TARSKI, FREGE AND THE LIAR PARADOX

Aaron Sloman, School of Computer Science,

The University of Birmingham, UK

(Originally in *Philosophy,* Vol XLVI, pages 133-147, 1971)

A. Introduction

A.1. Some philosophers, including Tarski and Russell, have concluded from a study of various versions of the Liar Paradox 'that there must be a hierarchy of languages, and that the words "true" and "false", as applied to statements in any given language, are themselves words belonging to a language of higher order'.[1] In his famous essay on truth[2] Tarski claimed that 'colloquial' language is inconsistent as a result of its property of 'universality': that is, whatever can be said at all can in principle be said in it, with an extended vocabulary if necessary. Thus, in English we can talk about English expressions, what they denote, what they say, whether what they say is true or false, and so on: English contains its own metalanguage. This universality enables us to construct sentences which say of themselves that they are false, and by applying the law of excluded middle to them we easily derive a contradiction. Tarski concludes that 'these antinomies seem to provide a proof that every language which is universal in the above sense, and for which the normal laws of logic hold, must be inconsistent' (op. cit, pp. 164-5). He then proposes to avoid such contradictions by the use of a hierarchy of languages such that statements about any one language can be made only in a different language at a higher level. This is, of course, one version of the more general programme of setting up artificial languages (or hierarchies of languages) whose syntactic formation-rules exclude as ill-formed or ungrammatical certain strings of symbols which are essential for the derivation of the logical and set-theoretical contradictions. A different programme, more congenial to philosophers who are not logicians, is to try to show that the 'deep' grammatical rules of a natural language themselves rule out the contradiction-generating phrases and sentences as ungrammatical and therefore nonsensical.

A.2. In this article I shall explore a third programme, extending some of Frege's ideas about sense and reference, so as to deal with the paradoxes at a semantic rather than a syntactic level. I shall argue that the peculiarity of paradoxical sentences is not that they are grammatically or syntactically ill-formed, nor that they are senseless, but rather that they lack what Frege would call a reference, namely a truth-value, without lacking a sense.[4] Since they lack a truth-value they and their negations cannot be used in the normal way to derive contradictions.

A.3. It is worth noting in passing that there is something paradoxical about Tarski's claim that colloquial languages are inconsistent. For if the principles of our language generate any statement of the form 'p &-p', then, according to a familiar logical metatheorem, the inference from any statement in the language to any other statement is valid for that language, and the language must be unusable. Yet we do not regard as valid the inference from 'The moon is not made of green cheese' to 'The sun will explode in 1975'. Why not? Is it that we are just too stupid or ignorant to see that the inference is valid? Why do we remain unconvinced and continue to use colloquial languages (as even Tarski does in his meta-meta-linguistic remarks) even after we have seen the logical demonstration (using paradox sentences) that such inferences are valid and our language inconsistent? Is it mere obstinacy? Nostalgia? Is there perhaps some disaster awaiting those who continue to use ordinary English in place of a hierarchy of formalized languages? Or should we perhaps conclude that since our language is inconsistent yet usable, inconsistency is not a serious defect? Or does the fact that our language is usable demonstrate that it is not inconsistent after all and that there must be a mistake in Tarski's alleged proof that it is?

I shall try to show that Tarski's argument is mistaken in its assumption that if *s* is a well-formed sentence of a colloquial language then *s* must have a truth-value and the disjunction of *s* and its negation must be true. I believe there is a still deeper mistake in attempting to apply such meta-mathematical concepts as 'consistency', 'rule of inference', 'axiom', to languages which can be used to perform the functions of natural languages. This mistake is discussed in the final section, in which an attempt is made to show how a modified concept of 'consistency' might be applicable to natural languages.

B. The solution

B.1. What are we to make of the problem sentences? My answer is that they merely illustrate the fact that the rules of a natural language are capable of generating meaningful expressions which necessarily fail to perform their intended function, or in Frege's terminology, expressions with a sense but no reference. I shall now attempt to explain what this means and how it differs from the claim that the rules are inconsistent.

B.2. Linguistic expressions and other devices can be classified into different categories according to the sorts of functions they are standardly intended to perform: let us call these *pragmatic*

categories. For instance, there are *referring devices* (such as proper names, demonstratives, definite descriptions, the present tense), whose standard function is to refer to or pick out some one individual (or group of individuals). There are *descriptive devices* whose standard function is to classify or describe individuals, or, if we include relational expressions, to classify or describe pairs of individuals, triples of individuals, etc. There are *assertive sentences* whose standard function is to convey information, i.e. something which is true or false. I shall not discuss other pragmatic categories, and what I have said about these is very much oversimplified for the sake of brevity.

