Chapter 1

ASPECTS OF PARADOX

1.1. The liar and its variants

The liar is the paradox of the sentence that denies its own truth. In at least the usual versions of the paradox we have a sentence that either directly or indirectly denies its truth. In a typical form, the liar concerns the sentence (L):

(L) is not true.

We can easily end up in contradiction about that sentence.

One line of reasoning that leads to contradiction relies on the schema

(T) **S** is true iff *p*.

To get an instance of the schema, we must replace the letter '*p*' with a declarative sentence and the letter '**S**' with a name of that sentence. The name replacing '**S**' may either consist in the sentence itself put inside quotation marks or be different. However, I will call (T) 'disquotational' in the sense that, in each instance, an expression (sentence) is mentioned on the one side and used on the other. The schema (T) appears to be a principle that characterizes the concept of truth, since it captures the idea that a sentence is true iff things are as it says they are.

One instance of (T) is the biconditional

(L) is true iff (L) is not true.

Assume that (L) is true. Then, because of the biconditional, it is not true. Hence, by *reductio ad absurdum*, we can deny the assumption: (L) is not true. Consequently, because of the biconditional again, it is true—contradiction.

Instead of relying on (T), we can reach a contradiction by invoking two rules of inference: the rule that allows us to infer from a declarative sentence to calling it 'true' and the converse rule, which allows us to infer from calling such a sentence 'true' to the sentence itself. In classical logic those two rules, which we can name 'true-in' and 'true-out' respectively, are equivalent to (T). For, in classical logic, accepting the inference from a sentence **A** to a sentence **B** is equivalent to endorsing the conditional ⌜If **A** then **B**⌝. And, of course, accepting a rule of inference amounts to accepting all the inferences that conform with it, and endorsing a schema amounts to endorsing all its instances. So true-in is equivalent

to the schema 'If p, then **S** is true', which captures the one direction of (T), and true-out is equivalent to the schema 'If **S** is true, then p', which captures the opposite direction. On the other hand, according to some other logical systems, accepting the inference from **A** to **B** does not by itself commit one to the conditional. Indeed, the logics presented in this book will be of that kind. Thus some theories of truth, which rely on such systems, endorse true-in and true-out but not (T). The theories developed in this book endorse both the rules and the schema.

The paradox has many versions. The sentence (L) is characterized by self-reference, but the paradox also arises without self-reference. In the sentences

(1) (2) is true
(2) (1) is not true,

we have circular reference and not self-reference, in that (1) contains a term denoting (2) and none denoting (1) while (2) contains a term denoting (1) and none denoting (2). By the schema (T), (1) is true iff (2) is true, and also (2) is true iff (1) is not true. Therefore, (1) is true iff it is not true, and the contradiction arises like before. Or again, take sentence

(3) There is a sentence that is written in line 25 of p. 1 and is not true.

Assume that the only sentence written in line 25 of p. 1 is (3). Then (there is a sentence that is written in line 25 of p. 1 and is not true) iff (3) is not true. By (T), (3) is true iff there is a sentence that is written in line 25 of p. 1 and is not true. Hence (3) is true iff it is not. But (3) does not refer to itself. It involves quantification over sentences, rather than reference to a sentence. The relation between (3) and itself is the relation between the sentence 'There is a person waiting at the stop' and Yannis Stephanou if he is waiting at the stop. Even if he is the only person waiting at the stop, the sentence does not refer to him, since it contains no name or other expression that denotes him.

(3) is one of the so-called *contingent* versions of the liar. In other words, things could be such that (3) was not involved in paradox, yet had the sense it also has actually. If one had not written (3) in line 25 of p. 1, but the sentence 'Naples is the capital of Italy', then (3) would not be involved in paradox; it would simply be true. It would also be true if both (3) and 'Naples is the capital of Italy' were written in line 25 next to each other. Of course, (L) and (1)–(2) are not contingent versions. A famous contingent example, due to Kripke [1984, pp. 54–55], consists of the sentences

(4) Everything Jones says about Watergate is true
(5) Most of the things that Nixon says about Watergate are untrue.

Let's imagine the following circumstances: Jones utters sentence (5) and says nothing else about Watergate; Nixon utters (4), and of the other things he says about Watergate, exactly half are true and the other half untrue. Then, by (T) and the circumstances, (4) is true iff (5) is true, and (5) is true iff (4) is not true. So the pair (4)–(5) becomes like (1)–(2). Yet in other circumstances no problem arises.

(4) and (5) could easily both be true, and they could easily both be false. They show that ordinary sentences about truth which in normal circumstances involve no contradiction can, in unusual but possible circumstances, become as paradoxical as (L). So the problem cannot be confined to a specifiable set of odd sentences, such as (L).

Some versions of the liar involve falsity rather than truth. Take the sentence

(6) (6) is false.

This sentence attributes falsity to itself. By (T), (6) is true iff (6) is false. This biconditional sounds paradoxical to many people, but in fact does not lead to contradiction unless supplemented with other principles about truth and falsity. One principle that seems reasonable is that if a sentence is false, then it is not true. So, by contraposition, if a sentence is true, it is not false. Assume that (6) is true; then by the biconditional it is false, and so, according to the principle just mentioned, it is not true. Hence, by *reductio*, the sentence is not true. Now assume that it is false; then by the biconditional it is true, so it is not false. Hence it is not false. (6) is neither true nor false. If we accept the principle of bivalence and believe that (6) is either true or false, we reach a contradiction. We may, however, reject bivalence and consider that (6) has no truth-value. Sometimes, the appellation 'liar' is restricted to a sentence like (6), and (L) is called 'a strengthened liar'. The idea is that in the case of (6) we can avoid contradiction by rejecting bivalence, but at least *prima facie* that option does not seem to help with (L).

In fact, rejection of bivalence does not seem to help with (6) either if we endorse the schema

(F) **S** is false iff not-p.

To get an instance of that schema, we must replace '**S**' with a name of a declarative sentence x and 'not-p' with a sentence that negates x. (F) captures the idea that a sentence is false iff things are not as it says they are. So it is as central to our understanding of falsity as (T) is to our understanding of truth. One instance of (F) is the biconditional '(6) is false iff (6) is not false', from which familiar steps lead to contradiction.

Also, if we adhere to (T) and accept that (6) is neither true nor false, we must be careful in our treatment of negation. If (6) is not false, then by (T)

(6′) (6) is not false

is true. But (6′) negates (6). So we must allow for sentences that are neither true nor false, but whose negation is true. We should abandon the principle that a sentence not-**A** is true iff **A** is false. We may instead invoke another principle of standard semantics: not-**A** is true iff **A** is not true. At any rate, if we accept that (6) is neither true nor false, we must abandon one or other of those principles, whether or not we adhere to (T). Otherwise, we shall be committed to saying that (6′) is both true and not true: it is true because (6) is not true, and it is untrue because (6) is not false.

There are paradoxes about truth other than the liar. One of them is Curry's paradox [Curry, 1942]. Let the symbol '$\perp$' abbreviate any absurd sentence you want, e.g., the sentence '$3+1=5$', and think about this:

(C) If (C) is true, then $\perp$.

