we don't actually exhibit, but only refer to indirectly, a role it shares with the apparatus of second-order ('propositional' or 'sentential') quantifiers. This similarity of function will turn out to be relevant to the diagnosis of the semantic paradoxes.

8

Paradoxes

1 The Liar and related paradoxes

The importance of the Liar paradox to the theory of truth has already become apparent; for Tarski's formal adequacy conditions on definitions of truth are motivated, in large part, by the need to avoid it. It is time, now, to give the Liar and related paradoxes some direct attention on their own account.

Why the 'Liar paradox'? Well, the Liar sentence, together with apparently obvious principles about truth, leads, by apparently valid reasoning, to contradiction; that is why it is called a paradox (from the Greek, '*para*' and '*doxa*', 'beyond belief').[1]

The Liar comes in several variants; the classic version concerns the sentence:

(*S*) This sentence is false

Suppose *S* is true; then what it says is the case; so it is false. Suppose, on the other hand, that *S* is false; then what it says is not the case, so it is true. So *S* is true iff *S* is false. Variants include indirectly self-referential sentences, such as:

The next sentence is false. The previous sentence is true.

and the 'postcard paradox', when one supposes that on one side of a postcard is written:

The sentence on the other side of this postcard is false

and on the other:

The sentence on the other side of this postcard is true.

[1] The 'paradoxes' of material and strict implication – discussed at length in ch. 11 – are, at worst, counter-intuitive, and not, like the Liar, contradictory; hence the scare quotes.

Another variant, the 'Epimenides' paradox, concerns a Cretan called Epimenides, who is supposed to have said that all Cretans are always liars. If a liar is someone who always says what is false, then if what Epimenides said is true, it is false. The Epimenides is, however, somewhat *less* paradoxical than the Liar, since it can be consistently supposed to be false, though not to be true (cf. Anderson 1970). There are also 'truth-teller' ('This sentence is true') and imperative ('Disobey this order') variants.

Other paradoxes involve 'true (false) of ...' rather than 'true (false)'. 'Heterological' means 'not true of itself'; so e.g. 'German', 'long', 'italicised' are heterological, while 'English', 'short', 'printed' are autological, true of themselves. Now, is 'heterological' heterological? Well, if heterological *is* heterological, it is not true of itself; so, it is not heterological. If, however, it is *not* heterological, it is true of itself; so, it is heterological. So 'heterological' is heterological iff 'heterological' is not heterological (Grelling's paradox).

Others again involve 'definable' or 'specifiable'. The number ten is specifiable by a name of one syllable, the number seven by a name of two syllables, the number seventeen by a name of three syllables. Consider, then, the least number not specifiable in fewer than twenty syllables. That number is specifiable in nineteen syllables, by 'the least number not specifiable in fewer than twenty syllables' (Berry's paradox). Let E be the class of decimals definable in a finite number of words, and let its members be ordered as the first, second, third ... etc. Now let N be the number such that if the nth figure in the nth decimal in E is m, then the nth figure in N is $m+1$, or 0 if $m = 9$. Then N differs from every member of E, and yet has been defined in a finite number of words (Richard's paradox).

Other paradoxes involve the concept of set. Some sets are members of themselves, while others are not (e.g. the set of abstract objects, being itself an abstract object, is a member of itself; the set of cows, not being itself a cow, is not). Now consider the set of sets which are not members of themselves. Is it a member of itself or not? If it *is* a member of itself, then it has the property which all its members have, that is, it is *not* a member of itself; if, on the other hand, it is *not* a member of itself, then it has the property which qualifies a set for membership in itself, so it *is* a member of itself. So the set of all sets which are not member of themselves is a member of itself iff it is not a member of itself (Russell's paradox). Other set-theoretical paradoxes include Cantor's paradox: no set can be larger than the set of all sets, but, for any set, there is another, the set of all its subsets, which is larger than it is; and Burali-Forti's: the series of all ordinal numbers has an ordinal number, Ω, say, but the series of all ordinals up to and including any given ordinal exceeds that ordinal by one, so the series of all ordinals up to and including Ω has the ordinal number $\Omega+1$.

This by no means exhausts the range of paradoxes to be found in the literature (cf. Russell 1908a, Mackie 1973 appendix, for more examples). I hope, however, that my list is sufficiently representative to illustrate the kind of problems with which a solution to the paradoxes must deal; the point of considering a number of variants is to enable one to check whether proposed solutions are sufficiently broad in scope.

'Set-theoretical' versus 'semantic' paradoxes?

