11

Truth, the Liar, and Tarskian
Truth Definition

GREG RAY

Alfred Tarski's work on truth has become a touchstone for a great deal of philosophical work on truth. A good grasp of it is critical for understanding the contemporary literature on truth and semantics. In this chapter, I will present a fresh interpretation of Tarski's view, one which aims to draw it out more fully in areas of philosophical interest. This has required extrapolation (e.g. drawing explicit implications for concepts and properties) and reverse engineering (e.g. introducing the notion of full conceptual warrant) for which I will not offer textual justification here. My purpose is to introduce Tarski's central ideas briefly and in the most tenacious way I can. It is my hope that this brief study will prove useful as a basis for further investigation.

1   Truth

Our topic of philosophical concern is truth, but we will be almost entirely concerned with *the concept of sentential truth*, that is the relational concept of (something's being) *a true sentence of* (some language). We will also have to deal with certain language-specific truth concepts, such as the concept of *a true sentence of English*. In addition to sentential truth, one can speak of *doxic truth* and *propositional truth* – these being the concepts of *a true belief* and *a true proposition*, respectively. Some think the concept of propositional truth more fundamental than others. We need not join this debate here, if we are careful to draw our conclusions with appropriate care.

The T-strategy

Underlying Tarski's work is a basic observation about sentential truth, namely that claims such as

'snow is white' is a true sentence of English if and only if snow is white
'neige est blanche' is a true sentence of French if and only if snow is white

seem quite obvious and unexceptionable, and they also state necessary and sufficient conditions for the truth of the sentences mentioned in them without appeal to any

further semantic notions. Tarski's starting point is the idea that such statements might be helpful in characterizing sentential truth. If *one* of these sentences gives necessary and sufficient conditions for the application of 'is a true sentence of French' to one sentence, then together a full set of such sentences could give necessary and sufficient conditions for the application of 'is a true sentence of French' to any sentence of French. Thus, if some finite way of expressing this infinite set of conditions could be found, we would have the makings of a definition of 'is a true sentence of French,' and this would be a start. The sense in which such a definition might characterize or 'capture' a truth concept will be an important question to take up later. Let us call the strategy of definition we've outlined, *the T-strategy*. The sentences we are concerned with are known as *T-sentences*.

DEFINITION A *T-sentence in English for* a language L, is any sentence which may be obtained from the *T-schema*,

s is a true sentence of L if and only if p,

by substituting for 's' a syntactic description in English of a sentence of L, and substituting for 'p' a translation into English of that same sentence of L.

I have characterized T-sentences expressed *in* English. There are obvious correlates for other meaningful languages. The T-strategy suggests that, if I wanted to characterize sentential truth for the Xanadic language (my *object language*) and I wanted to express myself in Polish (my *meta-language*), I should find some finite way of expressing the set of conditions given by the class of T-sentences in Polish for Xanadic. More generally, to make a definition realize the T-strategy for a language L in a language M, we will insist the definition satisfy the following condition of adequacy.

CONVENTION T All the T-sentences in M for L are theorems of that theory which consists of our definition statement (plus, perhaps, some axioms about syntax and sequences).

The problem of generality

There is no obvious generalization of the T-sentences – one which would be equivalent to stating that infinitude of sentences. So, the technical challenge of the T-strategy is finding a *finite* way of expressing what the relevant T-sentences express. Call this the *problem of generality*.

Tarski was not the first to see that something like T-sentences might be used to characterize a truth concept nor the first to tackle the problem of generality. Tarski seemed to think of certain familiar pronouncements about truth as unsatisfactory attempts at solving the problem. One such was Aristotle's famous dictum: *to say that what is, is not or that what is not, is false, while to say that what is, is or that what is not, is not, is true.* F. P. Ramsey pursued something like a T-strategy for doxic truth and expressed clearly the challenge of achieving generality.