We can derive an absurd conclusion if we invoke both the schema (T) and the rule of conditional proof, that is, the rule that allows us, when we have made an assumption **A** and drawn a conclusion **B** within its scope, to infer (outside the scope of the assumption) the conditional ⌜If **A** then **B**⌝. By (T) we have:

(C) is true iff (if (C) is true, then $\perp$).

Assume that (C) is true. In that case, taking the left-to-right direction of the above biconditional, if (C) is true then $\perp$. Hence, by *modus ponens*, $\perp$. Therefore, by the rule of conditional proof, if (C) is true then $\perp$. Thus, now taking the right-to-left direction of the biconditional, (C) is true. Hence, by *modus ponens*, $\perp$. We have concluded that $3+1=5$. As Curry's paradox does not involve the concept of negation, it refutes the idea that the paradoxes arise from some problem in that concept.[1]

Another paradox about truth is Yablo's [Yablo, 1985, p. 340, and 1993], which involves the following sequence of sentences:

(Sentence 1)	For every $n>1$, sentence n is not true
⋮	⋮
(Sentence m)	For every $n>m$, sentence n is not true
(Sentence $m+1$)	For every $n>(m+1)$, sentence n is not true
⋮	⋮

Assume that, for some number m, sentence m is true. Then, by (T), for every $n>m$, sentence n is not true; so, in particular, sentence $m+1$ is not true; hence, by (T) again, it is not the case that, for every $n>(m+1)$, sentence n is not true; then, there is a number $n>(m+1)$ such that sentence n is true, while it is also not true, which is absurd. Thus, by *reductio*, we may deny the assumption: for every number m, sentence m is not true. Therefore, sentence 1 is not true. But also, for every $n>1$, sentence n is not true. Hence, by (T), sentence 1 is true—contradiction. In contrast with the liar, Yablo's paradox does not seem to involve any circularity.[2]

---

[1] One may consider that, for any sentence **A**, ⌜$\neg$**A**⌝ means ⌜If **A** then $\perp$⌝ and conclude that Curry's paradox involves the concept of negation. In fact, ⌜$\neg$**A**⌝ does not mean that. First, it may be defined as ⌜**A** $\rightarrow \perp$⌝, but such a definition does not capture the meaning of '$\neg$' any more than the definition of ⌜**A** $\wedge$ **B**⌝ as ⌜$\neg$[$\neg$**A** $\vee$ $\neg$**B**]⌝ captures the meaning of '$\wedge$'. Definitions of that kind are just a device for decreasing the number of primitive symbols in a logical system. Secondly, '$\perp$' may mean '$3+1=5$', but ⌜$\neg$**A**⌝ does not mean ⌜If **A** then $3+1=5$⌝. Why should it mean that and not ⌜If **A** then $2=1$⌝? And if we do not give the symbol '$\perp$' any particular sentential meaning, we have given no meaning to the sentence ⌜If **A** then $\perp$⌝ and so there is no candidate for synonymy with ⌜$\neg$**A**⌝.

[2] Priest [1997] began a discussion about whether Yablo's paradox really involves no circularity. [Cook, 2014] is a book-length study of the paradox.

Yablo's paradox, Curry's and the liar are all semantic paradoxes, since the concept of truth is a semantic notion. There are paradoxes that employ other semantic notions, such as reference, satisfaction, and so on. One of them, akin to the liar, is Grelling's paradox [Grelling and Nelson, 1908, p. 307]. Let us stipulate that a predicate is *heterological* iff it does not satisfy itself; in other words, it is heterological just in case it is not true of itself. So 'long' is heterological, but 'short' is not; for 'long' is not a long predicate, but 'short' is a short one. At least for monadic predicates, the concept of satisfaction seems to be characterized by the following schema:

(S) For everything x, x satisfies **G** iff x is F,

where the letter '*F*' is to be replaced by such a predicate while '**G**' is to be replaced by a name of that predicate. For instance, anything satisfies 'long' iff it is long, and anything satisfies the predicate 'heterological' iff it is heterological. Thus if 'heterological' is heterological, then by the definition of 'heterological' it does not satisfy itself, and so by (S) it is not heterological. If, on the other hand, it is not heterological, then by the definition it satisfies itself, and so it is heterological. The predicate is heterological just in case it is not.

A satisfactory treatment of the liar should tell us, on the one hand, where the blame lies for the problem and, on the other, how we can overcome it. In other words, it should both offer a diagnosis and suggest a therapy.[3] Also, it ought to be able to deal with all the versions of the paradox and, if possible, with the other semantic paradoxes as well. Setting aside the approach which endorses contradictions and considers that the sentence (L) and the other paradoxical sentences both are and are not true, we can say that a satisfactory treatment should explain, on the one hand, where the blame lies for the production of contradictions and, on the other, how we can avoid them. We end up in contradictions relying on the schema (T) and using classical logic. (In the contingent versions, we also presuppose some facts that cannot be doubted and are to do with who says what, what is written where, and the like.) Much of the work that has been done on the paradox develops, in various ways, the idea that the blame lies in (T). My own work is included in the approach on which the blame lies in classical logic.

1.2. Propositions and truth-values

Those who consider that the blame for the problem lies in (T) do not, of course, need to reject the schema entirely. They can argue that we ought to restrict it: refuse to accept its instances that concern paradoxical sentences, but accept its instances that concern other sentences. Paradoxical sentences are just those in whose case we are led to contradiction. (L), (1) and (2) are paradoxical. (4) and (5), too, would be paradoxical if the appropriate circumstances obtained. That

[3]The distinction between the two tasks is essentially that made in [Chihara, 1979, pp. 590–591].

approach to (T) is particularly sensible if there are reasons to exclude paradoxical sentences from (T) which are independent of the fact that, in the case of those sentences, we are led to contradiction. But are there any such reasons?

1.2.1. No proposition expressed. One may argue that we should not apply (T) to sentences that express no proposition. In other words, we should not apply it to sentences that express no information about how things are.[4] The schema (T) is a formulation of the idea that a sentence is true iff things are as the sentence says they are. Obviously, the idea concerns only sentences that express information about how things are. Correspondingly, schema (T) should be accepted only to the extent that it concerns such sentences. But paradoxical sentences, one may continue, do not express propositions, do not convey information. Hence, we should not apply (T) to paradoxical sentences.

