend

None of those indented phrases is true; that’s what it means to be heterological. In general, a phrase P (e.g., ‘is ten words long’) is heterological just in case the following is *not* a true sentence when we plug P in for the dots:

> ‘…’ …

That’s the *defining test* for heterologicality. Plugging in ‘is a phrase ten words long’ for the dots gives us a false sentence; that’s why ‘is a phrase ten words long’ is heterological. Plugging ‘swamp muggle Cheney’ or ‘the ggdsi 55 **&&’ in for the dots gives you these two phrases:

> ‘swamp muggle Cheney’ swamp muggle Cheney
> ‘the ggdsi 55 **&&’ the ggdsi 55 **&&

These aren’t even sentences, let alone true sentences. So ‘swamp muggle Cheney’ and ‘the ggdsi 55 **&&’ are heterological as well! What ‘swamp muggle Cheney’, ‘the ggdsi 55 **&&’, ‘is a giraffe’, ‘bumbletomm here crumbletomm’, and ‘is ten words long’ have in common is the fact that none of them gives you a true sentence when you plug them in for the dots; that’s what it *means* to be heterological.

To sum up: if some weird person hands you a bit of language and asks you ‘Is this bit of language heterological or autological?’ you have an easy way to answer their question, a method that has two steps. Here are the two steps:

Suppose for example the bit of language in question is ‘Fred thinks laptop’. The *first* thing you do, step 1, is mechanically construct the following longer bit of language:

> ‘Fred thinks laptop’ Fred thinks laptop

Now the *second* thing you do, step 2, is figure out if you have come up with a true sentence: if you have, the original bit of language is autological; if you don't have a true sentence (because you have a false sentence or maybe just gibberish), then the original bit of language is heterological. That's all there is to it. This autological-heterological game may seem pointless, but we can see how to play it anyway.

To finally reveal Grelling's paradox, pretend that some weird person (such as myself) hands you the phrase 'is heterological' and asks you whether *that* bit of language heterological or autological. Well, is it true of itself, in which case it is autological? Or is it not, in which case it's heterological?

As instructed above, the *first* step in answering those questions is to mechanically construct the following bit of language:

> 'is heterological' is heterological

and see what happens. Call that immediately above indented sentence (or sentence-like thingie) 'Harry'. So Harry is a sentence (or bit of language), and not a person, dog, monument, or laptop. The *second* step is to figure out if we have constructed a true sentence. So, is Harry a true sentence?

Suppose for the moment that he is, and then we'll see where this assumption leads us. Clearly, since Harry is a true sentence (as we've just assumed for the sake of argument), he must say something (that's true), and just as clearly what Harry says (if anything) is that the phrase 'is heterological' is heterological. So, since Harry is the sentence that says 'is heterological' is heterological, and Harry is true (as we just supposed), it follows from our supposition that 'is heterological' is heterological. Thus, *if* Harry is true *then* 'is heterological' is heterological. We don't know that Harry really is true; all we've done is see what would happen *if* he were true.

We just said that if Harry's true then 'is heterological' is heterological; don't forget that result. But earlier we said that the defining test for heterologicality was this: a phrase P is heterological if you *don't* get a true sentence when you plug P in for the dots in

> '...' ...

Since we *did* get a true sentence when we plugged in 'is heterological' (we got the true Harry), and a phrase is heterological only when you *don't* get a true sentence, 'is heterological' fails to pass the defining test for being heterological. So it *isn't* heterological.

Thus, we have figured out this: *if* Harry is true, then 'is heterological' is both heterological (that was our "result" from before that I asked you not to forget) and *not* heterological (as we just proved in this paragraph), which is incoherent. Our assumption that Harry is true has led to a contradiction, the contradiction that 'is heterological' is both heterological and not heterological. Thus, our assumption that Harry is true must have been incorrect; so, Harry isn't true. No surprise there: Harry is just another crazy sentence.