B.3. Many expressions are composed of smaller units, some of which may be expressions in the same pragmatic category, others in different categories. The rules of the language specify *how* words, or other expressions, may be combined in order to generate more complex expressions, and also how the functions of complex expressions are determined by the functions, and manner of composition, of their components. Syntactic rules recursively generate various classes of expressions; semantic rules recursively specify how to identify the individuals, properties, states of affairs, or whatever (if there are any), correlated with the expressions; and pragmatic rules specify the standard kinds of linguistic functions performed by such expressions. (This is grossly oversimplified, for brevity.)

B.4. There are various ways in which an expression constructed in accordance with syntactic rules which normally generate expressions in a pragmatic category associated with a standard type of function may nevertheless fail to perform that function. A referring expression may fail to identify any individual. A descriptive expression may express a classification procedure which breaks down. An assertive sentence may fail to convey anything true or false. I call such expressions *unsuccessful.* The important point, for present purposes, is that there are different ways in which expressions can be unsuccessful, and that in order to understand these differences we must resist the temptation to lump them all together under some blanket heading such as 'nonsensical' or 'grammatically ill-formed'.

B.5. Firstly, the failure of an expression to do what it is supposed to do may be a result of some contingent fact: the world may be at fault rather than the expression itself. Thus, as is well known, definite descriptions with a perfectly clear sense such as 'the only nineteen-year-old poet earning a million pounds a year' may fail to refer uniquely because there simply happens to be no individual satisfying its description, or because there are several. A descriptive expression (e.g. 'has a temperature of 600° Centigrade') whose use presupposes that a number of different measuring devices happen to give similar results or assumes a theory may fail to classify objects for which those devices happen not to agree or whose behaviour refutes the theory. A sentence formed in accordance with rules which render it *capable* of saying something true or false may *happen* not to have a truth-value because of the contingent fact that a referring expression in it fails to refer, or because what it refers to happens to be a borderline case for some concept or an instance for which the concept breaks down. If the world had been different these expressions, with the sense they already have, would have functioned successfully.

B.6. Secondly, and more importantly for present purposes, there are cases where the failure of an expression to perform its standard function is non-contingent: the failure depends not on the world, but on that expression and its verbal context. To discuss these cases we require the general idea of a set of rules or operations which *normally* or *standardly* generate a certain type of result but which have 'unintended' by-products which are different, and may be described as *degenerate.* For instance, a large variety of arithmetical functions can be constructed using repeated multiplication, addition and subtraction, and normally the value of such a function depends on which number or set of numbers is taken as argument, but in some degenerate cases, such as the function

$$(x+1).(x-1)-x^2,$$

the same value is correlated with every argument, and therefore *calculation* of the value for the argument 375, for instance, though possible is unnecessary.

Another illustration is the algorithm for dividing an integer **n** into another integer **m** which normally generates a number **k** and a remainder **r:** the procedure is to multiply **n** first by 1, then by 2, then by 3, etc., until a result is reached which is greater than **m**. The second-last multiplier is then **k**, and the remainder **r** is **m-k.n**. A computer can be programmed to carry out instructions of the form 'divide **n** into **m**' in this way eventually printing out the numbers **k** and **r** and it will respond to the instruction no matter what numbers are mentioned in it. In particular, if it is instructed 'divide 0 into 66' it will set to work multiplying 0 first by 1, then by 2, and so on. But it will eventually have to be switched off, for otherwise it will go on for ever looking for a number **k** such that **0.(k + 1)** exceeds 66, or perhaps run out of tape or storage space. The instruction to the machine is perfectly well-formed: the machine knows what to do with it. It applies exactly the same procedure which, in 'normal' cases, generates a result, but in *this* instance the procedure is non-terminating. (A computer with a higher degree of sophistication might be able to discover the peculiarity, for instance, if it is programmed to look for various sorts of short-cuts for doing its jobs.)

Here we have a well-formed and meaningful but necessarily unsuccessful computer instruction (taking success to involve production of an answer).

B.7. Similarly it is clear that the following are degenerate referring expressions since they are quite

incapable of successfully referring to any individual if the words are taken in their normal senses: 'the man who came in after the last person to enter', 'the first wife of the oldest man ever to die a bachelor', 'the father of the person referred to by this phrase', 'the largest star never to have been referred to before ten years hence', 'the barber who shaves all and only those who do not shave themselves', 'the man referred to by the subject of the sentence: "The woman referred to by the phrase containing this sentence, is deaf" ', 'the largest thing referred to on this page'.