> Suppose a man believes that the earth is round; then his belief is true because the earth *is* round; or generalizing this, if he believes that A is B his belief will be true if A is B and false

otherwise. It is, I think, clear that in this last sentence we have the meaning of truth explained, and that the only difficulty is to formulate this explanation strictly as a definition. If we try to do this, the obstacle we encounter is that we cannot describe all beliefs as beliefs that A is B since the propositional reference of a belief may have any number of different more complicated forms. A man may be believing that all A are not B, or that if all A are B, either all C are D or some E are F, or something still more complicated. We cannot, in fact, assign any limit to the number of forms which may occur, and must therefore be comprehended in a definition of truth; so that if we try to make a definition to cover them all it will have to go on forever, since we must say that a belief is true, if supposing it to be a belief that A is B, A is B, or if supposing it to be a belief that A is not B, A is not B, or if supposing it to be a belief that either A is B or C is D, either A is B or C is D, and so on *ad infinitum.*   (Ramsey 1929: 9)

Tarski thought that he had an elegant solution to the generality problem, but recognized two significant obstacles – natural language and the Liar Paradox. We will discuss these in turn.

Conceptual status of T-sentences

Before we proceed, however, let us consider more closely the conceptual status of T-sentences. What one might hope to get out of a definition of 'is a true sentence of' (or 'is a true sentence of French') using the T-strategy depends on the status of these T-sentences. So long as the mentioned sentence is indexical-free, tenseless, and not vague or ambiguous, each T-sentence gives a *materially* necessary and sufficient condition for the application of the truth predicate in question to the mentioned sentence. What more can be said? We should not rush to claim that T-sentences express conceptual truths or are analytically true. There is, nonetheless, some interesting conceptual linkage between the concept of sentential truth and the T-sentences – a linkage which it is our current task to elucidate. Consider again a T-sentence such as

'neige est blanche' is a true sentence of French if and only if snow is white.

Such a sentence recommends itself to us, because it seems, roughly, that one who has the proper linguistic understanding knows it to be true. There is something to this idea. Let us begin with a case simpler than T-sentences.

DEFINITION   For language M and sentence, s, of M, we shall say that s has *simple conceptual warrant in* M iff one who understands s (as a sentence of M), is in a position to know (on non-truth-functional grounds) that if (1) each predicate of s is subserved in M by the concept it expresses in M, and (2) each singular referring term of s refers in M, then s is a true sentence of M.

Understanding s as a sentence of M, as used here, is meant to imply of the agent that he or she,

1. for each predicate, p, of s, grasps the concept, c, expressed in M by p and knows of c that it is expressed in M by p, and

2. associates with each singular referring term of s a condition and knows that it uniquely picks out the referent in M of the term if such there be, and nothing otherwise.

To explain what is intended by saying that a predicate is subserved by a concept, I will avail myself of the useful fiction that concepts come supplied with explicit *application rules* which say what sorts of things are supposed to be included or excluded by the concept.

DEFINITION   A concept, c, *subserves* a predicate, p, of language M iff for all x in the domain of discourse of M, (1) if the application rules for c imply that x falls under c, then p applies in M to x, and (2) if the application rules for c imply that x fails to fall under c, then p fails to apply in M to x.

Ordinarily, of course, if a predicate expresses a concept, that concept subserves the predicate. However, a concept could subserve a predicate which did not express it, and, just possibly, a predicate could express a concept that did not subserve it.

Sentences with simple conceptual warrant evidently include (1) the analytically true; (2) sentences free logics treat specially, such as 'if Vulcan is green, then Vulcan is green'; as well as (3) some more interesting cases involving vacuous names, such as, 'if Vulcan is a planet, then Vulcan is a heavenly body.' These sentences have exceptional conceptual credentials, though not all are guaranteed to be true. We note in passing that they are all of a sort that we would be entitled to rely on for the purposes of scientific theorizing – at least until such time that it became known that 'Vulcan' fails to refer.

T-sentences do not have simple conceptual warrant, but an extension of the same idea applies to them.

