Truth I: Introduction; and Correspondence

Webpage: https://rjh221.user.srcf.net/courses/truth/

Our aim here is not to say which things are true; nor to say how we know what is true; the first of those is a task for science, the second for science and epistemology. It's also not directly to answer a certain sort of sceptic who denies that there are any truths, or who denies that we should be concerned with believing them, though these are good philosophical questions (on them see Bernard Williams *Truth and Truthfulness*).

Rather, it is to give an account of what truth is: of what the property of truth (if there is one) consists in, or of what we mean by the predicate 'is true' or the operator 'It is true that'. There are different ways that we could do this. We might give an explicit, eliminative definition, but we might give something weaker, an axiomatization, or an even more circular sketch. Once we have this, it may help us with the other tasks of legitimizing or justifying, and perhaps of showing how we can know things to be true; or it may show those tasks to be misconceived. But that isn't our primary task here.

A second task, that has dominated much discussion of truth in the last 100 years or so, is to ensure that the account of truth given enables us to say something about the liar:

(1) Sentence (1) is false

which some thinkers have taken to show that the intuitive notion of truth is somehow flawed. We'll come to this in Lecture 4.

Standard accounts of truth have followed various lines: correspondence accounts (a sentence or proposition is true iff it corresponds to the facts); coherence accounts (a sentence is true iff it coheres with … what? other sentences? other true sentences? or iff it form part of the largest coherent block); pragmatic accounts (what is true is what we will believe at the end of inquiry; or what it is somehow useful to believe; or something like that); and minimalist, or redundancy, or disquotational accounts (there is nothing more to truth than that given by the equivalence of 'p' and '"p" is true' or something similar). We will look at these in turn.

CORRESPONDENCE THEORIES

Traditionally seen as foreshadowed by Aristotle: 'To say of what is that it is not, or of what is not that it is, is false, while to say of what is that it is, or what is not that it is not, is true.' (*Metaphysics* 1011b25) That sounds ok as far as it goes, but it's a little unclear how far it goes. Russell and Austin take things further, but unfortunately there the issue get tangled up with other things: with an account of belief in the case of Russell, and with an account of the illocutionary acts that are performed by truth talk in the other. We can use them though to illustrate one important distinction, that between theorists who think of correspondence

holding at the level of facts (what Kirkham calls 'correspondence as correlation'; Austin is a prime example), to those that want to decompose the sentences and facts into parts and then think that they correspond to each other at the level of the parts (what Kirkham calls 'correspondence as congruence'; Russell is a prime example).

Let's step back a bit. Correspondence theories take truth to consist in correspondence. So to say that a certain statement is true is *not* to say the same thing that would be said by that statement (though it may entail that); it is to say that the correspondence relation obtains. This is the focus of what Strawson denies in his response to Austin; and it is also, as we'll see in a couple of weeks, what the minimalist denies. But it is not clear what we should say about this issue directly. So let's look in more detail at the substance of the correspondence accounts.

Two obvious questions:
(i) What is the correspondence between? Sentences? Statements? And Facts? The World?
(ii) What is correspondence?

WHAT IS THE CORRESPONDENCE BETWEEN?

Taking the first of these first; if we take it to be between sentences and facts, can we individuate the facts properly? An argument against comes from the *Slingshot Argument* as given by Davidson, 'True to the Facts'.

Two assumptions: two descriptions 'the fact that p' and 'the fact that q' denote the same fact if (i) 'p' and 'q' are logically equivalent; or (ii) if 'p' and 'q' differ only in that a singular term occurring in 'p' has been replaced with a co-referring singular term in 'q'.

Then by (i), (1) and (2) denote the same fact since they are logically equivalent:

(1) Snow is white
(2) (The x, such that x is Putin and snow is white) = (The x, such that x is Putin)

and by (ii) (2) and (3) denote the same fact:

(2) (The x, such that x is Putin and snow is white) = (The x, such that x is Putin)
(3) (The x, such that x is Putin and grass is green) = (The x, such that x is Putin)

and then, running the initial argument in reverse, (3) and (4) denote the same fact:

(3) (The x, such that x is Putin and grass is green) = (The x, such that x is Putin)
(4) Grass is green

Then by transitivity of identity, (1) and (4) denote the same fact, so there is only one fact.

Is the argument valid? Let's focus first on the move from (2) to (3). Is the idea the Fregean one that 'snow is white' and 'grass is green' both denote the same thing, namely the true. That would certainly make the move from (2) to (3) valid, but in the context of this argument it is

surely question begging. Won't the proponent of fine-grain facts think that they denote different things, namely different facts?

Instead Davidson's claim seems to be that '(The x, such that x is Putin and snow is white)' and '(The x, such that x is Putin and grass is green)' are both singular terms, denoting Putin, and hence one can be substituted for the other.