Though some of these paradoxes had been known long before, they began to be of serious philosophical concern after Russell's discovery of his paradox. Frege had reduced arithmetic to sentence calculus, predicate calculus, and set theory. Russell, however, showed that his paradox was actually a theorem of Frege's system, which was, therefore, inconsistent. (Since Frege had hoped to supply foundations for arithmetic by reducing it to self-evident principles, the fact that his 'self-evident' logic axioms turned out to be contradictory was, naturally, a pretty severe epistemological shock; cf. ch. 1 §2.) The paradoxes cannot be dismissed as mere tricks or puzzles, for they follow from intuitively obvious set-theoretical principles and thus threaten the very foundations of set theory. In view of the fact that anything whatever is derivable from a contradiction, the consequences of paradoxes for a theory in which they are derivable are quite intolerable (but cf. ch. 11 §6 for further thoughts about 'p & $-p \vdash q$'). Russell's paradox operates as a key constraint on attempts to devise consistent set theories; the Liar paradox, similarly, operates as a key constraint on attempts to devise consistent semantic theories.

But this raises an important, though difficult, question. As the comment about the analogy between the role of Russell's paradox in set theory and the role of the Liar paradox in semantic theory suggests, it is possible to classify the paradoxes in two distinct groups, those which essentially involve set-theoretical concepts, such as '$\in$' and 'ordinal number', and those which essentially involve semantic concepts, such as 'false', 'false of ...', and 'definable'. In fact, it is commonplace to distinguish the *set-theoretical* and the *semantic* paradoxes (the distinction goes back to Peano; its currency derives from Ramsey's championship in 1925):

set-theoretical paradoxes	semantic paradoxes
(Ramsey: '*logical*')	(Ramsey: '*epistemological*')
Russell's paradox	Liar paradox and variants
Cantor's paradox	Grelling's paradox
Burali-Forti's paradox	Berry's, Richard's paradoxes
(Essentially involve 'set', '∈', 'ordinal number')	(Essentially involve 'false', 'false of', 'definable')

The second group is the one which is of immediate concern for semantic theory.

Russell himself, however, didn't think of the paradoxes as falling into two distinct groups, *because he thought that they all arose as the result of one fallacy*, from violations of the 'vicious circle principle'. If one supposes that some paradoxes arise because of some peculiarity of set-theoretical concepts, and others because of some peculiarity of semantic concepts, the classification into two groups will be acceptable; but if, like Russell, one thinks that the trouble lies in something deeper, common to all the paradoxes, one will find it misleading. It is hard to deny, I think, that all the paradoxes sketched do have a *prima facie* affinity with each other and that a solution to them all would surely be more satisfying than a solution to only some; and in view of this, the safest course seems to be *not* to beg, by concentrating exclusively on the 'semantic' paradoxes, questions which could be left open.

2 'Solutions' to the paradoxes

Requirements on a solution

Before attempting to assess the solutions which have been offered, it is wise, I think, to try to get a bit clearer just what would constitute a 'solution'. Well, what exactly is the problem? – that contradictory conclusions follow by apparently unexceptionable reasoning from apparently unexceptionable premises. This suggests two requirements on a solution; that it should give a consistent formal theory (of semantics or set theory as the case may be) – in

other words, indicate which apparently unexceptionable premises or principle of inference must be disallowed (the *formal* solution); and that it should, in addition, supply some explanation of *why* that premise or principle is, despite appearances, exceptionable (the *philosophical* solution). It is hard to make precise just what is required of such an explanation, but roughly what is intended is that it should be shown that the rejected premise or principle is of a kind to which there are independent objections – objections independent of its leading to paradox, that is. It is important, though difficult, to avoid supposed 'solutions' which simply *label* the offending sentences in a way that seems, but isn't really, explanatory. Further requirements concern the scope of a solution; it should not be so broad as to cripple reasoning we want to keep (the 'don't cut off your nose to spite your face' principle); but it should be broad enough to block all relevant paradoxical arguments (the 'don't jump out of the frying pan into the fire' principle); the 'relevant', of course, glosses over some problems. At the formal level, the latter principle urges simply that the solution be such as to restore consistency. Frege's response to the inconsistency found by Russell in his set-theory was a formal restriction which avoids Russell's paradox but still allows closely related paradoxes, and thus breaches this requirement (see Frege 1903, Quine 1955, Geach 1956). At the philosophical level, the 'frying pan and fire' principle urges that the explanation offered go as deep as possible; this, of course, is what underlies my hunch that a solution to *both* 'semantic' and 'set-theoretic' paradoxes, if it were possible, would be preferable to a solution local to one group.

The force of these requirements may perhaps be appreciated by looking briefly at some proposed solutions which fail to meet them.

It is sometimes suggested that the paradoxes be resolved by banning self-reference; but this suggestion is at once too broad and too narrow. It falls foul of the 'don't cut off your nose to spite your face' principle: for not only are many perfectly harmless sentences ('This sentence is in English', 'This sentence is in red ink') self-referential (cf. Popper 1954, Smullyan 1957), but also some mathematical argument, including Gödel's proof of the incompleteness of arithmetic, makes essential use of self-referential sentences (cf. Nagel and Newman 1959 and Anderson 1970); so that the consequences of a ban on self-reference would be very serious. And yet, since not all the variants of the Liar are straightforwardly self-referential (neither sentence in 'The next sentence is false. The previous sentence is true' refers to itself) this proposal is, at the same time, too narrow.