DEFINITION   For sentences M and L, and for t, a T-sentence in M for some language, L, (where t has form ⌜δ is a true sentence of λ iff Γ⌝), we shall say that t has *subtle conceptual warrant in* M just in case one who

(1)   understands t as a sentence of M,
(2)   *recognizes that the sentence denoted in M by δ is a sentence of the language denoted in M by λ,*
(3)   *understands the sentence denoted in M by δ as a sentence of the language denoted in M by λ.*

is in a position to know (on non-truth-functional grounds) that

if each predicate of t is subserved in M by the concept it expresses in M, and each singular referring term of t refers, then t is a true sentence of M.

Our T-sentences do have subtle conceptual warrant. The notion of subtle conceptual warrant aims to capture the special sense in which these sentences are conceptually underwritten. It grounds our feeling that example T-sentences are 'iron clad.' Let us say of a sentence of a language M that it has *full conceptual warrant* in M just in case it has either simple or subtle conceptual warrant in M.

Exactly specified languages

Tarski did not think that there was a well-defined class of T-sentences for natural languages like English. He thought that it was not clearly determined what was the basic vocabulary of English. Surely this is correct, since it is vague whether a new term in use, for example 'za,' should be thought of now as a term of English. Tarski also thought indexicality and tense presented difficulties for working directly with a natural language. These days, these are not seen as serious obstacles, because of the work of Donald Davidson (1967: 34) and others.

In the face of these obstacles, Tarski chose to pursue the T-strategy with regard only to languages with what he called an *exactly specified structure*. This would ensure that a language under examination had a well-defined primitive vocabulary and grammar, and this would help ensure that there could be a well-defined set of T-sentences for that language. Simplifying Tarski somewhat, let us say that to *exactly specify* a language, one must specify: a basic vocabulary, grammatical formation rules, the class of sentences, a set of axioms, and inference rules. If a language is exactly specified in purely syntactic terms, then it is said to be *formalized*.

The last two items on the specification list may seem objectionable. We do not think of languages as coming equipped with inference rules and axioms. The worry subsides, however, once we see that the inference rules and axioms in question are indeed determined (albeit not uniquely) by a meaningful language. First, logical relations between meaningful sentences obtain in virtue of what those sentences mean, and a set of inference rules is a way of codifying logical relations. One might also think of the set of inference rules as a way of identifying and specifying the meanings of the logical terms of the language. Either way, what we represent by including inference rules in a specification is determined by the language itself, not super-added. Second, it is a constraint on the axioms Tarski has in mind that they axiomatize the *conceptually assertible sentences* of the language – and these are, I propose, just the sentences of the language with full conceptual warrant. Thus, the axioms of an exact specification are also clearly determined by the language.

Using exactly specified languages makes Tarski's technical project more sure-footed, but makes our philosophical job harder. Since we have good reason to believe that there is no well-defined class of meaningful sentences for a natural language like English, there can be no exact specification of such a language. For this reason, care and reflection is necessary in considering any results we may obtain.

One further notion which we will make use of in the sequel is that of an *empirically assertible sentence.* Conceptually assertible sentences are ones which have exceptional credentials in virtue of which, special knowledge to the contrary, they may be 'treated as true' for the purposes of scientific and logical work. In an empirical language (e.g. a language suitable for expressing physical theory as opposed to the language of arithmetic) some sentences may be treated as true for the purposes of scientific theorizing not in virtue of their conceptual standing, but in virtue of being empirically confirmed. Keeping things as simple as possible, we will say that an *empirically assertible sentence* of a language is one which has met a certain (unspecified) standard of confirmation, and an *assertible sentence* of a language is one which is either conceptually or empirically assertible.

2   The Liar

Suppose, then, that we restrict further attention to languages with an exactly specified structure. The next obstacle to pursuing the T-strategy is more grievous. Considerations based on the Liar Paradox suggest that the T-strategy will lead us into inconsistency.

The Liar Argument

For the sake of argument, let us assume that the definite description 'the sentence with feature f' uniquely denotes the sentence which is *quoted in* sentence (a) below. Our argument will be given in (a fragment of) English. Also, for simplicity, we will suppose that 'L' refers to a language which looks and is structured just like a fragment of English and has no false cognates (so translation into English is transparent). Consider the following Liar Argument which begins with a T-sentence.