So apparently Harry is not true (maybe he's meaningless or not even a sentence; it won't matter). Now recall that Harry is the sentence, or bit of language

> 'is heterological' is heterological

So we're saying that Harry, i.e.,

> 'is heterological' is heterological

isn't a true sentence. Well, then that proves that 'is heterological' passes the defining test for heterologicality! After all, a phrase is heterological when you *don't* get a true sentence when you plug it in for the dots in

> '…' …

We plugged 'is heterological' in and got Harry, who we just said was not a true sentence. Thus, since 'is heterological' passes the test for heterologicality, we have just proven that it's *true* that

> 'is heterological' is heterological.

Well, that indented, *true*, sentence is just Harry again. So Harry *is* true after all, even though we already concluded that he *isn't* true. Excuse me while I blow my brains out in frustration….

That's Grelling's paradox. It does not have to do with truth directly, as the Liar does. Instead, it has to do with what's called *linguistic satisfaction.* When we say that 'is a giraffe' *applies to* or *is true of* Janice the giraffe, we are saying that she *satisfies* the phrase 'is a giraffe'. Similarly, Fido the dog satisfies 'is a dog' and the noun 'laptop' satisfies 'is a noun'. Just as in the case of the Liar Paradox our job with respect to Grelling's paradox is to discover the error in the argument for the contradiction; alternatively, we have to explain how contradictions (e.g., "X is both true and not true") can be true.

Mystery: The No-No Paradox

We move on to the No-No paradox. Yes, that's its real name nowadays (although it didn't go by that name a few centuries ago). We begin by considering some non-paradoxical sentences.

Suppose Fred knows that George is in the next room talking about quidditch. Suppose further that Fred thinks that George is an idiot when it comes to quidditch. As a result of that unkind opinion Fred says 'Everything George is saying about quidditch is false'. And suppose George said only the following three things about quidditch while he's been in that room:

The Chudley Cannons finished in first place last year.
Harry is an excellent seeker.
Hermione knows nothing about quidditch.

If Fred's remark 'Everything George is saying about quidditch is false' is true, then three things follow immediately: the Chudley Cannons didn't finish in first place last year, Harry isn't an excellent seeker, and Hermione knows something about quidditch. There's nothing paradoxical going on there. It is commonplace to express an opinion regarding the truth or untruth of someone else's opinion, and that's exactly what Fred is doing: expressing his opinion regarding George's opinions. But we can formulate some sentences that do just that and lead to paradox.

Return to Plato and Socrates. They are teaching in adjacent rooms, just like before. (This time no one is confused about which room people are in.) Socrates has the same teaching style as described earlier. Plato has adopted that teaching style. Each has a very low opinion of the other's philosophical acumen. In order to impress on his students how stupid Socrates is, Plato writes on his whiteboard the disrespectful

Socrates' sentence is not true

Call Plato's sentence 'Paul'; so the sentence on Plato's whiteboard is Paul. In order to impress on his students how stupid Plato is, Socrates writes on his whiteboard the disrespectful

Plato's sentence is not true

Call Socrates' sentence 'Sara'. In effect, Paul says that Sara is not true and Sara says that Paul is not true.

Right away there is something fishy here. In order to figure out if Paul is true, we need to look at what he says. Naturally, he says that Socrates' sentence isn't true. Okay, so in order to figure out if Paul is true we'll have to figure out if Sara is true.

Well, is Sara true? In order to figure out if Sara is true, we need to look at what she says. Naturally, she says that Plato's sentence isn't true. Okay, so in order to figure out if Sara is true we'll have to figure out if Paul is true.

It doesn't take a genius to see that we're never going to get anywhere in figuring out if Paul or Sara is true, as we'll just be going around in a circle forever. But the problem isn't that we're too stupid to figure it out: it seems that *even God* couldn't figure out whether Paul or Sara is true. The sentences go around in a circle as it were. Paul has no real meaning because he's trying to *acquire* his meaning from Sara; but Sara is trying to get her meaning from Paul! It seems that neither Paul nor Sara has any truth-value at all; the sentences are defective. Paul and Sara are absurd sentences in the sense that they don't really say anything at all in spite of not breaking any rules of grammar.

But if the sentences are so defective that they aren't true, then we're saying that Socrates' sentence isn't true and Plato's sentence isn't true. But isn't that *exactly* what Paul and Sara are saying!? Just look at them! And doesn't that mean that both of them are true after all??

That's the first part of the No-No paradox: the two sentences seem both true and not true. There's a second part as well.