The reasons why these expressions necessarily fail to refer are not all the same: in some the identifying description is inconsistent (e.g. the barber), in some there is, in effect, a non-terminating series of 'nested' functions ('the father of the father of the father of. . .'), in some these two faults are combined (e.g. the phrase containing a sentence). It would be of interest to compile a systematic classification of ways of non-contingently failing to refer, but this will not be attempted here.

B.8. As far as English grammar is concerned, the above illustrations all seem to be grammatically well-formed: the syntactic rules which generate successful referring expressions generate these illustrations as by-products. Further, I think we can say that in Frege's terminology (or an obvious extension thereof) they have a *sense* (Sinn) even though they necessarily lack a reference (Bedeutung): they are not meaningless or nonsensical phrases. Just as successful complex referring expressions express a procedure for identifying an individual (i.e. express a sense) so do these, and in both cases the procedure is determined by the meanings of the component expressions and their manner of composition: the difference is that in the degenerate cases the procedures have an inner structure which makes them incapable of producing a result (like the procedure for dividing 66 by 0, mentioned above), for instance because they are non-terminating or because one part of the procedure undermines the working of another part.

It is possible that examination of various cases will show that some breakdowns of procedures are more radical than others: for instance one could say that procedures whose applications necessarily grind to a premature halt are closer to the case of pure nonsense than the procedures whose applications are non-terminating. In a way, it is simply a terminological question whether we should say that the necessarily unsuccessful expressions have a sense, or whether we should coin a new terminology and talk about more or less complete[5] senses, or potential senses, or pseudo-senses. This does not make much difference for my present purpose, which is to show how unsuccessful *sentences* can be regarded as analogous to unsuccessful *referring expressions* (but I am reluctant to agree that *all* cases of necessarily unsuccessful reference are cases of a lack of sense, for it seems to me that the phrase 'the largest prime number between 24 and 28' has as much sense, in any useful sense of the word 'sense', as the phrase 'the largest prime number between 14 and 38'). Failure of reference does not constitute a paradox, or contradiction. Can something similar be said about sentences?

B.9. Frege's thesis that sentences denote truth-values is acknowledged to have provided a foundation for modern quantificational logic; yet his terminology has made it difficult for philosophers to take his theory seriously. It is too easy to object that to describe a *sentence* as naming or denoting anything is a misuse of language, or that what one *refers to* is what one is talking about, whereas we are not talking about truth-values whenever we make true or false statements. Terminological issues can obscure the power of Frege's claim that complex referring expressions and complex sentences alike can be (in many cases) analysed into function-signs and argument-signs, and that the relation between a referring expression and what it denotes has much in common with the relation between a sentence and its truth-value: things denoted, and truth-values, are alike arguments and values of functions. If Frege had invented a new, more general, term for what both relations have in common, instead of simply extending the terminology associated with reference,[4] he might have been more easily understood.

Further, if he had noticed more analogies than he did notice, his theory might have been more compelling: however, I believe his vision was blinkered by his over-riding concern with the foundations of arithmetic, which prevented his generalizing the concept of a 'function' even further than he did, along the following lines. The sense of a referring expression is or determines a procedure for identifying the individual referred to; and similarly the rules for constructing an assertive sentence generate a sense which is or determines a procedure for identifying the truth-value of what is asserted. (For brevity, cases where sense — and reference — varies with context are ignored here.) The outcome of either procedure, i.e. which individual or truth-value is identified, depends *in general* not merely on the sense, but also on how things happen to be in the world, for things might have been different from the way they are (have been, will be) so that, for instance, 'Britain's prime minister on 1st January 1970' might have identified a different person, and 'Rain fell in London on 1st January 1970' might have had a different truth-value. (This is where Frege would have had to generalize his concept of a 'function', to include contingent determination of values.)

Calling the way things (past, present and future) *are* 'the actual world', and noting that this is just one among many other possible worlds (i.e. which might have been actual), we can summarize so far by saying that the sense of a referring expression normally determines a many-one correlation from possible worlds to individuals and the sense of an assertive sentence normally determines a many-one correlation from possible worlds to truth-values. In Frege's terminology, the 'reference' in either case is the individual or truth-value correlated with the actual world. Next, I claim that there

are at least three kinds of 'degenerate' referring expression or sentence:

1. Those whose sense turns out to correlate the same individual or truth-value with all possible worlds, e.g. 'the age at death of the tallest person ever to die in his twentieth year' and 'All black things are black'.
2. Those whose senses fail to identify any individual or truth-value corresponding to the actual world, but would have succeeded had the world been different: this is *contingent* failure of 'reference'.
3. Those whose senses fail to identify any individual or truth-value in any possible world: this is the case of *necessary* failure.

