# 1

# Truth

*Jeffrey Ketland*

## Introduction: What Is the Philosophical Problem of Truth?

We each hold various beliefs, and assert various statements and propositions, on matters mundane, historical, scientific, and so on. One feature of such beliefs, statements, and propositions is that they may be *true* or *false*.[1] But what exactly does it mean for a statement, belief, etc., to be *true*? Intuitively, the *truth* of a statement consists in its representing the world correctly, or in the world being as the statement says it is. As explained below, this is a formulation of the *correspondence theory* of truth. However, philosophical questions immediately flood in. A preliminary question concerns the *kinds* of things that may be true or false: beliefs, claims, opinions, assertions, etc.[2] Call these *truth bearers*. (A truth bearer can be false, of corse.) Despite their apparent diversity, there are plausibly two basic kinds: *linguistic items* (e.g., statements, sentences) and *propositions*.[3]

What are the main philosophical issues here? First, it seems clear that *truth matters to us*. It is important to us whether the information we hear (or accept) is true or false. Much can turn on this. We postpone trying to answer why truth matters until the final section. A second question often causes confusion for the beginning student: how do we *determine* if a statement is true or false? Are there general *procedures* or *criteria* for determining whether a statement is true or false? Does truth consist in being *justified* in the right way? Call this the *Epistemological Question*. This question is *not* what philosophers intend when they discuss the philosophical problem of truth. Rather, they are interested in what it *means* to say, of a belief, statement or proposition, that it is true. That is, how is the concept of truth to be *analyzed*? Call this the *Analytic Question*. This question goes back to Plato and Aristotle, and earlier, and has ramifications throughout modern philosophy, affecting debates about the nature of existence, knowledge, meaning, reference and valid reasoning. In general, the epistemological question is a more difficult one to answer than the analytic question. The bulk of the discussion below focuses on the analytic question: what does "true" *mean*?

Next, we get a little clearer about what a definition is. A *definition* of a concept, or a word, is usually given by specifying conditions for that concept, or word, to apply to things. For example, we may define "bachelor" as follows:

A person x is a bachelor if, and only if, x is an unmarried adult male human.

We may also call this an *analysis* of the concept of being a bachelor. Analogously, we might look for a definition of truth, of the form:

(D) A is true if, and only if, . . . A . . .

where *A* is a truth bearer, and " . . . *A* . . ." indicates some condition that *A* satisfies.

Below, we assess a number of proposed definitions of truth, each listed as ($D_1$), ($D_2$), ($D_3$), etc. A point to bear in mind, however, is that one might reject the demand for such a definition, or analysis, of truth. For perhaps truth is simply a primitive, indefinable concept.

## Correspondence Theories of Truth

A correspondence theory begins with common-sensical formulations of the following sort: a statement is true just if it *agrees with reality*; or *represents reality as it is*; or *things are as it says they are*; or *says of what is, that it is*; or *designates an existing state of affairs*; or *corresponds to the facts* (or *to a fact*). These are, so to speak, correspondence locutions, and they seem to fall into two main kinds:[4]

($D_1$) A is true iff A says that such-and-such is the case, and such-and-such is the case.
($D_2$) A is true iff A corresponds to a fact.

The first of these, ($D_1$), is a descendant from Aristotle's formulation, "to say of what is, that it is, is true" (*Metaphysics*). We call it the *classical correspondence definition*. It forms the basis for the semantic conception of truth, developed by Alfred Tarski. The second, ($D_2$), has a long history too, and we call it the *correspondence-to-fact definition*. For the purposes of this section, we concentrate on ($D_2$), although the reader should bear in mind that ($D_1$) is an acceptable, and perhaps preferable, formulation of the correspondence theory.

The correspondence-to-fact definition says that truth involves a *correspondence relation* between truth bearer and a *fact*. Unless we are prepared to treat the notions of *correspondence* and *fact* as basic and primitive, it remains to elucidate them further. What is a fact?[5] Usually, truth bearers and facts are taken to be distinct kinds of entities.[6] Consider the statement "Edinburgh is north of London." Since it is true, the corresponding fact might be something like *Edinburgh's-being-north-of-London*. This is a "complex," whose constituents are Edinburgh, London, and the relation *North-of*. In modern parlance, such complexes are called *states of affairs*. Not every state of affairs is a fact; for some states of affairs *obtain*, some do not.[7] This leads to a definition of "fact" as "state of affairs that obtains." Then, the correspondence-to-fact view becomes:

($D_3$) A is true iff A corresponds to a state of affairs that obtains.

Thus, the statement "Edinburgh is north of London" is true iff the corresponding state of affairs (with its constituents, Edinburgh, London and *North-of*) obtains. Assuming the towns Edinburgh and London to be mind-independent entities, the truth of "Edinburgh

is north of London" depends on mind-independent reality. This is an attractive feature of a correspondence view, in that it permits truth to depend upon mind-independent reality.