- Suppose *Sara isn't true*. Well, since Paul says that *Sara isn't true*, then it sure looks like Paul is true. Thus, *if* Sara isn't true *then* Paul is true.
- Now suppose *Sara is true*. Well, since Sara is true and she says that Paul isn't true, Paul isn't true. Thus, *if* Sara is true, *then* Paul isn't true.

In sum: Sara is true exactly when Paul isn't true. That is, Sara and Paul have to have *opposite* truth-values, which of course conflicts with our argument above that neither one is true. In fact, it's a complete mystery how they could have opposite truth-values, as they seem to be on a par. Look at them once again: how could one of them be true while the other is false?

If you prefer brevity, here's the simplest case of the No-No paradox:

(A) Sentence (B) isn't true.
(B) Sentence (A) isn't true.

You should be able to figure out on your own how both '(A) is true' and '(A) is not true' lead to absurdities.

So much for the No-No paradox. What about solutions to our paradoxes?

The Commonsensical Solution

You're going to have to take my word on one matter: there isn't some *simple* mistake being made in the reasoning about the Liar, Grelling, and No-No paradoxes. It's just not true that there is some silly little mistake being made somewhere, a mistake that an intelligent and careful reasoner could uncover if she put her mind to it. After all, if there were such mistakes, then wouldn't they have been revealed by now by the many mathematicians, logicians, and philosophers who have been studying the paradoxes for so many years? Recall that in the case of the Liar paradox, it's been investigated for literally millennia.

Even so, there is a certain approach to the alethic paradoxes that is highly intuitive and for that reason alone never goes away. It's the solution that just about every smart person thinks of at some point when puzzling over the paradoxes. Roughly put, it's the idea that the paradoxical sentences don't really say anything: although they look like coherent claims, they are not. In order to understand this thesis we need to keep in mind the *type/token distinction*, which I'll go over now.

Here is a justly famous sentence:

> The quick brown fox jumped over the lazy dog.

The sentence is famous because it uses up all the letters of the English alphabet in a relatively compact manner. How many words are in that sentence?

There are two correct answers: nine and eight. The word 'the' appears twice, so you can either count it once or twice and you'll be right either way: the *single* word has *two* occurrences. The same thing holds in many contexts. Suppose that in a parking lot there are fifty Toyota Camrys and fifty Toyota Corollas. In one sense, there are two cars in the lot; in another sense there are one hundred. We disambiguate by saying that there are two *types* of cars in the lot, fifty *instances* of each type. The same holds for the fox/dog sentence: it has eight word types in it; seven of the types have just one instance each; one of the types has two instances. Call the instances *tokens*. So in the following sentence,

> The dog ran stupidly to the cliff while the wise owl and horse took one look at the cliff and backed away.

There are 22 word tokens and 17 word types. We can apply the type/token distinction to whole sentences as well. Consider this paragraph:

> Nikhet has fourteen pairs of shoes now. The best ones, the ones that are her favorite, are black boots with many clasps. She gets a lot of compliments on those boots! Last year she had just thirteen pairs of shoes, but she got some inexpensive flip flops to wear when walking to the nearest grocery store. Nikhet has fourteen pairs of shoes now.

There are two tokens of the sentence type 'Nikhet has fourteen pairs of shoes now' in that paragraph, one at the beginning and one at the end.

Some sentence tokens are printed on a page; they are made of ink. Other sentence tokens are spoken; they are acoustical disturbances in the air. Others are made of plastic and metal, such as the famous 'Hollywood' sign in California. Yet others are complicated hand gestures, as in sign language.

Imagine a conversation in which the father says, on two occasions, 'Alec is starving', where Alec is his four-year old son who is fidgeting in the backseat of the car. The first time he says it Alec protests that he isn't starving at all; he's upset because his sister Julia is bothering him. So the father's first sentence token was false. Later, Alec is acting up again and the father says, once again, 'Alec is starving'. This time he's right: Alec is starving and that's why he is being difficult in the backseat of the car. The second sentence token is true. So there were two tokens of the same sentence type, one true and one false.

This phenomenon of dual sentence tokens with different truth-values will be particularly important below.