The argument to a contradiction from the Liar sentence uses the assumption that 'This sentence is false' is either true or false; and so, unsurprisingly, it has often been suggested that the way to block the argument is to deny this assumption. Bochvar proposed (1939) to deal with the Liar by adopting a 3-valued logic in which the third value, 'paradoxical', is to be taken by the recalcitrant sentences. (See also Skyrms 1970a, 1970b, and ch. 11 §3.) This proposal, too, is in danger of being both too broad and too narrow: too broad, because it requires a change in elementary (sentence calculus) logical principles; and yet still too narrow, for it leaves problems with the 'Strengthened Liar' paradox – the sentence:

This sentence is either false or paradoxical

which is false or paradoxical if true, true if false, and true if paradoxical.

Another approach also denies that the Liar sentence is true or false, without, however, suggesting that it has a third truth-value, by arguing that it is not an item of the appropriate kind to have a truth-value. Only statements, it is argued, are true or false, and an utterance of the Liar sentence wouldn't constitute a statement. (See Bar-Hillel 1957, Prior 1958, Garver 1970; and cf. – *mutatis mutandis* with 'proposition' for 'statement' – Kneale 1971.) This kind of approach suffers, I think, from inadequate explanatoriness – it doesn't supply a suitable rationale for denying the offending sentences a truth-value. Even granted for the sake of argument that only statements or propositions can be either true or false (but granted *only* for the sake of argument – cf. ch. 6) one would need an argument why in the case of the Liar one does not have an item of the appropriate kind. After all, the Liar sentence suffers from no obvious deficiency of grammar or vocabulary. The minimum requirements would be, first, a clear account of the conditions under which an utterance of a sentence constitutes a statement; second, an argument why no utterance of the Liar could fulfil these conditions; third, an argument why only statements can be true or false. Otherwise, one is entitled to complain that the solution is insufficiently explanatory.

Russell's solution: the theory of types, the vicious circle principle

Russell offers (1908a) both a formal solution, the theory of types, and a philosophical solution, the vicious circle principle.

Nowadays, it is customary to distinguish in Russell's formal solution the simple and the ramified theory of types. *The simple theory of types* divides the universe of discourse into a hierarchy: individuals (type 0), sets of individuals (type 1), sets of sets of individuals (type 2), ... etc., and correspondingly subscripts variables with a type index, so that x_0 ranges over type 0, x_1 over type 1 ... etc. Then the formation rules are restricted in such a way that a formula of the form '$x \in y$' is well-formed only if the type index of y is one higher than that of x. So, in particular, '$x_n \in x_n$' is ill-formed, and the property of not being a member of itself, essential to Russell's paradox, cannot be expressed. *The ramified theory of types* imposes a hierarchy of orders of 'propositions' (closed sentences) and 'propositional functions' (open sentences), and the restriction that no proposition (propositional function) can be 'about', i.e. contain a quantifier ranging over, propositions (propositional functions) of the same or higher order as itself. 'True' and 'false' are also to be subscripted, depending on the order of the proposition to which they are applied; a proposition of order n will be true (false) $n+1$. The Liar sentence, which says of itself that it is true, thus becomes inexpressible, just as the property of not being a member of itself did in the simple theory. (I have simplified considerably; see Copi 1971 for a more detailed account.)

Russell himself, however, did not see the paradoxes as falling into two distinct groups; he believed that *all* the paradoxes arose from one and the same fallacy, from violations of what he, following Poincaré, called 'the vicious circle principle' (V.C.P.):

> 'Whatever involves *all* of a collection must not be one of the collection'; or, conversely, 'If, provided a certain collection had a total, it would have members only definable in terms of that total, then the said collection has no total'. [Footnote: I mean that statements about *all* its members are nonsense.] (1908a p. 63)

He states the V.C.P. in several, not obviously equivalent, ways: e.g. a collection mustn't 'involve', or, 'be definable only in terms of' itself. The V.C.P. motivates the type/order restrictions imposed upon the formal theory, by showing that what the formulae ruled ill-formed say is demonstrably meaningless. It is important that the very same philosophical rationale is given for both the simple and the ramified theories. Indeed, since Russell held that sets are really logical constructions out of propositional functions, he saw the restrictions of the simple theory as a special case of those of the ramified theory (cf. Chihara 1972, 1973).