If facts are obtaining states of affairs, what is *correspondence*? Correspondence might be understood as "conventional correlation" of truth bearer and state of affairs (c.f., the conventional correlation of *green-light signals* and *permission to proceed*, on foot or in one's car). For example, the sentence "this is a cat" is correlated with states of affairs that involve the presence of a cat, in the vicinity of the speaker; a statement made using this sentence, by a particular speaker in a certain context, is true just when one such state of affairs obtains in the vicinity of the speaker. This view takes account of the presence of context-sensitive expressions, such as "this," "here," "I," in speech acts. A difficulty with this view, however, is that it gives no indication how states of affairs get correlated with context-*in*sensitive claims, like "neutrinos lack mass" or "the French Revolution occurred in 1789."

Another view treats correspondence as a kind of "picturing" relation. To illustrate:

| | *Name* | | *Predicate* | | *Name* |
|---|---|---|---|---|---|
| Truth bearer: | "Edinburgh | | is north of | | London" |
| [correspondence] | $\Downarrow$ | | $\Downarrow$ | | $\Downarrow$ |
| State of affairs: | Edinburgh | $\rightarrow$ | *Being-north-of* | $\rightarrow$ | London |
| | *Object* | | *Relation* | | *Object* |

In a sense, the sentence and the corresponding state of affairs have the same "logical structure." The names in the sentence refer to the objects in the state of affairs (and the predicate in the sentence refers to the constituent relation). The truth bearer, in some sense, *pictures* its corresponding state of affairs, in analogy with how a map pictures, or represents, some region of territory. Truth bearers are therefore *representations* of reality.

A standard objection to the correspondence-to-fact view is that it leads to skepticism, the doctrine that reality is unknowable (for example, perhaps we are brains-in-vats, but don't realize it). To avoid skepticism, one might urge that in order to *know* a fact, one must be able to *directly cognize* the fact. But *mind-independent facts* seem so different from or mental states that we could never achieve this cognitive feat.

An obvious reply to this is that the correspondence view is a theory of *truth*, not knowledge. It answers the analytic question, not the epistemological question. A second reply is that, in any case, the correspondence theory itself seems to play no role in the argument for the unknowability of *facts*. The objection equally condemns or ability to know any *mind-independent* objects whatsoever: cats, stones, trees, electrons, comets, etc. Finally, note that the doctrine that cats, stones, trees, etc. (and facts about Edinburgh and London) are mind-independent is *not* assumed by the correspondence theory itself. Definitions ($D_2$) or ($D_3$) don't imply that the states of affairs are mind-independent. The correspondence theory is thus logically neutral on such questions.

Another objection is simply to that one should repudiate facts altogether. Facts, understood as obtaining states of affairs, are very unlike ordinary physical things: chairs, cups, rocks, fish, etc. One might accept the existence of the towns London and Edinburgh, and perhaps even the abstract relation *North-of*. But is there a further entity, *Edinburgh's-being-north-of-London*? Perhaps, fact-*talk* is merely a convenient manner of speech. Instead of "I am aware of the fact that *p*," one can say, "I am aware that *p*." Instead of "it is a fact that *p*," we say simply "*p*." The repudiation of facts need not entail that one cannot

make sense of truth. For example, the semantic conception of truth, discussed below, was presented as a correspondence theory, but avoids postulating facts, or sentence-to-fact correspondence.

If repudiation of *all* facts goes too far, maybe repudiation of *some* of the weirder ones is wise. For the correspondence-to-fact view requires a specific fact for *every* truth. Consider "London is not north of Edinburgh," which is true. If this corresponds to a fact, it must be *London's-not-being-north-of-Edinburgh*: some sort of "*negative*" fact. Is there such a thing? True statements containing "*not*" are just the start of the trouble, for there are compound statements containing "*or*," "*and*," "*if-then*," "*for all*," "*there is*," "*it is necessary that*," "*believes that*," and so on. When such a statement is true, is there always a fact? If "Sherlock Holmes does not exist" is true, is there a corresponding fact, the *non-existence-of-Sherlock-Holmes*? A way round this problem is to assume corresponding states of affairs only for the simplest sentences (the *atomic* sentences). Then, truth for the compound sentences (built up from these atomic sentences using "*not*," "*and*," etc.) may be defined using a "recursive definition," similar to the kind pioneered by Tarski.[8]

## Epistemic Theories of Truth

Around the turn of the twentieth century, several authors criticized the correspondence view and proposed to define truth in terms of some notion of *idealized justification* or *idealized rational acceptability*. Call these views *epistemic* theories of truth. Such views answer the analytic question (what is the meaning of "true"?) via a *prior* answer to the epistemological question. So, we point out first that there are various criteria we use to select which statements to accept and which to reject. These criteria involve observation, reasoning, mathematical proof, and so on. And second, it is proposed that the analytic question be answered by saying that a statement's *truth* consists in its *meeting these criteria*. The first point is not the point of dispute here, as it concerns matters of epistemology, not the definition of truth *per se*. The second is. For why should the fact that a statement meets certain epistemic criteria *imply* its truth? And why should its failing to meet these criteria imply its falsity?