As I said earlier, there is a proposed solution to the alethic paradoxes that utilizes the type/token distinction. It's the idea that the paradoxical sentence tokens are defective, and thus not true, even though they certainly look like ordinary grammatical sentences. If any approach to the paradoxes deserves the title 'commonsensical', this is it. There is a great deal to recommend this approach, especially when it carefully and consistently distinguishes sentence tokens from sentence types, claiming that while one token of a sentence type might fail to be true, a distinct token of the same type might be clearly true—just like in the case of Alec and his father. For one thing, the approach requires no significant alteration in commonsensical views about language or logic; many other theories have to jettison common sense in several places. Let us call it the *Token Approach*, as it trades on carefully distinguishing tokens from types.

The Token Approach's Treatment of the Paradoxes

I start by showing how the Token Approach deals with most of the standard alethic paradoxes, especially the Liar, Grelling's, and the No-No. Since we already introduced those paradoxes, this will involve some repetition; but the paradoxical sentences will be different enough to make it worth our while.

The following indented token string of self-referential symbols A,

> A isn't true

presents a problem. Suppose it's true. Well, then since it's true it must say something true. All it says, if anything, is that *A isn't true*; so I guess it must be the case that *A isn't true.* Thus, we just showed that if A is true (that was our supposition), then A isn't true. But that's silly: if A is true, then it can't be *not true* as well!

So maybe A isn't true; our supposition was false. That seems reasonable. Sentence A is ridiculous; it could hardly be true. A isn't true.

But just now with the immediately previous sentential token string of symbols from the previous paragraph, call it B, we concluded an apparently sound argument with a token of the sentence-type 'A isn't true'. How could we be right in writing *that* token string, B? That is, how could our token string B, at the end of the previous paragraph, be true? After all, token B is *exactly* like token A (the same words, the same meanings, the same order, the same sentential structure, etc.) and earlier we claimed that A *isn't* true. How can B be true while A isn't true? It can't. So, we have reached a contradiction.

At this point it's tempting to conclude that this shows that there is something seriously wrong with logic, or the expressive capacities of languages such as English, or the notion of truth, or something else equally dire. That's what most contemporary philosophers of logic conclude at this point. But there seems to be a much more attractive and seemingly less calamitous way out: although tokens A and B are composed of the very same words with the very same meanings in the very same order plugged into the very same sentential structure, they don't really say the same thing. In fact, whereas B says something perfectly true, A says nothing at all. Sentence tokens A and B differ in some way that generates the difference in meaning—although that way is hidden if one just looks at grammar and sentential structure (which is what most philosophers of logic do).

That's overstated a bit: token A has perfectly meaningful parts (e.g., the words 'isn't' and 'true') arranged in a perfectly standard sentential way. Token A is meaningful in *some* senses of 'meaningful': for instance, it has meaningful parts and a sentential structure. So it is wrong to say A isn't meaningful, as it has plenty of meaning. What it doesn't have, which B does have, is a *truth condition*; it fails to make a claim. It fails to make a claim even though it is meaningful in other senses of 'meaningful'.

The same idea applies to the No-No paradox. Consider this pair of token strings of symbols, the first one C and the second one D:

> D isn't true
> C isn't true

It seems as though C refers to D and D refers to C. On the face of it, by symmetry these two tokens should have the exact same semantic properties (each is doing nothing other than saying 'no' to the other). If one is true, the other has to be true too; they're on a truth-value par at the very least.

But suppose for a moment that C is true. If C is true, then it must say something true. If C says something, it says that D isn't true. So, since C is true and it says that D isn't true, it must be the case that D isn't true. So, D isn't true. In sum: *if* C's true, *then* D isn't true. Thus, if C is true, then C and D *differ* in truth-value. But that can't be right: in the previous paragraph we said that C and D don't differ in truth-value. Hence, we were wrong when we made the assumption, at the beginning of this paragraph, that C is true (as that assumption led to the false conclusion that C and D differ in truth-value). So that assumption was false. C isn't true.

But look at the token sentential string I just finished typing after giving my apparently sound argument; call that token string E. I have suggested to you that E is true; that's what I just proved in the previous paragraph. But E is *exactly* like D; same words, same meanings, same order, same structure, etc. So, since E is true and E and D are exactly the same, D must be true as well. But D can't be true! We said that C and D should have the same truth status, so we can't say, as we just said, that C isn't true while D is true.