If X is an individual or truth-value identified in some possible world by the sense of E (a referring expression or sentence), then I call the possible worlds in which X is identified the **X-conditions** for E. If E is a sentence, there are *truth-conditions* and *falsity conditions.* Thus, in case (1) the sentence E has only truth-conditions, or only falsity conditions: it is necessarily true or necessarily false. In case (2) the actual world is not one of the truth-conditions nor one of the falsity-conditions. In case (3) the sense of E determines no truth-conditions and no falsity-conditions. In effect, the sense of a sentence is a procedure which selects truth-conditions and falsity-conditions from the set of all possible worlds. In case (2) the *actual* world is not selected, while in case (3) *no* possible world is selected. I have discussed case (1) elsewhere[6] — cases (2) and (3) must now be examined more closely.

B.10. We have examined case (2), contingent failure of reference, for referring expressions. Similarly some sentences contingently fail to identity a truth-value. For, on account of such things as

i. our incomplete knowledge of what particular things exist in the world,
ii. our inability to make indefinitely fine discrimination, and
iii. our inability to anticipate all possible experimental results and future situations in which our present concepts could generate conflicts of criteria or borderline cases,

it is impossible for us to assign meanings to words in such a way as to *guarantee* that every well-formed sentence will, in every possible state of affairs, be either definitely true or definitely false. Not even formal logicians can construct a language like this with all the uses of natural languages. Instead, we tolerate indefiniteness, doubtful empirical presuppositions, etc., where they make little or no practical difference (how many drops per square yard per second are required for it to be raining?), and when serious borderline cases, conflicts of criteria, failure of presuppositions, etc., turn up we make *ad hoc* adjustments to the language (including what are called conceptual revolutions).

This difficulty might seem to be avoidable by simply specifying a set of truth-conditions for each sentence, leaving all remaining possible worlds to form its falsity-conditions. But since we cannot name each possible state of affairs (their variety being too rich) a criterion or procedure is needed for assigning every possible state of affairs to the truth-set or its complement: so the problem of borderline, or otherwise undecidable, cases arises again. We know from the history of science how difficult it is to *ensure* that our concepts are applicable come what may. In any case, there are obviously reasons why we need to be able to generate referring expressions whose success is a contingent matter, and, at least on Frege's analysis, sentences in which they are used as argument-signs must also be capable of being unsuccessful. Clearly sentences which *contingently* lack a truth-value cannot be condemned as senseless: for the fault lies in the (actual) world.

B.11. It is more difficult to admit that the rules of a language can also generate some sentences of type (3), having a sense but incapable of having a truth-value in *any* possible circumstances: the meanings of the words, the constructions used, and perhaps the context, generate a procedure for selecting a set of truth-conditions and a set of falsity-conditions, but both sets turn out to be empty, or to have only borderline or indeterminate instances. For example, the procedure for identifying the truth-value may fail to produce a definite outcome in *any* possible world because it never terminates (cf. B.6, above), or because the sentence contains a necessarily unsuccessful referring expression. On a Fregean theory, the following sentences, with their normal sense, are all incapable of having truth-values: 'This table is twice as long as I hereby say it is', 'The statement which says of itself that it is true, is false', 'The father of the man referred to by the subject expression of this sentence is dead', 'What this sentence states is true', 'The man referred to in the second half of this sentence is not the man referred to in the first half of this sentence'.

If we give up the blanket condemnation of such degenerate and apparently paradoxical sentences as *nonsensical,* we can open our minds to the exploration of the various *different* sorts of reasons why a procedure for correlating possible worlds with truth-values may be completely unsuccessful, even though the procedure is composed of steps, or generated by rules, which in other contexts work successfully.

For instance, the above sentences are clearly different from syntactically ill-formed strings like 'This table is between the door', 'It false that', 'This table is or but', etc. (Somewhere between or beyond ill-formed sentences and the previous examples seem to come so-called category-mistakes, such as 'Tuesday is between the door and the wall', 'Badly serviced machines sometimes polish prime numbers'—or perhaps even 'Colourless green ideas sleep furiously'. I shall not discuss these, nor analogous referring expressions.)

B.12. Let us look more closely at two examples of degenerate sentences of type (3) to see what has gone wrong,

> **(s)** What the sentence (s) says is true.

> **(t)** What the sentence (t) says is not true.

In general, there is nothing wrong with identifying something via a sentence which refers to or expresses it, provided the second sentence does so successfully. When the second sentence is, or refers back to, the first, there is a danger that the above proviso generates an infinite regress.