Indeed, one may offer the following argument in order to show that paradoxical sentences do not express propositions. It is not the case that (the sentence (L) is true iff it is not true). For if we have any biconditional of the form '*p* iff not-*p*', its negation is an instance of a logical law. Likewise, either it is not the case that (sentence (1) is true iff (2) is true) or it is not the case that (sentence (2) is true iff (1) is not true). Also, if the only sentence written in line 25 of p. 1 is (3), we must deny that (3) is true just in case there is a sentence that is written in line 25 of p. 1 and is not true. Other paradoxical sentences are similar. But then, paradoxical sentences have no truth-conditions. For if, e.g., (3) has a truth-condition, the condition can only be that there is a sentence that is written in line 25 of p. 1 and is not true. If paradoxical sentences have no truth-conditions, then they do not express propositions. A sentence expressing a proposition conveys information about how things are and so has a truth-condition.[5]

In my opinion, paradoxical sentences express propositions, although the opposite view is quite common; see, e.g., [Kripke, 1984, pp. 63–64]. The argument in the preceding paragraph ignores the possibility that classical logic may not apply to paradoxical sentences. Let's take (L). It is a negation of the following sentence:

---

[4] A *proposition* is such a piece of information. The information may be correct, even tautological; but also it may be wrong, even absurd. Declarative sentences, at least normally, express propositions. For example, the proposition that Kant is the most important German philosopher is expressed by the sentence 'Kant is the most important German philosopher' and every synonymous sentence of either English or another language. Interrogative and imperative sentences do not express propositions.

[5] One may claim that it is a misuse of 'true' to call a sentence 'true': only propositions can be appropriately so called, so all the sentences we have discussed present a category mistake. But one gains nothing by claiming that. For if only propositions can be appropriately called 'true', we can replace (L) and (T) with

(L^P) (L^P) does not express a true proposition

and

(T^P) **S** expresses a true proposition iff *p*.

We can similarly modify the other sentences we have discussed. Then, we shall face the various versions of the paradox again.

(L′) (L) is true.

If (L′) expresses a proposition, a piece of information, then (L) also expresses a proposition, the negation of the one expressed by (L′). It is clear what (L′) refers to: it refers to a certain series of words which is syntactically structured in a particular way and so is a sentence. It is clear what it says about that series: that it is a true sentence. So (L′) is an attribution of a property to a linguistic entity. How can it fail to express information about that entity?

The idea that paradoxical sentences do not express propositions ceases to be attractive as soon as we realize that there are propositions which are similar to paradoxical sentences and are paradoxical themselves. In order to reach a contradiction that concerns such propositions, we need the schema

(TP) *P* is true iff *p*.

We get an instance of (TP) when we replace the letter '*P*' with a term that refers to a proposition and the letter '*p*' with a sentence that expresses just that proposition. For example,

> The proposition expressed by the sentence 'Snow is white' is true iff snow is white.

The schema (TP) is the analogue of (T) for propositions.

We can see the sentence

(7) The proposition expressed by (7) is untrue.

Does (7) express a proposition? Yes, if there is a proposition to the effect that the proposition expressed by (7) is untrue. Is there such a proposition? Yes, at least according to the following argument from [Horwich, 1998, p. 41]: For any condition *C*, there could be someone believing that the proposition that satisfies *C* is untrue. Whatever could be believed by someone is a proposition. Thus, for any condition *C*, there is a proposition to the effect that the proposition that satisfies *C* is untrue. Hence there is a proposition to the effect that the proposition expressed by (7) is untrue. Moreover, there cannot be two propositions to that effect, so (7) expresses just one proposition.

It may be objected that unless we have a deflationary concept of a proposition, we should not accept that whatever could be believed by someone is a proposition; someone could believe that he was in tune with the universe, but surely there is no such proposition in any substantial sense of the term. My concept of a proposition is just the concept of a piece of information; it is not more substantial than that. Even if one doubts that whatever could be believed is a proposition, one should not doubt that if someone may tell someone else that things are a certain way, then there is a piece of information, and so a proposition, to the effect that things are that way. To tell someone that *p* is to convey information. And we may be told that the proposition expressed by (7) is untrue.

Let's call the proposition expressed by (7) 'Π'. According to (TP), Π is true iff the proposition expressed by (7) is untrue. Therefore, Π is true iff it is untrue. Π

is similar to the sentence (L). Just like (L), it refers to itself and denies that it is true. The only difference is that (L) refers to itself by means of a name, whereas Π refers to itself by means of a description; it refers to itself as the proposition expressed by a certain sentence. (On the other hand, (7) refers to itself by means of a name.)

We can also see the sentence

(8) There is a proposition that is expressed by (8) and is not true.

Does (8) express a proposition? Yes, as we can see by comparing (8) to the sentence

(9) Some proposition is untrue and is expressed by (8).

(8) and (9) are distinct sentences, since they are not made up of the same words. (9) does not refer to itself, and this may make it easier to study. It seems clear to me that (9) expresses one (and only one) proposition. It is the information that some proposition satisfies two particular conditions (it is untrue, and it is expressed in a certain sentence). But sentences (8) and (9) are synonymous; they have no difference in sense. The one results from the other through small changes in wording alone. Therefore, (8), too, expresses a proposition, just the one expressed by (9). (Incidentally, here is a moral: self-reference is not an aspect of the sense of a self-referring sentence, since it is possible for another sentence to have just the same sense without being self-referring.)

Let's call the (only) proposition expressed by (8) 'Σ'. Thus (there is a proposition that is expressed by (8) and is not true) iff Σ is not true. By (TP), Σ is true iff (there is a proposition that is expressed by (8) and is not true). Hence, Σ is true iff it is not. Σ is similar to sentence (3). Σ says that there is a proposition which, on the one hand, satisfies a certain condition and, on the other, is not true; and Σ itself is the only proposition that satisfies the condition in question. Correspondingly, (3) says that there is a sentence which, on the one hand, satisfies a particular condition and, on the other, is not true; and (3) itself is the only sentence satisfying that condition. The similarity gets up to the contingency in the satisfaction of the relevant condition: just as (3) might not be written in line 25 of p. 1, so Σ might not be expressed by (8), since (8) could have an entirely different sense from the one it actually has and so fail to express Σ.

The proposition expressed by the liar sentence (L) is also paradoxical. Let's call it 'Λ'. Λ is not a self-referring proposition, since it refers to a certain sentence and not to itself or any other proposition. In that respect, it is less similar to (L) than Π is. By the schema (TP) we have

Λ is true iff (L) is not true.

In order to reach a contradiction about Λ, we can invoke the principle

> For every sentence *S* and every proposition *P*, if *P* is the proposition expressed by *S*, then *S* is true iff *P* is true.

By that principle, (L) is true iff Λ is true, and so (L) is not true iff Λ is not true. Hence, Λ is true iff it is not.

When one realizes that there are propositions like Π and Σ, one has no motivation any more to assert that sentences such as (L) and (3) express no proposition. The idea that paradoxical sentences express no proposition is appealing when one hopes that it will lead to a principled treatment of the liar and kindred paradoxes. But we cannot transpose the idea into the case of paradoxical propositions, so we cannot use it to tackle the problem in their case, so we cannot use it to tackle the problem in general.[6]

1.2.2. Lack of truth-value. Varying the idea that paradoxical sentences express no proposition, one may invoke another reason to exclude them from schema (T) which is independent of the fact that, in the case of those sentences, we are led to contradictions. One may argue that we should not apply (T) to any sentence that has no truth-value. For if we do, then the right hand-side of the resulting biconditional will be the sentence itself, so it will be neither true nor false. But the left-hand side, which describes the sentence as being true, will be just wrong, that is, false. And then, according to the usual systems of three-valued logic, the biconditional will not be true; it will be neither true nor false. Paradoxical sentences, one may continue, have no truth-value. Hence we should not apply (T) to such a sentence.