The simplest criterion involves the statement's being *justified* (e.g., being supported by observational evidence). However, introductory epistemology corses explain that one must distinguish between a statement's being *true* and its being *justified* to some degree. Justified statements and beliefs are sometimes false; and there are truths we have no justification for believing. For example, we lack any justification for believing that Plato sneezed on his 30th birthday, and we also lack any justification for believing that he *didn't*. But logic alone tells us that *either* he did sneeze *or* he didn't sneeze. Hence, either the proposition, or its negation, is true. So, there is a truth we have no justification to believe. Hence, there is a gap between truth and justification. So, defining "A is true" as "A is justified" doesn't work. Still, might we hope to define truth in terms of criteria involving *idealized* justification? There is a baffling variety of such proposals, but we focus on three:

($D_4$) A is true iff A is verifiable, in principle.
($D_5$) A is true iff A belongs to the maximally coherent system (of beliefs).
($D_6$) A is true iff A would be accepted, in the ideal limit of rational inquiry, by anyone who investigates.

To use the technical jargon, ($D_4$) expresses *verificationism about truth*; ($D_5$) expresses the *coherence theory of truth*; and ($D_6$) expresses *long-run pragmatism*.

It is perfectly reasonable to accept statements that are justified either by observation or by logico-mathematical reasoning. So, "there is margarine in the fridge" is justified by observation – i.e., *looking* inside the fridge. "The period of a pendulum varies with the square root of its length" is justified by performing certain experiments. "There are infinitely many primes" is justified by a mathematical proof. Such justification procedures are examples of "verification." Definition ($D_4$) says: *A* is *true* just when it can be "verified" in such a way.

Such a view, however, has severe problems.[9] We *cannot* verify, by direct observation, the statement "the period of a pendulum varies with the square root of its length," for it is a *generalization*, and thus requires indefinitely many experiments. Furthermore, some statements accepted on the basis of observation are *mistaken* (consider the Müller-Lyer illusion). Furthermore, the view implies that *all* truths can, in principle, be verified. But perhaps there are statements, mathematical, scientific or historical, which are true but unverifiable, even in principle. For example, "Plato sneezed on his 30th birthday." Either this or its negation "Plato did *not* sneeze on his 30th birthday" is true, but is either verifiable?

The problem may be that the evaluation criteria are too restrictive. Presumably rational inquiry involves more than just direct, sensory, observation and logical reasoning. Perhaps the statements we ought to accept, as we proceed in inquiry, should form a *holistic, coherent, system*. Thus, though we cannot directly verify the pendulum law, it nonetheless *coheres* with the experiments we have done, and with further background laws of physics. So, perhaps the property of idealized justification we seek is this: *being an element of the maximally coherent system of beliefs* (or *statements*). The coherence definition ($D_5$) says: a belief (or statement) is true just when it belongs to this system. The intended notion of a maximally coherent system is not merely that of a complete consistent system, from logic.[10] Maximal coherence is meant to involve a richer property, wherein all the various beliefs or statements *mutually support one another*.

On a pure coherence theory, to be true is to belong to the maximally coherent system. The standard objection, from Russell, is that it is difficult to see why a maximally cohering system of beliefs is different from a complete, consistent, highly coherent *fictional story*. Complete consistent stories may contain falsehoods and omit truths. No matter how coherent Jane Austen's *Pride and Prejudice* may be, do we count its statements as true? In general, a proposition's belonging to a maximally coherent system need not entail its truth; and, conversely, its being true needn't entail its belonging to a maximally coherent system.

In order to deal with this objection, the most obvious thing is to include observational criteria, thus combining the coherence theory with verificationism. But still there are problems. Even if my *present* system of beliefs, conditioned by experience, is as coherent as possible, future experience may lead to further revisions. And why should *my* system be the same as *your* system? Somehow we must "aggregate" these systems, and consider their evolution into the future, under the guidelines of rational inquiry.

This motivates long-run pragmatism, advanced by C. S. Peirce. The notion of justification is *rational acceptability in the limiting case of inquiry*. The definition says that a proposition is *true* just when it is acceptable in the ideal limit. But do we have any reason to suppose that there *is* such a limit; that there will be convergence, amongst all who investigate? Perhaps our theories will forever be overthrown; perhaps they will forever be partial and incomplete.

Even if inquiry does gradually converge, to "Scientific Consensus," might it still not be the case that we are, in reality, sadly deceived brains-in-vats? We cannot simply *define* reality to be what "Scientific Consensus" says it is at the end of inquiry. If we do this, we have ruled out, by fiat, the possibility of radical error.