Something like this happens in the above two cases: both implicitly involve a sequence of nested functions which fail to terminate in any argument, so there cannot be a value for the outermost function. The second sentence has the additional failing of attempting at least implicitly to identify a proposition as being both the argument and the value of a function (negation), where the function in question is so defined as to have a value distinct from its argument. (Compare: (u) The father of the person referred to by (u).) Thus, by examining their semantic properties we can conclude that neither sentence can correspond to a truth-value, even though we can also derive a contradiction from the assumption that (t) has a truth-value. But without this assumption there is no contradiction. Thus there is no need to try to avoid a contradiction by postulating hierarchies of languages, or to search for 'deep' grammatical rules in natural languages, according to which such expressions are ill-formed.

If there is any real temptation for people to use such unusable sentences, it would be helpful if simple mechanical tests could be found for identifying them. But there is no reason to believe that there must exist a *universally* applicable decision-procedure; and this does not matter, as long as in *each* tempting case there is a way of telling whether an expression is degenerate or not. Why should not such an examination have to involve semantic as well as synactic considerations in some cases? Thus, it may be necessary to find out what is referred to by some part of a degenerate sentence, and if there are indexical words like 'this', then what is referred to may be different in different uses, so that a syntactically well-formed sentence may be degenerate in one use and successful in another. For instance, I assert:

> **(v)** The sense of sentence (t) determines no truth-value.

This implies something like:

> **(w)** What the sentence (t) says is not true.

But this apparently commits me to asserting (t), for there is no difference between (w) and (t). (This was pointed out by an anonymous referee of an earlier version of this paper.) There is, however, no real contradiction here, for despite the fact that the same words are involved, in my use of the words (w) I am commenting on a different sort of use of the sentence, so that what I say has a different *sense* from the original. To see this, notice that if it is now claimed that '(t)' was intended to name a use of the sentence of the sort I was making in (w), so that (t) is intended to have the same sense as (w), then my comments are withdrawn and I am no longer committed to (w)!

Compare this with the following: A makes a statement, whereupon B says to A 'What you have just said is false', whereupon C says to B 'What you have just said is false'. Clearly C is not making the same assertion as B, nor agreeing with him despite the use of exactly the same words. To deal fully with this, and other examples like 'This sentence does not express a sense', 'This sentence does not express a sense with a truth-value', or (adapting an example given by Quine in *The Ways of Paradox,* p. 9)

> '"Yields a sentence with no truth-value, when applied to itself" yields a sentence with no truth-value, when applied to itself',

would require a more detailed discussion of the way other factors besides the rules of the language used determine the sense of a referring expression or sentence. (I hope to complete a paper on this soon.)

B.13. I have tried in this section to show that parallels can be drawn between different kinds of failure of referring expressions and sentences. This seems to be an important part of the justification for Frege's theory that sentences *denote* truth-values. Further, it allows at least some paradoxical sentences to be disposed of as harmless without being labelled 'nonsense' or 'syntactically ill-formed'. Objections must now be considered.

C. Some objections and replies

C.1. The first main objection is that I have defined having a sense in terms of expressing a procedure and my notion of a procedure is still too vague: I have given no criteria for identity of procedures, so it is left unclear how to tell whether two expressions have the same sense or a different sense. Worse, I have talked about procedures of a type which *normally* or *standardly* produce an outcome of a certain sort and also 'degenerate' procedures of that type which are constitutionally incapable of producing the outcome, without showing in detail how to tell the difference between a degenerate procedure and something which is no procedure at all: and therefore it is not yet clear how I distinguish nonsense from unsuccessful sense. I accept this criticism, but cannot spell out a detailed answer here. Moreover it is not strictly necessary for my main aim, which is to draw a comparison between unsuccessful reference and failure of truth-value, and since the difficulty of drawing a clear boundary between sense and nonsense or of denying degrees of sense is common to both cases this supports the comparison.

C.2. My examples show that in a 'colloquial' language expressions of an identifiable pragmatic category may be syntactically well-formed and have a sense (or perhaps partial sense) and yet because of their semantic relations be incapable of performing the standard function of expressions in that category. It is possible to reply to this that such examples merely show the inadequacy of the grammar of ordinary language, since grammatical rules should generate only expressions which *can* perform their intended functions. For instance, Frege demanded that a language fit for 'scientific' purposes should not allow the construction of sentences with no truth-values or referring expressions without reference.[7] He therefore demanded that the grammatical rules and semantic rules of an adequate language should ensure that every well-formed name or sentence is successful. Perhaps this demand is reasonable for a language solely concerned with mathematics, but in general it is unreasonable, as already explained. A natural language needs an enormous variety of procedures for identifying individuals, for instance by means of their qualities, their spatio-temporal relations to other individuals, their kinship relations, their legal relations, and even their relations to other acts of reference (as in the phrase 'the man you referred to a minute ago'). This enormously varied means of identification of things referred to, including the use of metalinguistic expressions, gives colloquial languages great flexibility. The price of this flexibility is that the language cannot *guarantee* success of all referring expressions which are well-formed grammatically.