But why say that paradoxical sentences have no truth-value? Here it may be argued that paradoxical sentences are *ungrounded* and that ungrounded sentences, whether paradoxical or not, lack truth-value.[7] Sentence

(10) Snow is white

is not about truth or falsity, and its truth-value has consequences for other sentences. So

(10′) (10) is true

is true because (10) is true. The truth-value of (10) determines a truth-value for (10′). It also determines a truth-value for

(10″) (10′) is true
(10‴) (10″) is true
⋮ ⋮

as well as

(10+) Every sentence named by the numeral '(10)' with or without primes is true.

Sentences (10′), (10″), . . . , (10+) are all true because (10) is true. A sentence *s* is grounded provided either it has a truth-value without being about truth or falsity (or other semantic issues) or there are some other sentences (one or more) which

[6]Burgess and Burgess [2011, p. 119] argue that paradoxical sentences can be used in explanation of action and so should express propositions.

[7]Herzberger [1970, pp. 147–151] first introduced a notion of groundedness. The explanation provided here is different from his.

are not about truth or falsity and whose truth-values determine a truth-value for *s*. (10), (10′), (10″), . . . , (10+) are grounded. If the sentence 'Naples is the capital of Italy' were written in line 25 of p. 1, then the truth-value of that sentence, together with the fact that it is written in a certain line, would determine a truth-value for (3); this would be sufficient for (3) to count as grounded.

On the other hand, let's see sentence

(11) (11) is true.

(11) is called a *truth-teller*. It is not paradoxical. Neither the assumption that it is true nor the assumption that it is false leads to contradiction. But it is ungrounded; there are no sentences that are not about truth or falsity and determine a truth-value for (11). Let's also see sentences

(12) (13) is true
(13) (14) is true
(14) (12) is true.

They are not paradoxical either, but they are ungrounded, since none of them has a truth-value due to sentences that are not about truth or falsity (or other semantic issues). Whether (12) is true depends on whether (13) is true, and on whether (14) is true, but it depends on no sentence that does not involve truth. Things are similar if instead of a circle we have a descending sequence:

(15) (15′) is true
(15′) (15″) is true
⋮ ⋮

So it may be argued that if a sentence possesses a truth-value, then either the sentence is about the real world, as it were, or at least its truth-value is due to the truth-values of some sentences that are about the real world. And it is supposed that the real world is not to do with whether sentence so-and-so is true, or whether sentence so-and-so is false, but with whether snow is white, whether there are extraterrestrials, and the like. If so, ungrounded sentences possess no truth-value.

Indeed, paradoxical sentences are ungrounded. Whether (L) is true depends on whether it itself is not true, but it depends on no sentence that does not involve truth or falsity. Whether (1) is true depends on whether (2) is true, and thus on whether (1) itself is not true, but it depends on no statement that is not about truth. I know of no variant of the liar that concerns a grounded sentence. However, it is difficult to cling to the view that all ungrounded sentences lack truth-value. Take the ungrounded sentence

(16) If (16) is true, then (16) is true.

It has the form 'if *p* then *p*'. Sentences of that form are entirely tautological, so we should not doubt that they are true. (16) is, therefore, true. Or take the ungrounded sentence

(17) (17) is true and (17) is not true.

(17) is a contradiction, as it has the form 'p and not-p'. So I believe we should recognize it as false. Also see the following ungrounded sentences, which are based on criticism (3) from [Gupta, 1982, pp. 34–35]:

(18) (18) and (19) are both true

(19) It is not the case that (18) and (19) are both true.

The following principle is as reasonable as the law of non-contradiction:

> For any sentence S and any sentence S', if S' is a negation of S, then it is not the case that S and S' are both true.

(19) is a negation of (18). Hence it is not the case that (18) and (19) are both true. We have just concluded and accepted (19) itself. Once we have reached that point, it is difficult not to make the next small step, accepting that (19) is true. And of course if (19) is true, (18) is false. None of the examples (16)–(19) leads to paradox. If, now, one abandons the general view that ungrounded sentences lack truth-value, I do not see how one can any more sustain the claim that paradoxical sentences lack truth-value (sustain it, that is, without resorting to the fact that, in the case of paradoxical sentences, we are led to contradictions).

At any rate, if we accept that paradoxical sentences are neither true nor false, we must be careful with the question 'If a sentence is neither true nor false, what is the truth-value of calling it "true", and what is the truth-value of calling it "false"?' When we are considering sentences that are allegedly neither true nor false, but in which the concepts of truth and falsehood are not expressed, the most appealing answer to that question is 'Falsehood'. For example, if, following [Strawson, 1993], we accept that

(20) The king of France is wise

is neither true nor false, it is natural to say that both the statement '(20) is true' and the statement '(20) is false' are false. Indeed, that is what we must say if we endorse the schema (F). But when we declare (6) to be neither true nor false, we should not say that the truth-value of calling it 'false' is falsehood. By saying that, we shall inconsistently accept that (6) is false after all. And when we declare (L) to be neither true nor false, we should not say that the truth-value of calling it 'true' is falsehood. For, if so, then '(L) is true' is false, and hence (L), which negates '(L) is true', is true. So it may be better to consider that if a paradoxical sentence is neither true nor false, then to call it 'true', or to call it 'false', is also to say something that is neither true nor false. But then again, this view undermines the argument we saw at the beginning of the current section against applying (T) to sentences that have no truth-value.

Moreover, there is a serious difficulty with the view that (L) lacks truth-value. If it has no truth-value, (L) is not true (and, of course, it is not false either). I have just written (L) itself. The view that (L) has no truth-value commits us to asserting (L). Now, asserting (L) is incoherent. The problem is not that if one asserts '(L) is not true', one takes the first step towards a contradiction because in a short while one

will be compelled to add '(L) is true'. Such a step is taken by whoever is willing to apply the schema (T) or the rule true-in to (L), for either the schema or the rule allows us to infer from '(L) is not true' to '(L) is true'. But one may believe that (L) has no truth-value, so it is not true, and also believe that (T) should not be applied to sentences having no truth-value and that we should not infer from such a sentence to calling it 'true'. Then, one avoids contradiction. (And it is such a combination of views that we are currently discussing.) The problem is that at least provided a sentence is not ambiguous or indexical, it is incoherent to assert it and also assert that it is not true. (Ambiguity and indexicality induce complications, as we shall see in Chapter 3.)

The incoherence is not a contradiction; one does not assert both the sentence and its negation, nor does one assert that the sentence is both true and not true. But it is similar to a contradiction. When some philosophers assert '(L) is not true', they do both: they both assert a sentence—(L)—and assert that the sentence is not true. Of course, they do them by means of the same act, uttering (L), but at any rate they do them. The incoherence stems from the fact that truth is a basic norm governing assertion. If what is asserted is true, this counts as a success for the speaker: she is right. If it is false, this counts as a failure: she is wrong. Given how central truth is to our practice of evaluating assertions, whoever makes an assertion lays claim to saying something true. So it is incoherent to add that it is not true. The incoherence is independent of (T) and of any willingness to infer from a sentence to calling it 'true'.[8]

1.3. The problem in a formal setting

Propositional and first-order logic include no principle about truth. Their symbolic languages contain no symbol meaning 'true', while of course they contain symbols for negation, conjunction, disjunction, universal and existential quantification, etc. It is another matter that we use the concept of truth in the metalanguage in which we talk about a logical system and about its language.