(a)   'The sentence with feature f is not true in L' is true in L iff the sentence with feature f is not true in L.

(b)   'The sentence with feature f is not true in L' is identical to the sentence with feature f.

(c)   So, 'The sentence with feature f is not true in L' is true in L iff 'The sentence with feature f is not true in L' is not true in L.

It is worth stating carefully how this argument (sequence of sentences) poses a threat to reason. First, suppose you think that (a) and (b) represent beliefs that you hold. Then, certainly, (c) represents something that could be validly inferred from things you believe. But (c) is logically self-contradictory, and this suggests that your beliefs are in a sorry state indeed. You would be rationally compelled to conclude that you had a false belief. It is hard to see how (b) could be the culprit, so suspicion falls on (a). However, (a) could not represent a false belief you had, because we can prove (a) is not false:

> After all, a claim [like (a) which is of the form] $\ulcorner$A iff B$\urcorner$ can be false only if (i) A is true and B is false or (ii) A is false and B is true. Where A is $\ulcorner$'S' is true$\urcorner$ and B is S, these combinations cannot occur, for (i) if S is false, then the claim that it is true cannot be true and (ii) if S is true, then the claim that it is true cannot be false.   (Soames 1999: 51)

The Inconsistency Argument

Thus, the Liar Argument presents us with an intolerable situation – a genuine affront to reason. The Tarskian analysis of this situation is based on the following Inconsistency Argument. Let M be a fragment of English sufficient for giving the Liar Argument.

(1)   Sentence (a) is a conceptually assertible sentence of M. (Premise)

(2)   Sentence (b) is an empirically assertible sentence of M. (Premise)

(3)   The ordinary rules of logic apply in M (i.e. the rules of inference of M underwrite the usual deductive moves). (Premise)

(4)   Thus, the deductively inconsistent sentence, (c), is derivable from (a) and (b) by the rules of inference of M.

(5)   It follows that *the language M is inconsistent* in the sense that a deductively inconsistent sentence is derivable by the rules of inference of M from the assertible sentences of M.

This argument is not a problematical argument and its premises are ones that we have no reason at all to reject. Moreover, there *are* exactly specifiable languages for which these premises evidently hold, such as that fragment of English used in giving the Liar Argument earlier. For this reason Tarski held that an exactly specified language as much like English as possible would be inconsistent – a claim that has been a source of consternation and a subject of misinterpretation, for example Soames (1999).

Incoherence of the concept

Examination of the Inconsistency Argument reveals that one of the sentences, (a) or (b), must be *assertible but not true in M*. Again, suspicion falls only on (a). A simple argument showed that (a) cannot be false, so it is immediate that (a) must lack a truth value (and so on some understandings of belief you would surely have been mistaken to think it represented any belief you held).

Now, (a) is a T-sentence, and it is assertible because it has subtle conceptual warrant. Since it is certainly possible for an agent to satisfy the antecedent epistemological conditions for subtle conceptual warrant with respect to (a), we know by this that someone could be in a position to know of (a) that

if each predicate of (a) is subserved in M by the concept it expresses in M, and each singular referring term of (a) refers in M, then (a) is a true sentence of M.

We know that (a) is *not* a true sentence of M and it is evident that there is no reason to think that any singular term of (a) fails of reference. From these we infer that some predicate of (a) is not subserved in M by the concept it expresses in M. The only candidate is the predicative expression, 'is a true sentence of.' Thus, we are led to conclude that *'is a true sentence of' is not subserved in M by the concept, c, that it expresses in M, i.e. the concept of sentential truth*. How could it possibly happen that we have made a predicate express some concept, and yet, in spite of our intentions, that concept does not subserve it? The only conceivable way this could happen is if it were *strictly impossible* for the concept to subserve it. Such an impossibility is guaranteed if

the application rules for c imply that the pair ⟨'The sentence with feature f is not true in L', L⟩ *falls under* c, *and* the application rules for c imply that ⟨'The sentence with feature f is not true in L', L⟩ *fails to fall under* c.

The concept of sentential truth is, in a word, incoherent.