C.3. But now it may be objected that it is one thing to permit a language to generate expressions which are *contingently* unsuccessful, but quite another thing for the grammar to permit *necessarily* unsuccessful expressions. It would be pleasing if grammatical rules could be found which would generate all the expressions required for normal linguistic purposes without generating any which *necessarily* failed to do their job. It is not clear whether this is possible, but in view of the diversity of types of unsuccessful expressions it seems likely that any set of such rules would have to be messy and *ad hoc.* However, there is no reason why, apart from some kind of aesthetic consideration, a language whose grammar generates necessarily unsuccessful referring expressions with a clear sense should be regarded as *defective.* The fact that some expressions are unsuccessful does not prevent the remaining expressions from being usable.

C.4. At this stage it will be objected that even if unsuccessful *referring expressions* can be tolerated, the case is different with *assertive sentences.* There is a strong reluctance to be satisfied with a language whose rules generate problem sentences like 'The proposition hereby expressed is false', 'The proposition which says of itself that it is false, is true', 'The sentence "The sentence containing this sentence is true" is false', which clearly must be both true and false if they are either. We find it hard to agree that although these sentences break no grammatical rules, and express a sense (perhaps a degenerate sense), they necessarily lack a truth-value (e.g. because they express non-terminating procedures for determining a truth-value). Instead, our philosophical habits lead us to say that what a meaningful sentence expresses *must* be capable of being true or false, and since these sentences can be neither, they must be meaningless.

But when we look at such a sentence and find that it is apparently constructed out of meaningful components, we conclude that there must be unknown grammatical rules which it violates, and we thus start searching for the hidden 'logical grammar' or else conclude that, despite appearances, ordinary languages are useless and must be replaced by languages with a more selective grammar. But why not simply accept that an expression which breaks no formation rules can nevertheless fail to perform its standard type of function without thereby interfering with the usefulness of other expressions of the same sort ?

C.5. One obstacle is the belief that although we can at first ignore a problem sentence and its negation, we must regard their *disjunction* as true since this is an instance of the logically valid form 'p or not-p'. Tarski's argument to show that colloquial language is unavoidably inconsistent implicitly relies on the assumption that *every* sentence which is a substitution instance of a 'valid' formula of propositional logic expresses a truth (op. cit., p. 165). And from the truth of the disjunction a contradiction is easily derived. Briefly, the answer to this is that we call such a formula valid if we can show that every possible substitution of true or false propositions for the components yields a *true* proposition.

There is nothing in this to imply that the substitution of a proposition *without* a truth-value also yields a true proposition: rather, it yields a proposition without a truth-value. Propositional connectives, like 'not-' or 'or' (at least as understood in Frege's propositional logic) signify functions from truth-values to truth-values, as is clear from the fact that they are definable by means of truth-tables. But when a function lacks an argument it equally lacks a value. The function 'the square of . . .' determines no value for the argument *the largest prime number between 24 and 28,* since there is no such thing, $(x + 1).(x - 1) - x^2$, which, like (p or not-p), yields the same value for *every* argument, yields *no* value for the above nonexistent argument.

Similarly, when p lacks a truth-value, so does (p or not-p). Neither the truth-table for 'not' nor the truth-table for 'or' specifies, nor needs to specify, a truth-value for a complex proposition one of whose components lacks a truth-value. Of course, this does not mean to say that besides truth and falsity there is some mysterious *third* truth-value corresponding to the problem sentences (and others without a normal truth-value). For, to say that neither Mr. Smith nor Mr. Jones is my uncle is not to say that some third person, possibly called Mr. Neither-Smith-nor-Jones, is my uncle.

This is why, unlike Dummett,[8] I do not think the possibility of such truth-value gaps requires standard truth-tables to be supplemented by additional rows. To sum up: if S is one of our degenerate sentences without a truth-value, then so is (S or not-S): and from this no contradictions follows. This defence of colloquial language was apparently not noticed by Tarski.

C.6. Next, it may be objected that what I have said violates the cherished principle that if p is a proposition expressed by a sentence S, then the sentences 'p is true' and S are synonymous, or at least logically equivalent; for I have in effect implied that far from being equivalent they even have different falsity-conditions, for the former is false when the latter lacks a truth-value. The reply is that in view of the existence of indeterminate propositions there never was any good reason to cherish this principle and that any theory of truth which depends on it or implies it is therefore untenable.