## Intermezzo: The T-scheme and Material Adequacy

General definitions of truth are controversial. So, consider instead just the single proposition *that all men are mortal*. What does it mean to say of *this* proposition that it is true? Aristotle gives us a hint: a proposition is true exactly if things are as it says they are. So, this proposition is true if and only if all men are mortal. We are led to the following:

1 The proposition *that all men are mortal* is true iff all men are mortal.

For sentential truth bearers, the well-known example, from Tarski, is:

2 The sentence "snow is white" is true iff snow is white.

Sentences (1) and (2) are called *T-sentences*. They are "instances" of the general schematic principles:

The proposition that *p* is true iff *p*;
The sentence "*p*" is true iff *p*.

These are versions of what is known as the *T-scheme*.[11] To construct a T-sentence, we replace "*p*" by any declarative English sentence. There is, of course, no requirement that this sentence be true! That would be circular. So, the following is fine:

3 The proposition *that pigs can fly* is true iff pigs can fly.

Also, there is no (obvious) domain restriction on the sentences we may substitute in order to obtain T-sentences. They may involve any subject matter whatsoever. Thus:

4 The sentence "2 + 2 = 4" is true iff 2 + 2 = 4.
5 The sentence "torture is always wrong" is true iff torture is always wrong.

The T-sentences, (1)–(5), seem trivial or platitudinous. A complaint is that the T-sentences aren't *general* definitions, of the form (D). They don't tell us in general what it is to be true. They just tell us *one-by-one*, what it is for "snow is white" to be true, for "pigs can fly" to be true, and so on.

So, what exactly is the status of T-sentences? Consider "pigs can fly." We surely need empirical evidence to decide whether to accept, or reject, this biological hypothesis. But we do not need empirical evidence to know that this hypothesis is true if, and only if, pigs can fly. Hence, accepting a T-sentence is independent of particular empirical evidence. All we

need is to grasp are the relevant propositions involved, and the concept of truth. To use the jargon, T-sentences are *analytic*: we accept them in virtue of understanding the concepts they use.[12] But, a T-sentence is not a *general* definition, of the form (D). Rather, T-sentences are *partial definitions* of truth, each one specific to a particular truth bearer.

Suppose we wish to construct a *general* definition of truth, of the form (D). How should a proposed *general* definition be related to the *partial* definitions? Consider the following absurd definition of truth:

(D*) A sentence *A* is true iff *A* contains 27 letters.

Why is this absurd? The reason is that (D*) *does not imply the corresponding T-sentences*. In other words, one *cannot* show, from (D*), the following,

6 The sentence "snow is white" has 27 letters iff snow is white.
7 The sentence "2 + 2 = 4" has 27 letters iff 2 + 2 = 4.

And so on.

So, a proposed definition of truth counts as "correct" or "adequate" when it *implies* the corresponding T-sentences. Such a definition of truth is called *materially adequate*.[13]

## The Semantic Conception of Truth

Is it possible to construct truth definitions which *are* materially adequate? This was Tarski's aim in his 1935 article, "The Concept of Truth in Formalized Languages," presenting the *semantic* conception of truth, which he regarded as a version of the correspondence theory (although whether it is one remains controversial).

On the semantic conception, truth bearers are sentences, understood as strings of letters. For example, "fish swim" is the string "f," "i," "s," "h," " ," "s," "w," "i," "m." The truth or falsity of a string of letters only makes sense *relative to some language*. For example, "fish swim" is true in *English*, but could be false in another language. So, the semantic conception deals not with an absolute concept "*A* is true," but rather with a *relative* concept, such as "*A* is true in *English*," "*A* is true in *Spanish*," etc. In general, "*A* is true in *L*," where *L* is called the *object language*.

The object language might be a formalized language or it might be part of a natural language, such as *Spanish* or *Hindi*. The language *in* which we talk *about* the object language is allowed to be distinct from the object language, and is called a *metalanguage*. In the discussion below the metalanguage is *English*. For example, one may use *English* to talk about truth and falsity in *Spanish*.

Citing the classical correspondence definition, Tarski's version of the T-scheme is:

(T) The sentence *x is true in L* iff *p*.

A T-sentence is constructed by replacing "*x*" with a name of a sentence, and replacing "*p*" by the translation of the sentence. For example, if the object language is *German*, a possible T-sentence is:

8 The sentence "schnee ist weiss" is true in *German* iff snow is white.

This T-sentence might not be trivial or analytic for *you*. It will, however, be trivial or analytic for a bilingual *English* speaker who also speaks *German*. If the metalanguage contains the object language, we see the effect of what is called "disquotation":

9 The sentence "snow is white" is true in *English* iff snow is white.

The general procedure for constructing a Tarskian truth definition is as follows. First, one specifies an object language *L*, on the assumption that one may translate from *L* into the metalanguage; one next constructs, in the metalanguage, a definition of "*A* is true in *L*"; and finally, one proves that this definition is materially adequate.