We have the same solution here as before: although tokens E and D are exactly alike linguistically, they differ in that only the former makes a claim. This is the Token Approach to the alethic paradoxes, as it focuses on linguistic tokens, claiming that two tokens of the same (apparently non-indexical) sentence can have different truth conditions. We can use the Token Approach for many other paradoxical cases, like string of symbols F:

> F doesn't make a claim

Token string F is just as screwy as strings A, C, and D. F doesn't make a claim. But now look at my immediately previous sentential token string, G. I have suggested to you that G is true, but since it's exactly like F, F must be true too—or so someone might think. But no: just because F and G are exactly alike doesn't mean they have the same truth condition. The same treatment handles other, more clever sentence tokens such as 'This sentence token has no truth condition'.

F and G are tokens of the same linguistic type; the same holds for the pairs A/B and D/E. We want to say, in response to these alethic paradoxes, that the distinct tokens have distinct truth-conditions even though they are of the same type: although A, F, and D fail to have a truth condition, B, G, and E do and are true.

Here is a version of yet another paradox: Curry's paradox. Suppose the only thing written on a whiteboard is a token of the type 'If every sentence token written on the whiteboard is true, then 1 + 1 = 1'. Call that token H. Suppose H's antecedent is true. If it's true, then since H's antecedent says that every sentence token written on the whiteboard is true, it follows that every sentence token written on the whiteboard is true. But there's only one such token and that token is H. So, H is true. And we supposed that H's antecedent is true. Thus, by modus ponens its consequent is true. So, 1 + 1 = 1 (clearly, we could put anything in for '1 + 1 = 1', thereby "proving" anything). Since that's wrong, our initial assumption that H's antecedent is true must have been wrong. So, H's antecedent isn't true. If it's not true, then of course it's not true that every sentence token written on the whiteboard is true. Since there is just one sentence token written there, H, H isn't true. But earlier we said that H's antecedent isn't true. And because H is a conditional that means that H is true. Contradiction!

But just because H's antecedent isn't true doesn't mean H is true. Since H has no truth condition, it isn't true; so we don't care that its antecedent fails to be true as well. If there had been another sentence token on the whiteboard, no paradox need have resulted, which just shows that a sentence's semantic properties are dependent on surprising external factors.

So far, so good. But what have I just argued? I have argued that the one sentence token written on the whiteboard (i.e., H) isn't true. Since there is just one such sentence token, I've argued that *every* sentence token written on the whiteboard isn't true. Thus, I'll be happy asserting this: if every sentence token written on the whiteboard is true, then 1 + 1 = 1. (I know the antecedent is false, so I'm willing to put anything in the consequent of my material conditional.) And although the sentence token after the colon in the previous non-parenthetical sentence token—call it I—is type-identical to H, it is true while H isn't. That's the Token Approach again.

In the interests of being thorough and of seeing once again the importance of distinguishing types and tokens, let's take a look at Grelling's paradox. If some person hands you a linguistic *type* and asks you 'Is this linguistic type heterological?' you can answer their question in two steps. Suppose the bit of language in question is the type 'Fred thinks laptop'. Step one is this: consider the following longer linguistic type:

> 'Fred thinks laptop' Fred thinks laptop

Step two is this: figure out if the longer linguistic type is a true sentence type. If you *don't* have a true sentence type (because you have (a) a sentence type that is false, (b) a sentence type that is neither true nor false, or (c) a linguistic type that isn't even a sentence type), then the original linguistic type is heterological. If you *do* have a true sentence type, then the original linguistic type is non-heterological. Thus, the types 'is a giraffe' and 'noun' and 'giraffe' are heterological while the types 'contains English letters' and 'is written in a natural language' are non-heterological.

Now suppose you are asked to determine whether the type 'is heterological'—call that linguistic type M—is heterological. As instructed above, the first step is to consider the following longer type, call it N:

> 'is heterological' is heterological

Now suppose for reductio that N is true. Since N is a true sentence type, it must say something that's true, and it's clear that what N says (if anything) is that the type 'is heterological', M, is heterological. Thus, we have result R1: *if* N is true *then* M is heterological.