Of course, it is possible to *define* a new metalinguistic operator '. . . is true' as the identity truth-function (and that, in effect, is what Tarski did in connection with formal languages), but that does not explicate what 'true' means. To say exactly what it does mean would make this essay too long,[9] though I believe that its consistent use in natural language is possible and can be analysed satisfactorily, contrary to what Tarski claimed *(op. Cit.,* p. 165).

D. Some remaining problems

D. 1. There are many gaps in what has been said so far. This section is intended to close some of them. I hope it is clear that I do not intend to denigrate Tarski's work on formalized languages. My aim has been only to suggest that he gave up too easily the attempt to rescue *ordinary* languages from the charge of inconsistency, which is not to say that he should have refrained from discussing hierarchies of formalized languages: his mistake was very fruitful for symbolic logic.

D.2. I have talked freely about *the* rules of a language without explaining how we can tell what the rules of a language are. Which syntactic, semantic and pragmatic rules govern the language of any person or social group is a complicated empirical question: one can only put forward more or less tentative hypotheses and test them in the light of actual linguistic behaviour. Hence, the question whether the rules of a natural language (or some individual's language) actually contain some 'deep' provisos which our necessarily unsuccessful expressions violate is a complex empirical question. What I have mainly tried to show is that there is no *apriori* reason why grammatical or other linguistic rules should exclude such expressions.

D.3. Secondly, although I have tried to show that Tarski's attempt to derive a contradiction was invalid on account of his failure to notice the possibility of a truth-value gap, it might seem that since I am so tolerant of unsuccessful expressions of all sorts there is no reason why I should be concerned to avoid contradictions: the reader may now be wondering whether 1 would ever call any language inconsistent, or whether I would say that even the existence of inconsistencies does not matter since they can be noted as unsuccessful degenerate cases, then safely ignored.

Obviously there is no objection to *the formation* rules of a language generating inconsistent sets of statements, such as 'New York is in China', 'New York is not in China'. In general, the language does not assign truth-values to sentences, since that, so to speak, is left for the world to do, though special cases (by-products) do occur which must have the same truth-value no matter what the world is like: analytically true or false statements. Now, *given that a language works on the principle that in any one possible world an assertive sentence has at most one truth-value,* then the fact that the semantic rules have the consequence that some particular sentence is both true and false in some or all possible states of affairs must interfere with the working of the language. *If the language can only do its job provided that every sentence has mutually exclusive sets of truth-conditions and falsity-conditions,* then the existence of a sentence with overlapping sets must prevent the job being done.

But how can the italicized assumption be established? A full answer would require a detailed

analysis of the various functions of language. For the present suffice it to say that if sentences in a language did have overlapping sets of truth- and falsity-conditions then many valid forms of inference would no longer guarantee that if the premisses are true the conclusion is not false. For instance, if p can be both true and false, then the truth of p does not exclude the falsity of

     not-not-p.

So unless the double-valued sentences were of a recondite and easily recognized variety (e.g. sentences about sets of sets?) the usual method of deduction from true premises could not safely be used as a way of avoiding false conclusions: each sentence would have to be directly tested against the world. However, it is perhaps possible to envisage more or less bizarre languages which permit double truth-values and can nevertheless be used.

D.4. But such a language would not be susceptible to a Fregean analysis: for in Frege's theory every sentence is the result of substituting argument-signs into a function-sign and a function is, by definition, a method of correlation which gives a *unique* value for a given argument set. Hence, if any putative sentence does turn out to have more than one truth-value, that means its main function sign does not express a *function* but some other sort of multi-valued correlation, and thus we do not have a proper sentence after all since it is not composed from the right sorts of symbols. To sum up: if we mean by an inconsistent language one whose rules permit a statement to be both true and false (to have overlapping truth-conditions and falsity-conditions) then first, a language properly constructed on Fregean principles could not be inconsistent, and secondly, an inconsistent language would appear to have a number of inconvenient features but might nevertheless be usable provided the inconsistency was not too radical.

D.5. Are colloquial languages inconsistent? As already remarked, only complicated empirical investigation can (perhaps) reveal whether or not they are constructed on Fregean principles. But evidence for inconsistency would be the existence of a sentence which clearly *must* have both truth-values. Besides the versions of the Liar paradox already dealt with there are many well-known sentences which seem to be incapable of being true or false without being both, for instance: 'The set of all non-self-containing sets contains itself', 'The property of not being a property of oneself is not a property of itself', and other logical and semantical paradoxes.

If it can be shown that these sentences contain unsuccessful referring expressions, or that for some other reason the semantic rules do not, after all, assign any truth-value to them (e.g. because they express non-terminating procedures for identifying a truth-value), then they, like the Liar, can be disregarded.[3] In particular, they cannot be used as premises of sound arguments, and this is the answer to the questions raised in A.3.