3   Tarskian Truth Definition

Evidently the T-strategy invites inconsistency since it is a T-sentence that sets up the Liar Argument. Nonetheless, Tarski has the idea that the strategy still might be usefully carried out by further restricting attention to exactly specified languages for which not all the assumptions of the Inconsistency Argument hold. Specifically, Tarski proposed that we can do this if we only consider object languages, L, which are *not* semantically closed. Where a *semantically closed* language, L, is characterized loosely as one which (1) has the resources to denote its own expressions, and (2) has the resources to predicate truth in L of those expressions. The crux of the matter is that a semantically closed language is one in which a liar sentence (like 'the sentence with feature f is not true in L') can be formed, and this is a sort of thing we are now aiming to avoid. A complete set of T-sentences *for* such a language must include a T-sentence for that liar sentence, and thus, any language, M, in which we could pursue the T-strategy would be one in which the T-sentence for the liar sentence was an assertible one, that is premise (1) of the Inconsistency Argument would hold. So long as we stick with languages that are not semantically closed, however, we effectively avoid this.

I am simplifying. The existence of sentences that form *Liar chains*, means that there are variants on the Liar and Inconsistency Arguments which will make the task of identifying the languages suited for Tarski's definitional project trickier yet (cf. Kripke 1975: 54–5; Yablo 1993).

Truth definitions

We have now (let us suppose) identified a class of exactly specifiable languages for which we might still hope to carry out the T-strategy. As stated earlier, Tarski's insight was that the problem of generality could be solved by employing the (now very familiar) technique of recursive definition. Tarski proceeds by example, showing how to give a recursive definition meeting Convention T for the language of the calculus of classes (a quantified language ranging over sets and having a single predicate term expressing the subset relation). Note, the language in which the definition is expressed is, perforce, *expressively richer* than the object language, since the former has sentences that translate all those of the object language, *as well as* the resources to denote the expressions of the object language.

To give an example definition here, we will use a language, L, which has a two-place predicate, '$\subseteq$', for the subset relation, plus logical terms for negation, conjunction, and quantification, and some individual variables.

DEFINITION   Let an *L-sequence*, f, be a function from the variables of L into the domain of discourse of L.

DEFINITION   For a variable, α, of L, let an *α-variant* of an L-sequence f be any L-sequence, f′, which is just like f except possibly for the value f′ assigns to α.

DEFINITION   For all L-sequences, f, and every formula, σ, of L, f *L-satisfies* σ iff

if σ is of the form $\ulcorner \alpha \subseteq \beta \urcorner$ for some variables α and β, then f(α) is a subset of f(β),

171

if σ is of the form ⌜ψ & θ⌝ for some formulas ψ and θ, then f L-satisfies ψ and f L-satisfies θ,

if σ is of the form ⌜ ̚ψ⌝⌜ for some formula ψ, then f does not L-satisfy ψ, and

if σ is of the form ⌜∀α ψ⌝⌜ for some variable α and formula ψ, then every α-variant of f L-satisfies ψ.

DEFINITION   For all sentences φ of L, *φ is a true sentence of L* iff φ is L-satisfied by every L-sequence.

This is not the definition we are looking for, but it is a simple matter to transform our recursive definition of L-satisfaction into an explicit definition and combine it with our truth definition to yield the following explicit truth definition.

DEFINITION   (explicit): For all sentences φ of L, *φ is a true sentence of L* iff for all L-sequences, g, ⟨g,φ⟩ is a member of the least set, X, such that for all L-sequences, f, and L-formulas, σ,

if σ is of the form ⌜α ⊆ β⌝ for some variables α and β, then ⟨f,σ⟩ ∈ X iff f(α) is a subset of f(β),

if σ is of the form ⌜ψ & θ⌝ for some formulas ψ and θ, then ⟨f,σ⟩ ∈ X iff ⟨f,ψ⟩ ∈ X and ⟨f,θ⟩ ∈ X,

if σ is of the form ⌜ ̚⌝ψ⌜ for some formula ψ, then ⟨f,σ⟩ ∈ X iff ⟨f,ψ⟩ ∉ X, and

if σ is of the form ⌜∀α ψ⌝, then ⟨f,σ⟩ ∈ X iff for every α-variant, f′, of f, ⟨f′,ψ⟩ ∈ X.