For example, suppose *L* is a language with just two sentences, *X* and *Y*, whose translations are "dogs bark" and "fish swim." There are only two T-sentences, namely,

10 *X* is true in *L* iff dogs bark.
11 *Y* is true in *L* iff fish swim.

A materially adequate truth definition for *L* may be given as follows:

12 *A* is true in *L* iff [(*A* is *X* and dogs bark) or (*A* is *Y* and fish swim)].

(The reader might try to show how to infer (10) from (12).)

For object languages of any serious interest, however, one cannot write down such definitions, as there are infinitely many sentences to deal with. For example, suppose that *L* contains the logical connective "*not*" and one basic sentence, say *X*, whose translation is "dogs bark." Then *L* has infinitely many sentences: *X*, *not-X*, *not-not-X*, *not-not-not-X*, etc. Because of this, one cannot write down a truth definition like (12). Instead, one gives what is called a *recursive definition*, as follows:

13 *X* is true in *L* iff dogs bark.
14 *not-A* is true in *L* iff *A* is not true in *L*.

This recursive definition is adequate. The method may be generalized to include other logical connectives, such as "*and*," "*or*" and so on.

When the object language contains names, predicates, connectives, and quantifiers (the phrases "for any" and "there exists"), the situation becomes more complicated. One must first define two auxiliary semantic concepts: *reference* (or *denotation*) and *satisfaction*. Satisfaction, roughly, is the relation of a predicate, such as "loves," to the things it applies to. For example, a pair of objects [*a*, *b*] *satisfies* the predicate "loves" if, and only, if *a* loves *b*. Reference is the semantic relation that holds between a name and what it stands for. For example, the name "Bertrand Russell" refers (in *English*) to the philosopher Russell himself.

A final point. One might think that *English* contains its *own* truth predicate: i.e., a predicate meaning "is true in *English*." However, this assumption leads to a paradox, the notorious Liar Paradox. Informally, consider the so-called liar sentence "this sentence is not true,"

which attributes *untruth* to *itself*. Call the liar sentence G. Informal reasoning leads to a contradiction. For, G is equivalent to "G is *not* true." But the T-scheme tells us that G is equivalent to "G is true." So, we conclude "G is true" is equivalent to "G is *not* true." A contradiction! In short, the T-scheme is *inconsistent*. Tarski drew several conclusions from this, as well as using it to prove some powerful mathematical results.[14] In particular, the conclusion that the common-sense concept of truth is inconsistent.

We next turn to some objections to the semantic conception of truth.

A preliminary objection is that the semantic conception deals with truth for sentences, not propositions. But perhaps propositions are basic, and we should define *sentence*-truth in terms of *proposition*-truth. For example, as follows: a sentence is true relative to some language iff the *proposition* that it expresses, relative to that language, is true. This is attractive, but not without problems. The main problem is that it is rather unclear what propositions really are; some philosophers simply reject them, in favor of sentences, whose syntactical structure is much clearer.

A second objection is that Tarski shows how to define "*A* is true in *L*," a language-relative notion of truth, but not an absolute notion, "*A* is true." Thus, the single, univocal, notion of truth has fragmented into seemingly unrelated concepts: "true-in-*English*," "true-in-*Spanish*," etc. In reply, note that it just does not make sense to talk of sentences as being merely true or false. For sentences, their truth must be relative to a language.[15]

A third, and rather threatening, objection concerns a lacuna. The semantic conception seems not to *explain* the semantic notions involved: reference, satisfaction and truth. For example, *German* contains the noun "schnee," whose translation in *English* is "snow." A Tarskian truth theory for *German* thus contains the partial definition,

15 The word "schnee" refers in *German* to snow.

This is a semantic fact about *German*. But it doesn't give any indication as to *why* the noun "schnee" refers, in *German*, to snow. The point generalizes to other semantic concepts. One might argue that a Tarskian theory should be extended, adding a separate *theory of reference*, which explains *why* expressions refer to whatever they do, perhaps in terms of how expressions are used, and causal connections between speakers, the expressions they use, and the referents of the expressions.

A forth objection concerns whether Tarskian semantic methods can be *generalized* to real-life natural languages, which exhibit a variety of poorly understood features, including more complicated ways of constructing sentences, context-sensitive expressions, evaluative predicates, and phenomena such as ambiguity and vagueness. One cannot summarize the overall situation easily, but there is a large amount of work in semantic theory, generalizing Tarski's approach to many of these phenomena.[16]

Finally, is the semantic conception a correspondence theory? This is disputed. Tarski himself stated that it was, and others followed him. The semantic conception of truth is based on the classical correspondence definition ($D_1$), rather than the correspondence-to-fact definition ($D_2$). So, *if* the classical correspondence definition is a correspondence theory, then surely so is Tarski's. For ($D_1$) meets the *correspondence intuition*: truth depends on how reality is. The point of difference is that a Tarskian definition of truth doesn't introduce *facts*, and doesn't introduce a *sentence-to-fact correspondence relation*.