On the other hand, we often enrich the language of first-order logic with various additional symbols and, with their help, develop formal theories on various topics. Such a theory is not part of logic, but presupposes some system of logic which affords the rules of inference that we follow when, in developing the theory, we construct proofs. For example, we introduce the predicate '$\in$' and use it in set

[8]Traditional emotivists and prescriptivists were happy to say things like 'Murder is evil' but deny that such sentences are true. Were they incoherent? If their semantic views on ethical terms were correct, there was no incoherence. For if a sentence such as 'Murder is evil' is a command or manifests an emotion, like an interjection, they did not perform an assertion when uttering it. It is not possible to assert a command or interjection, although it is of course possible to say it in earnest. If, on the other hand, their semantic views were incorrect, and their ethical utterances were assertions, then their position may be considered incoherent.

theory. In most formal theories, the logical system that is presupposed is classical logic.

Likewise, we can enrich the language of first-order logic with a truth-predicate 'T' (that is, a predicate that means 'true' or something similar) and some symbols expressing syntactic concepts, and we can develop a formal theory of truth in the enriched language, *Lang*. The theorems will include some general principles about truth. For example, they may include principles like 'If a sentence is T, then its double negation is T' and 'A sentence and its negation are not both T'. Also, if *Lang* possesses terms referring to its own sentences, the theorems may include many or all instances of the schema

$$(\mathrm{T}') \qquad T\mathbf{s} \leftrightarrow p,$$

where the letter 'p' is to be replaced with a sentence of *Lang* while '$\mathbf{s}$' is to be replaced with a term of *Lang* referring to that sentence.

In the formal theory of truth, one needs to be careful to avoid incoherence due to liar-like paradoxes. The liar and kindred paradoxes are the obstacle to developing a coherent and plausible theory. Contradictions arise more easily than one would think at first sight. They are inevitable if the language *Lang* allows us to form sentences like (L), the system of logic that is presupposed is classical logic, and also, for every sentence in the language and for every term in the language that refers to that sentence, the corresponding instance of the schema (T′) is a theorem. *Lang* may well allow us to form sentences like (L), that is, sentences which, in some way, refer to themselves and deny that the truth-predicate applies to them. Forming such sentences will be possible if the language contains symbols for certain syntactic concepts.

To see that, let's say that *Lang* possesses the function symbols '$\frown$', 'δ' and 'σ' (where '$\frown$' is two-place, 'δ' one-place, and 'σ' three-place) and infinitely many individual constants. The domain of the variables is the set of the expressions (that is, the strings of symbols of the language, including strings that consist in a single symbol). For every symbol of the language, there is one and only one individual constant that is a name of it. In particular, the constants 't', 'd' and 's' refer to the symbols 'T', 'δ' and 'σ' respectively, while the constant 'n' is a name of the symbol '$\neg$' of negation. Also, the constants 'v' and 'w' refer to the variables 'x' and 'y' respectively, while the constant 'u' is a name of the constant 'v'. The symbol '$\frown$' signifies the function which assigns to any expressions x and y the expression that consists of x and y in that order. So the term '$(t\frown v)$' refers to the expression 'Tx'. Each expression has one and only one *canonical designation*. For example, the canonical designation of the expression '$\neg Tx$' is '$((n\frown t)\frown v)$'. Canonical designations are of course themselves elements of the domain. The symbol 'δ' signifies the function that assigns to each expression x the canonical designation of x. So the term '$\delta((n\frown t)\frown v)$' refers to the term '$((n\frown t)\frown v)$'. Finally, the symbol '$\sigma$' signifies the function which assigns to any expression x, any symbol y and any expression z the expression that results from z by replacing y with x. So

the term '$\sigma wv(t^\frown v)$' refers to the expression that results from 'Tx' by replacing the variable 'x' with 'y'; in other words, it refers to 'Ty'.

Given all these, sentence

(21) $\neg T\sigma\delta((((((n^\frown t)^\frown s)^\frown d)^\frown v)^\frown u)^\frown v)v((((((n^\frown t)^\frown s)^\frown d)^\frown v)^\frown u)^\frown v)$

is like (L). To see that, let's turn to the expression that the term '$((((((n^\frown t)^\frown s)^\frown d)^\frown v)^\frown u)^\frown v)$' refers to, namely, the expression

(22) $\neg T\sigma\delta xvx.$

(22) is not a sentence, since it contains two free occurrences of the variable 'x'. If the truth-predicate means 'true' and not simply something similar, then (22) means 'the expression that results from the expression x by replacing the variable "x" with the canonical designation of the expression x is not true'. (21), now, tells us that the expression that results from (22) by replacing the variable 'x' with the canonical designation of (22) is not true. But (21) is the expression that results from (22) by replacing the variable 'x' with the term '$((((((n^\frown t)^\frown s)^\frown d)^\frown v)^\frown u)^\frown v)$'. And that term is the canonical designation of (22). Thus (21) refers to itself and tells us that it is not true. The difference from (L) is that (L) contains a name of itself, whereas (21) contains (from the symbol 'σ' up to its end) a composite term that describes it.

So let's say that *Lang* possesses a sentence

(23) $\neg T\mathbf{s},$

where $\mathbf{s}$ is some term that refers to (23). (21) is such a sentence. Then, $\ulcorner T\mathbf{s} \leftrightarrow \neg T\mathbf{s}\urcorner$ is an instance of the schema (T′), but those who are constructing the theory of truth are compelled, if they persist in classical logic, to exclude it from the theory. Indeed, owing to their adherence to classical logic, they will include in the theory the negation of the instance, $\ulcorner\neg[T\mathbf{s} \leftrightarrow \neg T\mathbf{s}]\urcorner$. This restriction is placed on (T′) in a theory of truth constructed in accordance with Tarski's guidelines when the language of the theory allows sentences like (23) to be formed.

If one wants to justify omitting the instance of (T′) that concerns (23) from the theory, and wants to justify it without invoking simply the need to avoid contradiction, one may say that the predicate 'T' does not mean 'true' in general. More precisely, one may specify a language L such that, even if every sentence of L is a sentence of *Lang*, still some sentences of *Lang*, such as (23), are not sentences of L. L will be like that if, for example, it is confined to the sentences of *Lang* which do not contain the predicate 'T'. Then, one may say that 'T' means 'true sentence of L' and add that, since 'T' does not concern sentences outside the language L, such as (23), we have no reason to accept the instances of schema (T′) for those sentences.