But when applied to M the defining test for heterologicality is this: M is heterological if and only if you *don't* get a true sentence type when you construct a type according to the recipe in step one above. The type constructed in step one was just N. Thus, M is heterological if and only if N isn't a true sentence type. Thus, we have result R2: if M is heterological, then N isn't true (that is half of the biconditional of my previous sentence). But R1 and R2 together entail that if N is true then N isn't true. Thus, our supposition that N is true was mistaken.

Thus, N isn't true. But M is heterological if and only if

> 'is heterological' is heterological

isn't true. But that's just N, which we said wasn't true. So M definitely passes the test for heterologicality. M is heterological. That is, we have just proved that

> 'is heterological' is heterological.

But that's just N! So N is true after all! Contradiction.

The Token Approach advocate accepts the preceding argument all the way up to but not including ‘So N is true after all!’ All the argument up to that point proved is that *a* token of N is true; it did not prove that *all* tokens of N are true. Indeed, earlier on we saw that some tokens of N aren’t true. Thus, type N isn’t true or false because it has true tokens and non-true tokens.

The lesson from these arguments is this: whenever you read anything regarding the alethic paradoxes, always always *always* insist on knowing whether the sentences under discussion are tokens or types. If the author fails to disambiguate (which is the usual state of affairs), then you have the task of doing it yourself and evaluating the resulting multitude of interpretations.

Of course, there are puzzling sentence types that don’t seem to generate contradictions. For instance, suppose the only linguistic token on the whiteboard in room 101 is ‘All the sentence tokens on the whiteboard in room 101 are true’; call that token O. Clearly, the Token Approach will claim these tokens fail to have truth conditions. The evidence backing up this assessment is different this time, though. To see this, suppose we have two tokens, T1 and T2, where T1 is in room 1 and says ‘All the sentence tokens in room 2 are true’, and T2 is in room 2 and says ‘Snow is white’. T1’s truth condition is determined by that of T2, assuming they have truth conditions at all. So it is easy to see how the truth condition of one token *comes from* that of another. It is clear that token O is pretty similar to token T1. And yet, while T1 does get a truth condition, via T2’s truth condition, O fails to get one. Again, this is the commonsensical thing to say about “truth telling” tokens such as O. Most any non-philosopher will assert that ‘This sentence is true’ is just as “screwy” as ‘This sentence is false’: neither one really says anything at all. Naturally, one can just say that token O is true without encountering a contradiction, but there is no good reason to do so. Just because such a position generates no contradiction supplies no good reason to think it’s true! The obvious thing to say about O, or about related sentences such as ‘This sentence token is true’, is that they have no content, they fail to say anything.

Indeed, we have already seen what the token approach advocate says about liar-type tokens that don’t generate contradictions. Consider again tokens C and then D:

D isn’t true
C isn’t true

If we go by the rule ‘Does it generate a contradiction?’ then all we can conclude is that one of C and D can have a truth condition, but not both. We could say that D is true but C has no truth condition. No contradiction results. The same holds for saying that C is true and D has no truth condition. But the commonsensical thing to say is that both C and D are screwy: neither has a truth condition, neither “says” anything at all. Whether or not the Token Approach is successful, no one can deny that it’s intuitive.

Needless to say, the Token Approach would be wonderful if it were true, as it requires no revision to commonsensical views of language and logic. It merely requires keeping careful attention to type-token distinctions and an open mind about contextual factors. However, the Token Approach seems to have two serious problems, as I'll now show.

The Context-Dependence Objection

Here is the first problem with the Token Approach: it needs a detailed and convincing explanation of *how* two apparently non-indexical tokens of the same sentential type can differ in truth conditions even though they don't differ in individual word meanings or order. Consider 'The sentence token on the right side of the whiteboard doesn't make a true claim'. When someone puts a token T1 of that sentence type on the right side of the whiteboard (and there's no other sentence token over there or anywhere else on the whiteboard), then we have a paradoxical sentence token. The Token Approach advocate will end up saying, in articulating her response to this token sentence, a token sentence T2 of the very same type as T1: same words, same order, same structure, same meanings. How on earth could those two tokens T1 and T2 differ in truth-value given their sameness in all relevant semantic properties? Surely they have the same truth condition. Call this first problem the *Context-Dependence Problem.*