It is widely accepted that these sorts of definitions do satisfy Convention T, and so we surely have a definition suitable for a predicate which expresses the concept of *a true sentence of L.*

Translingual truth predicates

Tarski succeeded in following the T-strategy to its completion, solving the problem of generality by showing how to give a recursive definition, and this rises to the technical challenge we identified at the outset. However, the sample definition Tarski gave is apt only for a predicate expressing the monolingual concept of *a true sentence of the language of the calculus of classes.* What about the relational concept of sentential truth? It has often been said in criticism of Tarski that he showed "how to define 'is a true sentence of L' for fixed L," but failed to show us how to define the relational 'is a true sentence of' – the implication being that it is only the latter that expresses a concept in which philosophers are *really* interested.

Yet, Tarski himself evidently thought that his technique could be generalized. Indeed, there would seem to be no barrier, in principle, to the construction of a definition suitable for a two-place predicate, like 'is a true sentence of' *provided that* the language, M, in which the definition is to be given not be one for which the premises of the Inconsistency Argument hold. This indirectly imposes a constraint on the object lan-

guages over which the target truth predicate can range. There may be real difficulties posed if the range of object languages is infinite, but such difficulties do not show that there are no suitable metalanguages, M.

The complaint we are considering is really misplaced. What is true is that the language in which a relational truth predicate were defined in the Tarskian way would have to be a language with a restricted domain of discourse. However, this is no more than the demand for consistency requires.

4   Discussion

The question of analysis

Does a Tarskian definition like the one just described provide a conceptual analysis of the concept of sentential truth? Certainly not. One of Tarski's main aims in giving a definition is to ensure consistency with empirical facts. But the upshot of the Inconsistency Argument is that the concept of sentential truth is incoherent in such a way that anything that might pass for an *analytical* definition would surely not be consistent in this way. So, it is very plainly not on the Tarskian agenda to provide an analytical definition. In fact, Tarski understands himself to be *defining the set* which is the extension, not a predicate at all, so the definitional part of his project is not about giving meanings (Coffa 1991: 293–6).

Still, it may come as a surprise that even a Tarskian definition for a semantically open language like Quadling could not at least give us an analysis of the humble concept of *a true sentence of Quadling*. Nonetheless, the denial of such analytic status is implied by a family of arguments promulgated in the literature. If successful, these arguments would show that if you *introduced a new predicate* using a Tarskian definition, this predicate would not mean the same thing as an antecedently meaningful truth predicate that expressed one of our truth concepts. These arguments proceed by comparing an example T-sentence to the result of performing definitional substitution on that T-sentence using a Tarskian truth definition. The arguments seek to impugn the Tarskian definition by finding telltale differences between the two sentences that point to differences in meaning. Philosophers have claimed differences in logical status, modal status, subject matter, and informativeness. Arguments of this sort can be found in Putnam (1985: 63–4); Soames (1995: 253–4). John Etchmendy (1988: 56–7) seeks to use this sort of argument to draw a further conclusion, namely that Tarskian definitions do not give any information about the semantics of their target language, appearances notwithstanding. If this were correct, Tarski (1936) would not have contributed to theoretical semantics in the way he is widely thought to have. For critical discussion of these arguments, see Davidson (1990: 288–95); Garcia-Carpintero (1996); Heck (1997).

Even if these arguments are correct, it would be a mistake to conclude as some do that Tarskian definitions are merely extensionally correct or that Tarski thought only as much. The appropriateness of the *T-strategy* depends on there being a significant conceptual connection between the concept in question and the T-sentences, but Tarski's

notion of *definition* does not itself require 'giving the concept.' Yet, since Tarskian definitions realize the T-strategy, they inherit something of that conceptual connection. *If one satisfied the knowledge conditions set forth in the definition of subtle conceptual warrant, and one knew that the language M was so chosen that the truth predicate would be subserved by the concept it expresses, then one would be in a position to know that each of these T-sentences is true in M.* Ditto for the Tarskian truth definition statement. For this reason, we might with some justice say that a Tarskian truth definition for L 'captures' the concept of *a true sentence of L*, even though the definition is not a concept-giving one.