The formal theory of truth will proceed otherwise if those who are constructing it prefer to address the liar and its variants by deviating from classical logic. Let's say again that *Lang* possesses paradoxical sentences like (L), (1)–(2), etc. The system of logic presupposed by the theory will no longer be classical logic. On

the other hand, the theory should include all instances of schema (T′) or at least endorse both the rules true-in′ and true-out′; these are the rules that allow us to infer from a sentence of *Lang* to calling it '*T*' and from calling such a sentence '*T*' to the sentence itself. What interest would the theory hold if it deviated from both classical logic and the principles that characterize the concept of truth? If we deviate from those principles, we can tackle the paradoxes without abandoning classical logic.

Indeed, it is preferable if the theory includes all instances of (T′) and is not confined to the rules. Schemas like (T′) and rules like true-in′ and true-out′ are equally basic to truth. It would be unrealistic to consider that the rules are more central to our grasp of the notion than the schemas or that the latter are more central than the former. So if we confine ourselves to the rules, we have already restricted the principles that characterize the concept of truth. Once we are willing to tackle the paradoxes by diverging from classical logic, there is no motivation for imposing such a restriction.

What will the non-classical logic presupposed by such a theory be like? Unless we are willing to endorse contradictions '*p* and not-*p*', the logic should not permit the move from a biconditional of the form '*p* iff not-*p*' to a contradiction. So we shall be able to assert, without contradicting ourselves, that (23) is true iff it is not. If, on the other hand, our theory incorporates contradictions, the logic should not lead to triviality; it should not permit the move from '*p* and not-*p*' to an arbitrary sentence of the theory's language. Classical logic permits that move; that is the rule called *principle of explosion* or *ex contradictione quodlibet*. Moreover, as we have seen, there are many variants of the liar. Correspondingly, the language may well allow us to form various paradoxical sentences and not only sentences like (L) and (23). So the non-classical logic should be able to combine, without leading to triviality, with the instances of schema (T′) that concern various paradoxical sentences. After we describe the non-classical logic, we ought to prove that the logic does not lead to contradiction, or at least does not lead to triviality, when combined with all the instances of (T′). Such proofs tend to be difficult.

There is no reason to diverge from standard logic excessively, that is, more than is required in order for the paradox to be tackled without any restriction on (T′), true-in′ or true-out′. Here, the approach that endorses contradictions is at a disadvantage, since our logical tradition is strongly opposed to them. Contradictions have been harshly condemned in both philosophical and mathematical reasoning and explicitly rejected in almost all the history of logic. So endorsing them is a radical move. It is more radical than, e.g., rejecting the law of excluded middle (the schema '*p* or not-*p*').[9] Also, there is no need to throw away rules of classical logic which do not lead to contradiction when they combine with (T′) and its attendant

[9] Aristotle [1957, Book Γ, 1005b22–23] famously calls a version of the law of non-contradiction 'the most certain of all principles'.

rules. The maxim here is to be revisionist only in so far as the task in hand requires us to be.[10]

There are various statements about truth that do not lead to paradox, seem obvious at first sight and are not instances of (T′) or other analogous schemas; for example, there is the principle that if a sentence is true, then its double negation is true too. It is good if our theory of truth includes such principles. Still, one needs to be careful because those principles are frequently akin to various logical laws (which do not involve the concept of truth). So if the non-classical logic that our theory incorporates and presupposes does not include the law of excluded middle, then it will be strange for the theory to include the principle that, for every sentence, either it or its negation is true. If, again, the non-classical logic includes the law of non-contradiction (the schema 'not-[p and not-p]'), then it will be strange if the theory does not include the principle that for no sentence is it the case that both it and its negation are true.

1.4. Tarski and Kripke

Tarski's work in the 1930s and Kripke's work in the 1970s have been the two most influential formal approaches to truth. Tarski showed us a way in which we can talk about truth in a formal setting without succumbing to paradox. We can construct a hierarchy of symbolic languages $L_0, L_1, L_2, \ldots$. In L_0 there is no predicate for truth. For every n, L_{n+1} is an extension of L_n; the sentences of L_n are also sentences of L_{n+1}. (Actually, in a Tarskian hierarchy, L_{n+1} may not contain the sentences of L_n themselves, but translations of them. For simplicity, I here suppose that it contains the sentences of L_n.) For each sentence of L_n, there is in L_{n+1} a singular term referring to that sentence. But L_{n+1} also has a predicate, T_{n+1}, which does not exist in L_n and which we can intuitively interpret as applying to the true sentences of L_n and to nothing else. Indeed, Tarski showed us how we can define T_{n+1} in L_{n+1}, rather than have it there as a primitive predicate—but for our purposes we need not focus on that aspect of the hierarchy. Thus L_{n+1} is a metalanguage for L_n. There is no other predicate for truth in the hierarchy. Each one of the languages $L_1, L_2, \ldots$ is accompanied by some axioms that allow us to prove theorems formulated in it. Particularly, in each L_{n+1}, we can prove all instances of the T-schema for L_{n+1}:

$$T_{n+1}\mathbf{s} \leftrightarrow p.$$

To get an instance of that schema, we must replace 'p' with a sentence of L_n and '$\mathbf{s}$' with a singular term of L_{n+1} referring to that sentence.[11]

[10] It is possible to endorse (T′) but not true-in′ or true-out′. But then, one should reject either *modus ponens* or the rule that permits inferring from a biconditional to one of its constituent conditionals. Both alternatives seem to be excessive deviations from classical logic.

[11] Tarski [1983] works with only two linguistic levels, the object-language and the metalanguage. So a Tarskian hierarchy is an extension of his teachings.

For any sentence in a language L_n in the hierarchy, we can say, within the hierarchy, something to the effect that that sentence is true: we can go to L_{n+1} and use T_{n+1}. But no sentence in the hierarchy leads to paradox. In particular, no sentence denies its own truth in a paradoxical manner. First of all, it may be that no sentence in the hierarchy contains a singular term which refers to that same sentence; whether this is so depends on what exactly are the languages $L_1, L_2, \ldots$. We have, however, seen that it is rather easy to achieve self-reference in a formal language. So there may be somewhere in the hierarchy a sentence $\ulcorner \neg T_m \mathbf{s} \urcorner$ where $\mathbf{s}$ refers to this same sentence. But, even in that case, there will be no problem. Since the sentence contains T_m, it is not part of any language before L_m. So no biconditional about it is an instance of the T-schemas for languages up to, and including, L_m, for those T-schemas concern sentences of languages before L_m. And if the T-schema for any language L_n concerns our sentence, among others, then the sentence is part of L_{n-1}, so $n > m$. An instance of that schema for our sentence will be

$$T_n \mathbf{s} \leftrightarrow \neg T_m \mathbf{s}$$

(or perhaps it will have a singular term other than $\mathbf{s}$ on the left-hand side). As there is a different truth-predicate on each side, such a biconditional leads to no contradiction. And, irrespective of what we can prove from the axioms accompanying the languages of the hierarchy, no contradiction about $\ulcorner \neg T_m \mathbf{s} \urcorner$ can be derived by intuitive reasoning: since the sentence is not one of L_{m-1}, T_m does not apply to it. So what it says about itself is not paradoxical, but just right: it is not a true sentence of L_{m-1}.