In response to this problem, someone could say that the tokens differ in truth conditions because they are context-dependent in some *hidden* way. Of course, the two tokens T1 and T2 contain *non-hidden* context-dependent terms (e.g., 'right side'), but those context-dependent terms have the same meaning and referent in each token (regardless of how we understand 'meaning' and 'referent'), which means that this non-hidden context-dependency makes it unsuitable for generating a difference in truth conditions for the two tokens. We need to posit a *hidden* context dependency that accounts for the crucial truth-conditional difference in T1 and T2.

By my lights, this hidden-context-dependency hypothesis on behalf of the Token Approach should strike most philosophers as insufficiently grounded. There certainly is *nothing whatsoever* in the constituents or form of the two tokens that suggests there is some hidden context-dependent term or other linguistic element that accounts for why the one token T2 has a truth condition while the problematic token T1 does not. Is there any *independent evidence* that there is a hidden context-dependent element that accounts for the difference in truth condition? At this juncture the Token Approach advocate could argue as follows:

> If T1 had a truth condition, then as the simple argument above shows we would reach a contradiction. So, T1 must not have a truth condition even though T2 does. And that means that either T1 or T2 must have a hidden context-dependent element. Maybe we didn't see the context-dependent element before, but I just gave a rock-solid argument that there *has* to be one there anyway. This is just an *indirect* proof of the context-dependency, as I have yet to locate the dependency, but it's a proof all the same.

One should admit that that *looks* like a fine argument. But it has two serious problems that in my view make it inadequate.

First, this argument-speech assumes, in its first conditional premise, that no other solution to the paradox is forthcoming (short of dialetheism, the view that some contradictions are tru). This is not an ideal premise. Second, the argument obviously fails to give us even the slightest clue as to what the hidden contextual element is that is generating the difference in truth condition. If we had some good independent linguistic evidence that T1 or T2 were covertly context-dependent *and*, what is more, that that context-dependence accounted for the truth-conditional difference in T1 and T2, then we would be in business. Note that it does no good to just point out *other* pairs of type-identical token sentences that differ in truth conditions due to peculiar context-dependent elements that the sentences don't wear on their sleeves. For instance, epistemologists like to think that different tokens of 'Fred knows the bank is open on Saturdays' can have different truth conditions depending on various fairly subtle and hidden context dependencies. The same holds for less controversial sentences, such as 'Michael Jordan is tall'. That's fine, but we need to see the covert context-dependency *in T1 or T2* or we are just whistling past the graveyard.

In response to the two counterarguments of the previous paragraph, the advocate of the Token Approach can start out by saying that her solution is the best one anyway. After all, look at the extreme measures embraced by the alternatives! We could conclude that there are true contradictions (dialetheism)…or we could just say that there are some context dependencies that are hard to locate. We are already very familiar with subtle context dependencies (again, see the cases from epistemology and the philosophy of language generally), so why not just suppose that there are some more of them? It seems like an awfully reasonable thing to conclude, given our very limited knowledge of semantics coupled with the drastic nature of the alternatives (dialetheism isn't the only alternative response to the paradoxes, but I assume that most philosophers will agree that all of them seem highly counterintuitive at points, even if as a matter of empirical fact level-headed philosophers can learn to live with them).

This should strike one as a reasonable response to the Context-Dependence Problem, at least when construed as an *initial* response to it. But once again, I don't think it's anywhere near adequate: there is an argument that it is *no longer* a reasonable response. Suppose I come home from work but an hour later I can't find my keys. I seem to vividly remember putting my keys in my desk, so I look there. After eight hours of searching every nook and cranny of the desk, I can't find the keys in the desk. At this point, I've got to give up the key-are-in-the-desk hypothesis, even though that certainly was the most reasonable hypothesis initially. In the case of the alethic paradoxes we have examined all sorts of context sensitivity in language and we are aware of nothing at all that can handle the whiteboard sentence case. So, even though the Token Approach should get our vote as the hypothesis that is *initially* most reasonable, it has lost that happy status.