## Deflationism

The proposition *that snow is white* is true iff snow is white. Thus, to assert the *truth* of this proposition is equivalent to asserting the proposition itself. Similarly, to claim that "snow is white" is true is equivalent to claiming that snow is white; and so on. To assert "*A* is true" is equivalent to asserting *A*. These equivalences are encapsulated by one or other version of the T-scheme. The Tarskian semantic conception took (a consistent version of the) T-scheme as an *adequacy condition* on definitions of truth. But given that the T-sentences are platitudes, an attractive suggestion is that the concept of truth is *fully* captured by the T-scheme alone. If correct, perhaps nothing more, or little more, needs to be said. The problem of truth has been *deflated*: we arrive at *deflationism*.[17] If this is right, the view that truth has any sort of "nature," requiring metaphysical analysis, is a philosophical error – a muddle.

Deflationary suggestions were made by Frege, Ramsey, Ayer, and Wittgenstein. An early version noted that since "*A* is true" is equivalent to *A*, the predicate "true" might seem *redundant*. However, this is too quick, as there are other contexts where it is not so obvious how to eliminate the predicate "true," a point emphasized by both Tarski and Ramsey.

Still, the T-sentences are analytic platitudes about truth, and deflationism tries to exploit this to the maximum. While there is no exact consensus on what deflationism is, beyond some rather unclear claims that truth is not a property, or that the problem of truth is a "muddle," there are several deflationary theses commonly defended. First, that the notion of truth is, really, a *logical* notion; second, that the sole *reason* for having a truth predicate in a language consists in its *logical utility*, third, that the theory of truth is *neutral* on non-truth-theoretic matters; and forth, that the concept of truth plays no *essential role in explanations*.[18]

To explain the claim that truth is a logical notion, consider the logical expression "*and*." To understand "*and*" is to know how to reason with it. One may infer "*A and B*" from the two assumptions *A* and *B*. One may infer *A* from "*A and B*," one may also infer *B*. To understand "*and*" is just to understand these logical rules. With truth there is an analogy. From *A*, one may infer "*A* is true"; and from "*A* is true," one may infer *A*. Thus, there are logical rules for reasoning with the truth predicate, and these seem analogous to the logical rules for reasoning with other logical notions, "*and*," "*not*," "*or*" and so on.

This brings us to the second claim, concerning the logical usefulness of a truth predicate. Suppose that one is so impressed with Harvey's knowledge that one wishes to endorse *everything* Harvey says. If one had a lot of spare time, one could begin asserting the following "infinitely long" statement:

16 If Harvey says that penguins waddle, then penguins waddle; *and* if Harvey says that fish swim, then fish swim; *and* if Harvey says that plastic is edible, then plastic is edible; . . . and so on.

In a sense, (16) is an "infinite conjunction" of statements of the form "If Harvey says that *p*, then *p*." However, note that with the predicate "true," one may simply say:

17 Everything Harvey says is true.

Thus, using the truth predicate and the rules for reasoning with it, one may re-express a certain *infinite* conjunction (16) as a single *finite* statement (17). Similarly, even if we do not know *what* John said, we may still *repudiate* it by saying, "what John said is not true." Even if we do not know specifically what the Dalai Lama said, we may indirectly *endorse* his statements by saying "whatever the Dalai Lama says is true." The T-scheme accounts for this logical utility of the truth predicate. Furthermore, in order for the truth predicate to have this logical utility, nothing more is required beyond the T-scheme: there is no need for talk of facts, correspondence, or notions of justification.

The third claim concerns the neutrality of the T-scheme. Again, there are technical results which confirm this. Accepting the T-scheme makes no difference to one's background views on non-truth-theoretic matters. One may accept the T-scheme (more exactly, a consistent version) irrespective of whether you think "electrons are *mind-independent entities*" or you think "electrons are *logical constructions from sense-data*." The T-scheme is thus metaphysically neutral.

The final deflationary claim is that, while a truth predicate has a certain logical utility, the predicate plays no essential explanatory role whatsoever. Truth is thus "insubstantial," in some sense. A popular argument for accepting an empirically successful scientific theory, is that the best explanation of its empirical success (i.e., its making true predictions), consists in the theory itself being *true*. Does not the truth of the theory explain the truth of the predictions? The deflationist may reply, however, that uses of the notion of truth may be *eliminated* from particular explanations, by using the T-scheme. For example, we observe energy release when uranium-235 is subjected to irradiation by neutrons. The best explanation, one might say, is that Einstein's theoretical law "$E = mc^2$" is *true*. However, the phenomenon is just as easily explained by the ground-level claim that $E = mc^2$. If this is right, there is no need to bring in truth: truth is *dispensable* in scientific explanations.