The Tarskian hierarchy shows how we can talk about truth in a formal context without getting entangled in paradox. But it constitutes no suggestion about what goes wrong when we reason in a natural language and do get entangled. It is a therapy, not a diagnosis. Tarski [1983, pp. 164–165] considered that natural language is universal (it allows us to talk about any topic whatsoever) and for this reason inconsistent. It is sometimes objected to that view that a language is not the kind of thing that can be inconsistent. A theory or an assertion may be inconsistent, but not a language; see, e.g., [Burge, 1984, pp. 83–84]. One way to make Tarski's view more specific and less open to the objection is to say that the word 'true' has an incoherent meaning in natural language in that the unrestricted schema (T) is part of the meaning of 'true' and engenders contradictions. But again, it is not clear what it means to say that a schema, like (T), is part of the meaning of a word. We can clarify the claim by putting it as follows: willingness to reason in accordance with (T) is what understanding the word 'true' consists in, or at least it is part of what understanding 'true' consists in; the word has an incoherent meaning in that mastery of the word involves readiness to think in a way that, in some cases, is bound to lead to contradictions; so, to avoid contradictions, we must somewhat change the meaning of the word by modifying the reasoning that constitutes what it is to understand it.

It is not clear that that diagnosis is correct. For one thing, as we have seen, we can derive contradictions by relying not on (T), but on (F) or (S). So the diagnosis should be extended to semantic expressions other than 'true'. More importantly, it can be argued, as we shall see in the next chapter, that what lies at the root of the paradox is our adherence to classical logic.

Thanks to [Tarski, 1983, Section 5] we can also show, relying on Gödel's method of encoding sentences with numbers, that the language of arithmetic does not contain its own truth-predicate. To be precise, the (formal) language, L_A, of first-order arithmetic possesses no predicate whose extension in the standard model of arithmetic is the set of numbers that encode sentences of the language which are true in the model. For if there is such a predicate in the language, then using Gödel's methods we can formulate in L_A a sentence which effectively says that its numerical code does not belong to the extension of the predicate. If the code belongs to the extension of the predicate in the standard model, then, by our hypothesis about which is the extension, the sentence turns out true in the model, so, by what it says, the code does not belong to the extension; and if the code does not belong to the extension, then, by what the sentence says, it turns out true in the model, so, by our hypothesis about the extension, the code belongs to the extension. In other words, the assumption that there is such a predicate in the language leads to contradiction.

This result can be extended to formal languages that have greater expressive resources. In particular, let us enrich L_A with a monadic predicate 'T' and consider a standard first-order model M which does not differ from the standard model of arithmetic except that it also makes an assignment to the new predicate. Then, the extension of 'T' in M cannot be the set of numbers that encode sentences of the enriched language which are true in M.

Also, an analogue of the result concerns our language *Lang*, which possesses no arithmetical expressions but has the predicate 'T' and the syntactic devices that allow sentence (23) to be formed. If *Lang* has a standard first-order model which is *normal* in the sense that the domain of quantification contains the expressions of *Lang* while the syntactic devices receive the intended interpretations, then the extension of 'T' in the model will not be the set of sentences of *Lang* which are true in the model. For if that is the extension, then (23) will be true in the model iff it is not. Indeed, the reasoning in this and the two preceding paragraphs does not presuppose any principles involving the usual concept of truth. Such principles may be considered open to doubt. Instead of the usual concept, the reasoning makes use of the relativized notion of truth-in-a-model, which is defined with precision in logic.

Now, Kripke [1984] showed that if we enrich L_A with a monadic predicate 'T', we can construct a model M which does not differ from the standard model of arithmetic except that it also makes an assignment to the new predicate and in which the extension of 'T' *is* the set of numbers that encode sentences of the language $L_A +$ 'T' that are true in M. So we can say that $L_A +$ 'T', when interpreted

by means of M, contains a truth-predicate appropriate for its own sentences. M is not a standard first-order model, though, because what it assigns to 'T' is not a single subset of the domain, but two disjoint subsets, the extension and the antiextension of the predicate.

An atomic sentence $\ulcorner T\mathbf{s}\urcorner$ is true in M if the object (number) that the singular term $\mathbf{s}$ refers to belongs to the extension of 'T'; it is false in M if the object belongs to the antiextension of 'T'; and it is neither if the object belongs to neither the extension nor the antiextension. The truth-values of compound statements are calculated in accordance with Kleene's strong three-valued tables [Kleene, 2009, p. 334], but in variants of M we follow certain other schemes for handling sentences that are neither true nor false. In any case, a negative sentence $\ulcorner\neg\mathbf{A}\urcorner$ is true if $\mathbf{A}$ is false, false if $\mathbf{A}$ is true, and neither if $\mathbf{A}$ is neither. The antiextension of 'T' in M contains, on the one hand, the numbers that encode sentences of $L_A + $'$T$' which are false in M and, on the other, the numbers that do not encode sentences of that language. Thus the elements of the domain that belong to neither the extension nor the antiextension of 'T' are the numbers that encode sentences of the language which are neither true nor false in M.

Again, using Kripke's methods, we can construct a model M' for *Lang* which is normal in the sense explained above and in which the extension of 'T' is the set of sentences of *Lang* that are true in M'. Like M, M' will differ from a standard first-order model because it assigns 'T' both an extension and an antiextension. The antiextension of 'T' will contain, on the one hand, the sentences of *Lang* that are false in M' and, on the other, the elements of the domain which are not sentences of *Lang*.

If we interpret the language $L_A + $'$T$' by means of the rules that Kripke discusses for calculating truth-values, we can still formulate a liar-like sentence and other paradoxical sentences in it. The liar-like sentence now effectively says that its numerical code belongs to the antiextension of 'T'. The presence of such sentences leads to no contradiction. The paradoxical sentences are provably neither true nor false in M or in any of its variant models. Similarly, (23) is neither true nor false in M'. If we interpret *Lang* by means of Kripke's rules, (23) effectively says about itself that it belongs to the antiextension of 'T'.

If we interpret $L_A + $'$T$' or *Lang* like Kripke, then we cannot formulate in the former language a sentence which effectively says, of its own numerical code or any other number, that it does not belong to the extension of 'T', nor can we formulate in the latter language a sentence which effectively says, of itself or any other sentence, that it does not belong to the extension of 'T'. That is a consequence of the way in which 'T' is interpreted, by means of an extension and an antiextension, and the way in which negation is treated. For any singular term $\mathbf{s}$, $\ulcorner T\mathbf{s}\urcorner$ tells us that the denotation of $\mathbf{s}$ belongs to the extension of 'T', since that is the condition whose satisfaction by a model is necessary and sufficient in order for $\ulcorner T\mathbf{s}\urcorner$ to be true in that model. On the other hand, $\ulcorner\neg T\mathbf{s}\urcorner$ tells us that the denotation of $\mathbf{s}$ belongs to the antiextension of 'T', since that is the necessary and sufficient

condition in order for $\ulcorner\neg T\mathbf{s}\urcorner$ to be true in a model. And no sentence in $L_A + \text{`}T\text{'}$ or *Lang* tells us that the denotation of **s** does not belong to the extension of 'T'.