The Revenge Objection

Now for the second problem with the Token Approach: it appears to fail for some paradoxical sentences of the liar family. It works fine for the so-called "strengthen liar", 'This sentence is not true' (the "ordinary liar" is just 'This sentence is false'), but it appears to fail for others. Consider string token X:

> No token of X’s type makes a true statement.

Now: does X make a true statement (or “express a true claim”, or however one wants to put it)? Suppose it does; call this our “initial assumption”, which I’ll refer to below. Since it makes a true statement we can ask what true statement it makes. Apparently, it makes the true statement that no token of X’s type makes a true statement. Thus, it’s true that no token of X’s type makes a true statement. That is,

> No token of X’s type makes a true statement.

Call the immediately previous indented string token Y. We just proved that it makes a true statement, relying on the assumption that X makes a true statement. And yet it’s a token of X’s type. Thus, there is a token of X’s type that makes a true statement. Thus, it’s not the case that no token of X’s type makes a true statement. But this directly contradicts what we concluded with the true Y. So, our “initial assumption” that X makes a true statement must have been mistaken. X doesn’t make a true statement. Fine; it’s a screwy sentential string so it’s no surprise that it doesn’t make a true statement. No paradox yet.

Now consider token Z indented below and bear with me a moment while I repeat some of what I said above (the repetition is needed in the argument below):

> No token of X’s type makes a true statement.

Suppose for a moment that token Z makes a true statement. The true statement it makes, if any, is that *no token of X’s type makes a true statement*. Thus, it must be true that no token of X’s type makes a true statement. And yet, Z makes a true statement and is a token of X’s type (just look at Z and X; clearly they are of the same linguistic type). Hence, it’s not the case that *no* token of X’s type makes a true statement, as we just found one (i.e., Z) that makes a true statement. We have reached a contradiction, based on the assumption that Z makes a true statement. Thus, that assumption must have been wrong. Thus, Z doesn’t make a true statement. Again, this is no surprise, as Z is screwy.

Things start to get interesting when we realize that Z is an *entirely arbitrary* token of X’s type. And we proved, in the immediately preceding paragraph—call it the *proof paragraph*—that it doesn’t make a true statement. In fact, we could use that paragraph-proof for *any* token of X’s type, any one at all! That is, whenever you give me a token T of X’s type, I can just slap down the proof paragraph to show that T doesn’t make a true statement (all I have to do is replace ‘Z’ with a name for T in the paragraph-proof). Here are three more tokens of that type:

> No token of X’s type makes a true statement.
> No token of X’s type makes a true statement.
> No token of X’s type makes a true statement.

We can generate three proofs for those three tokens of X's type that show that none of those tokens makes a true statement. You give me a token of X's type, and I slap down a copy of the proof paragraph, thus showing that that token doesn't make a true statement. Thus, we can conclude from all these proofs that *no token of X's type makes a true statement* (since for each token there is a corresponding proof that it doesn't make a true statement). That is, we have just proved that

> No token of X's type makes a true statement.

Call that immediately preceding token string S. S makes a true statement, as we just *proved* it. But obviously it's a token of X's type. Thus, at least one token of X's type—namely, S— makes a true statement. But that contradicts the statement made by S, which we proved to be true! Sigh. Call this the *Revenge Problem.*

Here is a potential way around the Revenge Problem: although the little argument for the truth of S looks impeccable, it's really flawed because S doesn't make a true statement. Although the proof for S is rock solid in *some* sense, paradoxically enough its conclusion, S, has no truth condition at all. This is an odd result, for sure, but we can live with it because we have always known that the alethic paradoxes were going to have strange but true consequences.

But it seems that I should be able to argue inductively: I've proven that *every* token of X's type that I've seen thus far fails to make a true statement; I have an absolutely perfect proof-schema that I can apply to *any* token of X's type that will definitively prove that that token doesn't make a true statement either; thus…. And of course what comes after the 'thus' should be 'No token of X's type makes a true statement'. Why can't I state the conclusion of my sound proof so that it's true (i.e., makes a true statement)? We are supposed to swallow the idea that a valid argument with all true premises isn't sound?

Conclusions

What we have done is this: we have discovered that some very simple, seemingly clearly true claims about truth lead straight to contradiction.