But these are definite descriptions. On a Russellian account, these, of course, are not singular terms; and there is no reason to think that substituting co-denoting definite descriptions as Russell understands them will give rise to sentences that express the same fact. Surely the fact that the winner of the 2008 US presidential election was Obama is different to the fact that the winner of the 2012 election was Obama, even though those two descriptions are co-denoting.

Suppose alternatively that we reject the Russellian account, and treat descriptions as singular terms. The problem then is that Davidson's other principle, that logical equivalents express the same fact, seems questionable. The move from (1) to (2), which would be valid on the Russellian account now looks to be a move that does change the subject: from a claim about snow being white, to an identity statement about Putin. So Davidson's concern here doesn't' seem very worrying.

[*For those who are really keen on this stuff.* Stephen Neale shows (following Gödel) that a somewhat better slingshot argument can be constructed if one replaces assumption (i) with the weaker assumption (i*) that a description will refer to the same fact if "Fa" is substituted for "a = the x such that (x=a and Fx)" and *vice versa* (see Neale *Facing Facts;* for a good summary see John MacFarlane's review of the book in the on-line *Notre Dame Philosophical Reviews*). But there are still many ways of understanding facts that reject either (i*) or (ii) or both. For instance, this is the case for most truth maker theories: every true sentence must be made true by something, i.e. a fact. Such accounts have their problems (how do we deal with disjunctive facts, or negative existentials?), but the central point here is that they do not seem to be theories of truth at all. Rather, they use a prior understanding of truth to provide a constraint on what makes sentences true.]

WHAT IS CORRESPONDENCE?

Move now to the second issue, that of how to understand correspondence, and to a more general worry: is our grip on a sentence *corresponding* to a fact any different to our grip on it being *made true* by the fact? If not, have we actually got anywhere? For now we are presupposing the notion of truth in defining correspondence, so it cannot be used to define truth without going in a circle.

We might wonder whether we should take the idea of correspondence so seriously. Austin suggests it shouldn't be taken too seriously. Kirkham, endorsing this, gives a useful schema which he thinks can be applied to different correspondence accounts:

(t) t is true iff $(\exists x)((tRx) \& (x \text{ obtains}))$

Here the variable 'x' ranges over facts or states of affairs; the variable 't' and the relation R will be filled in differently by different theorists. So for Russell 't' ranges over beliefs, and R is something like 'is the belief that'. For Austin, 't' ranges over sentences, and R is something like the 'means that' relation.

But still: what is it for a fact to obtain? That sounds like a somewhat technical notion; how might we explain it? To say that the fact that grass is green obtains is presumably to say that grass is green; and isn't that to say that it's true that grass is green (with a truth operator) or that 'grass is green' is true (with a truth predicate)? Isn't there a worry that we have just pushed the bump in the carpet a little further along? Rather than defining truth in terms of something else, we have, in a circular way, defined it in terms of itself.

One possible response is to turn from approaches that think of the correspondence holding at the level of facts (recall this is what Kirkham calls 'correlation'), to those that want to decompose the facts into parts, and then think that sentences correspond to facts at the level of the parts (recall this is what Kirkham calls 'congruence'). Perhaps if we approach things at that level, we won't get this circularity.

Hartry Field suggest just this: rather than looking to the correspondence of true sentences to the facts, looks for the correspondence of names to things, and of predicates to properties, and build an account of truth out of that. (See his 'Tarski's Theory of Truth'). To see this we'll need to look at Tarski's account; but that is fitting anyway, since in many places (though not all) Tarski describes himself as a correspondence theorist.

TARSKI

Tarski held that a good definition of truth must meet two conditions. To understand the first, we need to distinguish between the language whose truth predicate we are talking about, and the language that we are giving the theory in; call the first 'the object language' and the second 'the metalanguage'. We'll see the need for this later when we come to look at the liar paradox: in brief, Tarski held that the only way to avoid paradox is to insist than no language can contain its own truth predicate.

Now we can state Tarski's first condition on a good definition of truth: it must be 'materially adequate': it must generate, in the metalanguage L_1 each instance of the T-schema for the object language L_0:

X is true in L_0 iff P

where 'X' is replaced by a name of a sentence in L_0, and 'P' is replaced by a translation of that sentence into the metalanguage. The metalanguage may be an extension of the object language (it may be the object language together with the truth predicate of that language) but it need not be. So, for instance, while we will get T-sentences like:

'Snow is white' is true in English$_0$ iff snow is white.

which look trivial, we will also get sentences like:

'La neige est blanche' is true in French$_0$ iff snow is white.

which is clearly not trivial. In fact the first sentence is not trivial either; it just looks it because we know English.

Secondly, the definition must be 'formally correct'. This is a harder notion to get clear on. For a start the definition must contain, on the right hand side, no undefined semantic notions. If truth is a dubious concept, the only way to legitimate it is to define it in less dubious terms. For Tarski, who defined himself as a physicalist, this meant that ultimately truth should be defined in purely physical terms.