Thus, if we want to say that a given sentence of $L_A + \text{`}T\text{'}$ is true in $M$, we can make our point in $L_A + \text{`}T\text{'}$ provided we interpret that language by means of M; $\ulcorner T\mathbf{s}\urcorner$, where **s** is a term for the numerical code of that sentence, will make the point. Likewise, if we want to say that a given sentence of $L_A + \text{`}T\text{'}$ is false in $M$, then $\ulcorner\neg T\mathbf{s}\urcorner$ will make the point provided we interpret the language by means of $M$ and treat negation like Kripke. But if we want to say that a given sentence of $L_A + \text{`}T\text{'}$ is not true in M, we cannot make our point in $L_A + \text{`}T\text{'}$ interpreted through $M$. This fact is an expressive restriction that Kripke [1984, pp. 79–80] admits. It does not falsify the claim that the language $L_A + \text{`}T\text{'}$, when interpreted by means of M, contains a truth-predicate appropriate for its own sentences, but it blunts the force of that claim. Things are similar with respect to *Lang* and M'.[12]

1.5. The project

The aim of this book is to formulate a formal theory of truth that avoids contradictions. The theory is meant to sanction schemas and inference-rules about truth which appear platitudinous. It is not intended to provide a philosophical account of the concept or the property of truth. Claims about the concept tell us whether and how the notion of truth can be analysed, under what conditions one possesses it, whether it has any characteristic role, and so forth. They may take on a semantic form and attempt to elucidate what the word 'true' means. Claims about the property of truth tell us whether it has an underlying nature, whether it is a relational property, and so on. If to be true is to correspond to a fact, then truth is a relational property, and so is it if being true consists in cohering with other truth-bearers. It seems to me that when apparent platitudes about truth, in combination with classical logic, lead to contradictions and triviality, avoiding such

[12] In M, the numerical code of a sentence **A** of $L_A + \text{`}T\text{'}$ belongs to the antiextension of 'T' just in case the numerical code of $\ulcorner\neg\mathbf{A}\urcorner$ belongs to the extension. Thus we can vary M and interpret the language by means of a model, N, in which 'T' has the same extension as in M but no antiextension. Although N does not differ from standard first-order models in assigning two sets to 'T', the treatment of atomic sentences $\ulcorner T\mathbf{s}\urcorner$ is still non-standard: $\ulcorner T\mathbf{s}\urcorner$ counts as true in N iff the number denoted by **s** belongs to the extension of 'T', but it counts as false in N iff the number either codes for no sentence of $L_A + \text{`}T\text{'}$ or codes for a sentence **A** of the language such that the number coding for $\ulcorner\neg\mathbf{A}\urcorner$ belongs to the extension of 'T'. Compound statements are treated as in the case of M. Like before, the extension of 'T' is the set of numbers that encode sentences of $L_A + \text{`}T\text{'}$ which are true in $N$, but if we want to say that a given sentence of the language is not true in $N$, we cannot make our point in $L_A + \text{`}T\text{'}$ interpreted by means of N. If we know that **s** denotes a number coding for a sentence **A** of $L_A + \text{`}T\text{'}$, then $\ulcorner\neg T\mathbf{s}\urcorner$ tells us that the code of $\ulcorner\neg\mathbf{A}\urcorner$ belongs to the extension of 'T', so it indirectly says that $\ulcorner\neg\mathbf{A}\urcorner$ is true in N, not that **A** is not true in the model. We can similarly vary M'.

unwelcome consequences is a task more urgent than developing a philosophical account of the concept or the property of truth.[13]

Nevertheless, readers may be interested to know that the account I prefer is minimalist. It seems to me that one possesses the concept of truth just in case one appreciates that an assertion, sentence, belief, proposition, etc. is true iff things are as it says they are. The main way to appreciate that is a willingness to reason from a claim to the conclusion that the claim is true and from the premiss that a certain claim is true to the claim itself. Such reasoning accords with rules like true-in and true-out. Assenting to biconditionals such as 'The proposition that whales are mammals is true iff whales are mammals' or 'The sentence "Whales are mammals" is true iff whales are mammals' is another way to manifest possession of the notion of truth. Appreciation is a matter of degree, and possession of a concept is not an all-or-nothing affair. I recognize that the notion of truth plays an important normative role in assertion and belief, but I would argue that that is not an essential feature of the notion. As regards the property, I would agree with Horwich [1998, pp. 1–5] that it has no underlying nature, so there is no essence of truth awaiting discovery. Those views, however, will not be defended in this book.

The theory that sanctions the platitudes about truth will tackle the paradoxes by diverging from classical logic. It will diverge not only by omitting some classical principles, but also by contradicting others: it will include biconditionals of the form '*p* iff not-*p*', whereas 'not-(*p* iff not-*p*)' is a classical tautology. Once we are ready to deviate from standard logic, we have no motivation to distinguish between the object-language and the metalanguage. So there will be no such distinction; the theory will come in more than one version, and the versions will differ in their languages, but each one will treat its truth-predicate as appropriate for evaluating the sentences of its own language. The main platitudes about truth are schemas like (T)—although in fact (T) is subject to some restrictions which we shall see in Chapter 3 and which are not relevant to the languages we shall discuss. Each version of the theory will contain biconditionals like the instances of (T) for all the sentences of its language. The theory will also vindicate other platitudes, such as the principle that if a sentence is true, then its double negation is true too and the principle that for no sentence is it the case that both it and its negation are true.

The project is descriptive and not normative. The theory's truth-predicate, which will be '*T*' in all the versions, has the usual, ordinary sense of the word 'true', and the theory is meant to contain statements that do hold and not merely some that it is permissible to endorse if one wants one's position to be coherent. It is not

[13]For the distinction between properties and concepts see, e.g., [Putnam, 1991, pp. 197–198 and 1979, pp. 305–307]. Properties and concepts are individuated differently. Since water is the substance H_2O, it is natural to consider that the property of containing water is the same as the property of containing H_2O. Still, the concept of containing water is other than the concept of containing H_2O.

meant as a suggestion about how we should modify or improve the usual sense of 'true'.[14]

In any case, why should we tackle the paradoxes by diverging from classical logic? Why are we wrong in accepting some apparently obvious logical principles and not in accepting certain seemingly platitudinous principles about truth?

[14] '*T*' will be predicated of sentences, and, as mentioned in fn. 5 above, one may claim that, in its ordinary sense, 'true' does not apply to sentences [Strawson, 1993, pp. 60–65 and Baker and Hacker, 1984, pp. 180–190]. Outside philosophy and logic we do not normally call sentences 'true', but we frequently describe someone's words as true, and then what is so described is a token of a sentence. If 'true', in its usual sense, applies to sentence-tokens, then it also applies to sentences (sentence-types). It does not apply to all sentences, though; as we will see in Chapter 3, the one-place predicate 'true' is not appropriate for evaluating ambiguous or context-dependent sentences, and the word is ordinarily one-place. The reason we do not normally call sentences 'true' outside technical discussions must be, not that the word is inapplicable to sentences, but that a great many sentences of natural languages are ambiguous or context-dependent.