Deflationism

The Tarskian view is often associated with deflationist views of truth. According to some, the central tenet of *deflationism* is that there is no property of truth. If this is what is meant by deflationism, then the Tarskian view of truth that we have outlined is most certainly deflationist. In fact, Tarski was in possession of the best possible reason for endorsing this deflationist thesis. Tarski's view of truth delivers a simple argument to the conclusion that there is no *property (relation) of sentential truth*. The argument I have in mind relies on two general principles:

1. For any binary relation, R, necessarily, for any pair, x and y, either x is R-related to y or x fails to be R-related to y.
2. For any R,p,M, if R is the relation of sentential truth, and p is a predicate which expresses in M the concept of sentential truth, then for every x and y in the domain of discourse of M, p applies in M to $\langle x,y \rangle$ iff x is R-related to y.

Item (1) states a conceptual truth about properties; that is just the sort of thing that a property was supposed to be. (2) articulates how the concept of sentential truth and the property thereof would be related to a predicate that expressed that concept. Simply put, a predicate is supposed to be 'underwritten' by the property (if any) associated with the concept the predicate expresses.

Now, we reason as follows. By way of contradiction, suppose there is a property of sentential truth, R. Then it is easy to see that (1) and (2) enable us to infer that any predicate expressing the concept of sentential truth will have a proper extension, but we know this is not so. Another way of seeing the point is to see that our supposition, together with (1) and (2) ensure that premise (a) of the Liar Argument is true in M, and we know that cannot be so. Therefore, by reductio, there is no property of sentential truth. This reasoning can be repeated for monolingual truth properties such as *being a true sentence of French.*

Thus, it is not hard to muster a deflationist conclusion from the Tarskian view as we have developed it here. However, sometimes 'deflationism' is associated with the rather more nebulous idea that truth is not a philosophically significant notion. Nothing we have said suggests that one who held Tarski's view should be a deflationist in *this* sense. Indeed, it looks like the view may be committed in quite the opposite way. Tarskian definitions don't give an analysis of our concept of sentential truth nor even of its sub-

174

sidiary monolingual notions. This *suggests* that there may be something to these concepts other than what the T-sentences codify. Working out a cogent *disquotationalist* version of deflationism could be seen as the attempt to resist this suggestion. Space prohibits further discussion here, but a *locus classicus* of the debate is Field (1987), and a useful collection in this area is Blackburn and Simmons (1999). For an extended discussion of the disquotationalist proposal, see David (1994).

Making truth safe for science

Even though his analysis of the Liar pointed to the incoherence of the concept of sentential truth, Tarski nonetheless saw value in carrying out his definitional project in a restricted context. By doing so, he showed that one *can* appeal to the notion of sentential truth in the conduct of inquiry without fear of introducing inconsistency. In so far as the restricted class of languages in which this applies is broad enough for the conduct of scientific inquiry (is it?).

Tarski succeeds in *making truth safe for science.* Note that success in this does not require that Tarskian definitions be analytical, nor, *contra* Field (1972), that they offer any sort of physicalistic reduction of semantics.

5   Conclusion

We have presupposed throughout our discussion that there *is* a concept of sentential truth and that, by dint of certain linguistic intentions, terms like 'is a true sentence of' express it. But we should now step back and examine this presupposition. To tell the story we have, we tacitly employed a conception of concept according to which, by making the term 'true' express a certain concept, we would make it the case that the word 'true' is *supposed to work a certain way,* it is *supposed to* apply to certain things and it is *supposed to* fail to apply to certain other things, *whether the term for whatever reason actually succeeds in applying (failing to apply) to those things or not.* We appealed earlier to the fiction that concepts were things that came equipped with explicit application rules to make this idea concrete. On this conception of concept, there is nothing funny about speaking, as we have, of an incoherent concept.