Let us turn now to some objections to deflationism.

A major objection is that the T-scheme, unless restricted, is *inconsistent*. It leads to the Liar Paradox. It is unclear what deflationism has to say. If consistency is wanted, some T-sentences must be rejected. The problem of explaining *which* ones is non-trivial. The deflationist might, on the other hand, settle for an inconsistent theory of truth. But the cost is high, as it requires unappealing revisions in logic. (The semantic conception, whatever its faults or lacunae, is *not* inconsistent. It was part of Tarski's intention to develop a consistent theory of truth.)

The T-scheme may, in fact, be weakened (in various ways), to restore consistency. The objection now is that it becomes too *weak* to give a *usable* theory of truth. We would like to be able to say, in general, that:

18 For any sentence *A*: *not-A* is true iff *A* is not true.
19 For any sentences *A* and *B*: *A-and-B* is true iff both *A* and *B* are true.

Principles like these are used in or reasoning all the time. However, a deflationary theory of truth based on the T-scheme (or a consistent version) does *not* imply the generalizations (18) and (19). Note that these sorts of generalizations are always built-in to a Tarskian semantic theory of truth.

A third objection to deflationism concerns the *normative* dimension of truth: *true* beliefs are what we *aim* to believe, or what we *ought* to believe. Or cognitive inquiries are guided by a normative rule of the form,

20 Aim to believe a proposition if and only if it is *true*.

It might seem that this normative feature is not accounted for by deflationism. However, perhaps the deflationism can reply to this objection as follows. The *formulation* of this rule as a single statement is really just an example of the logical utility of truth predicate, which has already been explained. The single rule, (20), is equivalent, by the T-scheme, to a *schematic rule*, of the form

21 Aim to believe that *p* if and only if *p*.

So, the truth predicate allows us to reformulate the *schematic* normative rule as a single normative rule. And the schematic normative rule (21) doesn't appear to involve truth at all, at least not explicitly.

## Concluding Remarks

This chapter has covered quite a lot of material, and I must apologize that certain logical technicalities have entered as we moved beyond the more basic material on the correspondence and the epistemic theories. However, this is virtually unavoidable, as all important work in philosophy concerning truth since the last 1960s is of a nature similar to the material in Sections 4, 5 and 6 above. However, it is hoped that the interested reader may take the broad survey above as a useful starting point for further study.[19]

Finally we return to the question raised at the start, concerning *why truth matters*. The correspondence theorist may answer this as follows. Truth matters because truth involves *agreement with reality*, and it is *reality* that matters to us. In general, it matters to us whether food norishes us, or loved ones are unharmed, or surroundings are safe, or prospects are good, and so on. Various political and social phenomena also matter to us. To a physicist, the nature of the physical world matters. To a historian, the past events matter. In short, truth matters because reality matters.

### Further Reading

Three excellent general resorces are the textbook Kirkham (1992) and the anthologies, Blackburn and Simmons (1999) and Lynch (2001). The anthologies are comprehensive and the editorial contributions are very clear. Two shorter expository pieces are Haack (1978: ch. 7) and Glanzberg (2006), at the online *Stanford Encyclopedia of Philosophy*, which also has several valuable truth-related articles. Three further expository books are Engel (2002), Künne (2003) (which contains much interesting historical material), and Blackburn (2006). Each of these has a rather deflationary approach.

Regarding specific accounts of truth, David (1994) defends a correspondence-to-fact view against deflationism, while Horwich (1990; 1999) defends a version of deflationism. The semantic conception is set out in Tarski (1935; 1944). The coherence theory is defended in Walker (1989) while Putnam (1981) defends a form of pragmatism. The identity theory is defended in Dodd (2000).

Beyond these pointers, the literature on truth is simply huge.