Alternatively, it may be possible to show that natural languages do not conform to Fregean principles of semantics, so that some degenerate sentences can correspond to *both* truth-values. (Compare paragraph D.4: this possibility is currently being explored by Maurice Tennant.) If there are such sentences, then many natural languages are, after all, inconsistent, though for the reasons mentioned it is still not clear how serious a fault that is![10]

University of Sussex, 1971. (Author now at The University of Birmingham, UK)

NOTES

[1]B. Russell, *An Inquiry into Meaning and Truth,* Pelican Books, 1962, p. 17.

[2]A. Tarski, 'The concept of truth in formalized languages', translated in *Logic, Semantics, Metamathematics,* Clarendon Press, 1955. All page references are to this article. Readers are advised to acquaint themselves either with the (non-technical) introduction and first section, or else with Tarski's shorter essay 'The Semantic conception of truth', in H. Feigl and W. Sellars (eds.), *Readings in Philosophical Analysis,* Appleton-Century-Crofts, 1949.

[3]This article was partly stimulated by a paper on 'Paradoxicality' by L. Hollings, who also helped by criticizing an earlier version, as did C. J. F. Williams, P. Williams, N. Everitt and Carolyn Stone. Hollings has attempted to carry out the programme mentioned in my concluding paragraph. Unfortunately his paper is not yet published.

[4]Frege. 'On sense and reference' and other papers in *Translations from the Philosophical Writings of Gottlob Frege,* by P. Geach and M. Black, Basil Blackwell, 1960. Some aspects of his theory are elaborated further in 'The Thought, A Logical Enquiry', translated in *Mind,* 1956 and *in Philosophical Logic,* edited by P. F. Strawson, O.U.P., 1967. See especially pages 62-5 of *Translations.*

[5]This terminology is suggested by Hollings. See note 3

[6]In 'Functions and rogators', in *Formal Systems and Recursive Functions,* edited by J. N. Crossley and M. A. E. Dummett, North Holland, 1965. The terminology of this paper was unfortunate. The basic aim was to show that the concept of a rule or principle of correlation does for function-signs what Frege's concept of 'sense' does for names. In the present article I shall not use 'function' to refer to a set of ordered pairs.

[7]For example, see pp. 63, 159, 167 of *Translations.*

[8]M. A. E. Dummett, 'Truth', in *Proc. Aristotelian Soc.,* 1958-9, reprinted in *Truth,* edited by G. Pitcher and *in Philosophical Logic,* edited by P. F. Strawson. The situation is a bit more complicated than may appear at first, depending on how *entailment* is defined. For if to say that p entails q is to say that all the truth-conditions of p are included among truth-conditions of q, and if 'not' simply exchanges truth-conditions and falsity-conditions, then (a) p can entail q without not-q entailing not-p (though 'q is not true' must entail 'p is not true'), and (b) not-p is not equivalent to 'p is false'.

[9]The outlines of an analysis can be found in section B above, especially B.9-10. Dummett (see note 8) has criticised Frege's theory for failing to explain the asymmetry between *truth* and *falsity,* and this criticism would apply equally to my extension of Frege's theory. The reply is (a) that in most of their logical and semantic properties truth and falsity are perfectly symmetrical, and (b) that the lack of symmetry can be completely explained in turns of a pragmatic or communicative convention that, except in special contexts, a complete sentence may be uttered only if that sentence corresponds to the value T. The opposite convention would generate a language grammatically identical, but with every sentence expressing the contradictory of what we understand by it. The whole syntactic and semantic apparatus described in this paper could be embedded in a kind of game in which no assertions aie made and in which T and F were perfect duals, provided the above conventions were not involved. It is even arguable that such a game could be learnt by people who had never learnt a language in which true and false assertions could be made. There are further uses of the words 'true' and 'false' not accounted for in the main text or in these remarks.

[10]Since writing this article, I have learnt through Robin Stanton that ideas similar to mine have been developed in connection with computer languages. (See Saul Gorn's 'The identification of the computer and information sciences: Their fundamental semiotic concepts and relationships' in *Foundations of Language,* Vol. 4, No. 4, Nov. 1968.) Some of the ideas of B.9-10, above, are also closely related to, though developed independently of, Saul Kripke's 'Semantical considerations on modal logic', in *Acta Philosophica Fennica,* Fasc. XVI, 1963. Kripke's discussion is much more extensional: he is apparently not concerned with identification *procedures.* I am currently trying to make both discussions more 'realistic' by replacing the concept *of a possible world* with the concept of *a possible extension of a part of the actual world.*