However, philosophers have commitments about these things and there is a history to the use of the word 'concept.' Some will hold that concepts are more akin to properties as I have characterized them. That is to say, some will hold that a concept is a thing that partitions the things of the universe into two classes – those that fall under the concept and those excluded by it – and if if there isn't such a partition, you don't have a concept. Earlier reasoning showed, however, that there can be no such partition when it comes to sentential truth. Thus, if we use 'concept' in this narrower sense, the Tarskian view will certainly force us to say that there is not even a concept of sentential truth. Instead, we should have to say, there is only a kind of predicate which, by dint of our linguistic intentions, is *supposed* to work this way and is *supposed* to work that way, when in fact *nothing could possibly work* this way *and* that way.

With such dramatic philosophical conclusions as these in the offing, it is easy to see how Tarski's work could become a centerpiece of the discourse on truth.

175

References

Coffa, J. A. (1991) *Semantic Tradition from Kant to Carnap: to the Vienna station*. Cambridge: Cambridge University Press.

David, M. (1994) *Correspondence and Disquotation*. New York: Oxford University Press.

Davidson, D. (1967) Truth and meaning. Reprinted in Davidson, D. (1984) *Inquiries into Truth and Interpretation* (pp. 17–36). Oxford: Clarendon Press.

Davidson, D. (1990) The structure and content of truth. *Journal of Philosophy*, 87, 279–328.

Etchemendy, J. (1988) Tarski on truth and logical consequence. *Journal of Symbolic Logic*, 53, 51–79.

Field, H. (1972) Tarski's theory of truth. *Journal of Philosophy*, 69, 347–75.

Field, H. (1987) The deflationary conception of truth. In G. MacDonald and C. Wright (eds.), *Fact, Science and Morality* (pp. 55–117). Oxford: Blackwell Publishers.

Garcia-Carpintero, M. (1996) What is a Tarskian definition of truth? *Philosophical Studies*, 82, 113–44.

Heck, R. Jr. (1997) Tarski, truth, and semantics. *Philosophical Review*, 106, 533–54.

Kripke, S. (1975) Outline of a theory of truth. Reprinted in R. L. Martin (ed). (1984), *Recent Essays on Truth and the Liar Paradox* (pp. 53–81). Oxford: Oxford University Press.

Putnam, H. (1985) A comparison of something with something else. *New Literary History*, 17, 61–79.

Ramsey, F. P. (1929) On truth. Published posthumously in N. Rescher and U. Majer (eds.) (1991), *Episteme* (vol. 16). Dordrecht: Kluwer Academic Publishers.

Soames, S. (1995) T-sentences. In W. Sinnott-Armstrong, D. Raffman and N. Asher (eds.), *Modality, Morality, and Belief: Essays in Honor of Ruth Barcan Marcus* (pp. 250–70). Cambridge: Cambridge University Press.

Soames, S. (1999) *Understanding Truth*. New York: Oxford University Press.

Tarski, A. (1936) The establishment of scientific semantics. Reprinted in translation in J. Corcoran (ed.) (1983), *Logic, Semantics, Metamathematics*. 2nd edn. (pp. 401–8). Indianapolis: Hackett.

Yablo, S. (1993) Paradox without self-reference. *Analysis*, 53, 251–2.

Further Reading

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

Chihara, C. (1979) Semantic paradoxes: a diagnostic investigation. *Philosophical Review*, 88, 590–618.

Horwich, P. (ed.). (1994) *Theories of Truth*. Aldershot: Dartmouth.

Kirkham, R. L. (1992) *Theories of Truth: A Critical Introduction*. Cambridge, MA: MIT Press.

Tarski, A (1933) On the concept of truth in formalized languages. Reprinted in translation in J. Corcoran (ed.) (1983), *Logic, Semantics, Metamathematics*. 2nd edn (pp. 152–278). Indianapolis: Hackett.

Tarski, A. (1944) The semantic conception of truth and the foundations of semantics. *Philosophy and Phenomenological Research*, 4, 341–75.