## Notes

1 Throughout, we concentrate on truth. A sentence is *false* if and only if its negation is true.
2 Sometimes non-linguistic non-propositional entities are called "true" – e.g., a "true friend," a "true Scotsman," etc. We set such uses aside, and concentrate on truth as applied to statements, beliefs, etc.
3 A proposition is the *content* of a statement, or the *content* of a mental state. This allows us to say that sentences with the same meaning express the same proposition, and that what you believe to be the case is exactly the same as what I believe.
4 Henceforth, we use "iff" as an abbreviation for "if, and only if."
5 Some authors, even scientific authors, use the word "fact" to mean roughly "statement accepted on the basis of observation." This is *not* what we mean, for such statements may be *false*, and therefore do *not* correspond to any fact.
6 The term "truth maker" has been suggested for what makes a truth bearer true. We note here that there is a theory of facts which *identifies* facts with *true propositions*. Thus, the correspondence relation between true proposition and fact is the *identity relation*. This view is called *the identity theory*. See Dodd (2000) for a defense of the identity theory.
7 One might think that there are no states of affairs that *don't* obtain. If so, facts are states of affairs, period.
8 There is an objection to the correspondence view (the "Slingshot Argument") which concludes that *every* truth corresponds to the *same* fact: the *Big Fact*. The reader may consult Neale (2001) for details.
9 It is unclear how to *verify* statements about certain topics, such as morality or religion. The verificationist might regard such statements as neither true nor false, or meaningless.
10 *Consistency* means that for no statement *A*, can one prove both *A* and *not-A*; while *completeness* means, for any statement *A*, either one can prove *A*, or one can prove *not-A*.
11 A *scheme* is a kind of linguistic frame into which various sentences may be substituted. Some authors write "schema."
12 There, however, is a fly in this ointment, which is that some T-sentences are false, because of the "Liar paradox," briefly mentioned below. However, one can restrict the set of T-sentences, and together they may be understood to *implicitly* define truth.
13 As an application, the epistemic definitions ($D_4$), ($D_5$), and ($D_6$) are *not* adequate in this sense.
14 In more detail, the result is known as Tarski's *Indefinability Theorem*: if a consistent language *L* is "sufficiently rich," the concept of truth in *L* is not itself definable *in L*. If the metalanguage for *L* contains a definition of truth in *L*, the metalanguage is, in some sense, "richer" than the object language.
15 To be more exact, the truth value of a sentence – a string of symbols – is relative to an interpretation of those symbols.
16 Consider a natural language *L* with context-sensitive expressions, such as "I," "now," and "here." The semantic theory is modified as follows. The notion of *truth* (in *L*) is replaced by the notion of *truth* (in *L*), *relative to certain parameters*. These parameters specify the speaker, the time, and the location of a speech act.
17 Two popular forms of deflationism are *disquotationalism* (truth bearers are sentences) and a *minimalism* (truth bearers are propositions). For our purposes, they needn't be sharply distinguished.

18 See Horwich (1999: 262) for a clear summary of what he takes his version of deflationism to involve.

19 We have, alas, not discussed debates about meaning, relativism/rationalism, statements possibly lacking truth values (e.g., moral statements; vague statements), and possible revisions of classical logic. We have not discussed the technical work (some devoted to studying the semantic paradoxes) of increasing relevance, particularly to debates about deflationism, which have dominated the recent philosophical literature

## References

Blackburn, S. (2006). *Truth: A Guide for the Perplexed*, London: Penguin.

Blackburn, S. and Simmons, K. (eds.) (1999). *Truth*, Oxford: Oxford University Press.

David, M. (1994). *Correspondence and Disquotation: An Essay on the Nature of Truth*, Oxford: Oxford University Press.

Dodd, J. (2000). *An Identity Theory of Truth*, London: Macmillan.

Engel, P. (2002). "Truth," in J. Shand (ed.), *Central Problems of Philosophy*, Chesham: Acumen Publishing.

Glanzberg, M. (2006). "Truth," *The Stanford Encyclopedia of Philosophy* (online).

Haack, S. (1978). *Philosophy of Logics*, Cambridge: Cambridge University Press.

Horwich, P. (1990; 1998). *Truth*, Oxford: Oxford University Press.

Horwich, P. (1999). "The Minimalist Conception of Truth," in Blackburn and Simmons.

Kirkham, R. L. (1992; 1995). *Theories of Truth*, Cambridge, MA: MIT Press.

Künne, W. (2003). *Conceptions of Truth*, Oxford: Oxford University Press.

Lynch, M. (ed.) (2001). *The Nature of Truth*, Cambridge, MA: MIT Press.

Neale, S. (2001). Facing Facts, Oxford: Clarendon Press.

Putnam, H. (1981). *Reason, Truth and History*, Cambridge: Cambridge University Press.

Russell, B. A. W. (1967) [1912]. *The Problems of Philosophy*, Oxford: Oxford University.

Tarski, A. (1935). "Der Wahrheitsbegriff in den formalisierten Sprachen," *Studia Philosophica*, I, 261–405; tr. "The Concept of Truth in Formalized Languages," in Tarski 1956.

Tarski, A. (1944). "The Semantic Conception of Truth and the Foundations of Semantics," *Philosophy and Phenomenological Research* 4: 341–376; repr. in Blackburn and Simmons and in Lynch.

Tarski, A. (1956). *Logic, Semantics, Metamathematics: Papers by Alfred Tarski from 1922–1938*, ed. J. H. Woodger, Oxford: Clarendon Press.

Walker, R. (1989). *The Coherence Theory of Truth*, London: Routledge; extracts repr. in Lynch.