A DEFINITION OF TRUTH FOR A FINITE LANGUAGE L_A

If there were only a finite number of sentences for which the truth predicate were defined, this would be fairly straightforward. Consider a language L_A that contained only the sentences:

snow is white
grass is green
snow is green
grass is white

then we could define truth for that language as follows:

S is true in L_A iff
S = 'snow is white' and snow is white
or S = 'grass is green' and grass is green
or S = 'snow is green' and snow is green
or S = 'grass is white' and grass is white

The right hand side of the definition doesn't contain any semantic terminology: it doesn't contain any mention of truth or reference or anything like that; it refers only to notions of snow, grass, white and green. (These may not be physicalistically acceptable; but if not, then they should be reduced to physicalistically acceptable notions, or banished from the language; at any rate, it isn't the notion of truth that is causing the problem.) Of course, it doesn't tell you *which* of these sentences is true; but a theory of truth isn't meant to teach you what colours things are. And again although it may look trivial, that is only because the metalanguage and the object language are the same. Suppose that the object language was a fragment of French, and the metalanguage English. Then this would tell you something substantial about French.

But our languages aren't finite: whilst the basic elements are, they can be used to generate an infinite number of sentences. So no list will do the job. Moreover, lists leave out important generalizations. For instance, we want our theory of truth to register the fact that a

conjunction will be true iff both of the conjuncts are true, that the predicate 'is white' applies to white things, and so on.

A DEFINITION OF TRUTH FOR AN INFINITE LANGUAGE L_B WITH PREDICATES AND TERMS

Non-logical vocabulary
The one-place predicates 'is white' and 'is green'
The terms 'snow' and 'grass'

Logical vocabulary
'&'

Sentences

If P is a predicate and t is a term, ‹Pt› is a sentence

If A and B are sentences, ‹A & B› is a sentence

Denotation
'snow' denotes snow
'grass' denotes grass

Application
The predicate 'is white' applies to an object iff it is white
The predicate 'is green' applies to an object iff it is green

Truth

An atomic sentence ‹Pa› is true in L_B iff the predicate ‹P› applies to the object denoted by ‹a›

A sentence ‹A & B› is true in L_B iff A is true in L_B and B is true in L_B

Tarski's actual definition was for a language containing quantification. It was here that his real technical achievement was made, and it makes things considerably more complicated. But philosophically the basic ideas remain the same. For those wanting to get a sense of it, I recommend Soames *Understanding Truth* but I will not discuss it here.

DOES TARSKI'S ACCOUNT MEET HIS OWN CONDITIONS?

These definitions meet the material adequacy condition: it is clear that every T-sentence for the language L_A and L_B can be derived from this definition (this is a bit less obvious for a quantified language; see Soames p. 73 for a discussion).

Do they meet the formal correctness condition? That is controversial. Certainly truth is defined in terms of denotation and application, and they do not have ineliminable semantic vocabulary on the right hand sides of their definitions. But they are defined by means of a list (this is especially clear in the definition of L_B).

Is that to understand the notion of denotation? Hartry Field ('Tarski's Theory of Truth') argues that it isn't. We need an understanding of denotation that allows us to go beyond a simple list; that allows us for instance, to apply it to new cases (to know whether to put them on any list). Following Kripke, Field suggests that we should apply something like a causal theory of names to understand denotation; a term denotes an object iff there is the right sort of causal chain between them. We might say the same about predicates: for a predicate to pick out a property is for there to be the right sort of causal chain between them.

But even if we can define the appropriate causal chains in the right way (something that Kripke himself was sceptical of), the problem remains for the idea of *application*, which is also defined by means of a list. This is the heart of Field's complaint. His parallel: suppose we had an account of the chemical notion of valence that said:

> E has valence n iff E is potassium and n = +1, … or E is sulphur and n = -2 etc.

We wouldn't think that was adequate to explain valence. We still need to say what it is for the notion of valence to apply to an element. Field seems to think that something like a causal account can help us with this, but it is quite unclear how. We are back with the same worry that we had, when talking about about what it is for a fact to obtain. The natural thing to say is that a property P applies to an object a just in case Pa; that is, just in case it's true that Pa, or 'Pa' is true. But then we have just come round in a circle.

We are left with two obvious possibilities. Either we can try to say something more substantial about what truth is. That, in effect, is what the pragmatists and coherence theorists do. Or we can give up on the idea that there is a substantial notion of truth at all. That is why attempts to give one always end up bringing us round in a circle. That is the minimalist approach.

(There is a further worry: what does Tarski's account tell us about truth in natural language? Tarski thought that he had captured the intuitive idea of truth, but he insisted that his notion only applies to artificial, hierarchical, languages, to avoid the risk of paradox. But if it doesn't apply to natural languages, is it our intuitive notion after all? We'll come back to this when we look at the Liar.)