At both the formal and the philosophical levels, Russell's account runs into difficulty. Formally, there is some danger that Russell has cut off his nose to spite his face; the restrictions avoid the paradoxes, but also block certain desired inferences. Remember that Russell was trying to complete the programme, begun by Frege, of reducing arithmetic to 'logic', i.e. to sentence calculus, first-order predicate calculus, and set theory. However, the type restrictions block the proof of the infinity of the natural numbers, and the order restrictions block the proof of certain bound theorems. In *Principia Mathematica* these proofs are saved by the introduction of new axioms, respectively, the axiom of infinity and the axiom of reducibility; this ensures the derivability of the Peano postulates for arithmetic; but the *ad hoc* character of these axioms lessens the plausibility of the claim that arithmetic has been reduced to a *purely logical* basis. Still, it could be thought that these difficulties, though they cast doubt on the feasibility of Russell's logicism, don't necessarily show his solution to the paradoxes to be misguided.

But one's suspicions are confirmed by difficulties at the philosophical level. In the first place, the V.C.P. certainly isn't stated with all the precision that might be desired; and it is correspondingly difficult to see what, exactly, is wrong with violations of it. Ramsey commented that he could see nothing objectionable about specifying a man as the one with, say, the highest batting average of his team – a specification apparently in violation of the V.C.P. Not *all* the circles ruled out by the V.C.P., he urged, are truly vicious (notice the analogy to the difficulties in the proposal to ban all self-referential sentences).

However, despite these difficulties, Russell's diagnosis and solution have continued to be influential; later, in §3, I shall argue that Russell's approach is, indeed, in certain respects, on the right lines. But my immediate concern is with other solutions which resemble Russell's in interesting ways. His diagnosis is echoed in Ryle's approach. Ryle 1952 argues that 'The current statement is false' must be unpacked as 'The current statement (namely, that the current statement...[namely, that the current statement...{namely ...etc.}]) is false', and no completely specified statement is ever reached. Like Russell, Ryle thinks that the 'self-dependence' of the Liar sentence somehow robs it of sense. Mackie 1973 agrees with Russell and Ryle that the problem lies in the Liar's 'vicious self-dependence', but prefers to say, for the good reason that the Liar sentence is apparently correctly constructed from *bona fide* components, that the upshot is not meaninglessness but 'lack of content'. However, since he is careful to distinguish 'lack of content' from lack of meaning *and* from lack of truth-value, one is left somewhat at a loss to understand just what lack of content is lack *of*. And Tarski's approach to the semantic paradoxes, to which I turn next, has some significant similarities (observed by Russell 1956; and cf. Church 1976) to the Russellian hierarchy of orders of propositions.

Tarski's solution: the hierarchy of languages

Tarski diagnoses the semantic paradoxes (to which his attention is restricted) as resulting from the two assumptions:

(i) that the language is semantically closed, i.e. contains (a) the means to refer to its own expression, and (b) the predicates 'true' and 'false'
(ii) that the usual logical laws hold

and, being reluctant to deny (ii) (but cf. the comments, above, on Bochvar's proposal) denies (i), proposing as a formal adequacy condition that truth be defined for *semantically open* languages. So Tarski proposes a *hierarchy of languages*:

the object language, O,
the metalanguage, M,
which contains (a) means of referring to expressions of O and (b) the predicates 'true-in-O' and 'false-in-O',
the meta-metalanguage, M′,
which contains (a) means of referring to expressions of M and (b) the predicates 'true-in-M', and 'false-in-M',
etc.

Since, in this hierarchy of languages, truth for a given level is always expressed by a predicate of the next level, the Liar sentence can appear only in the harmless form 'This sentence is false-in-O', which must itself be a sentence of M, and hence cannot be true-in-O, and is simply false instead of paradoxical.

Though the appeal of Tarski's theory of truth has won this proposal a good deal of support, there have also been criticisms of its 'artificiality'. The language hierarchy and the relativisation of 'true' and 'false' avoid the semantic paradoxes, but they seem to lack intuitive justification independent of their usefulness in this regard. In other words, Tarski's approach seems to give a formal, but not a philosophical, solution. The reason Tarski gives for requiring semantic openness, is, simply, that semantic closure leads to paradox. There *is* an independent rationale for the relativisation of 'true' and 'false' to a language – that Tarski is defining 'truth' for sentences (wffs), and one and the same sentence (wff) can have a different meaning, and hence a different truth-value, in different languages; but this rationale does not supply any independent justification for insisting that 'true-in-L' always be a predicate, not of L, but of the metalanguage of L.

Intuitively, one does not think of 'true' as systematically ambiguous in the way Tarski suggests it must be. Perhaps this counter-intuitiveness would not, by itself, be an overwhelming consideration. But Kripke (1975) points out that ordinary ascriptions of truth and falsity cannot even be assigned *implicit* levels. Suppose, for instance, that Jones says:

All of Nixon's utterances about Watergate are false.

This would have to be assigned to the next level above the highest level of any of Nixon's utterances about Watergate; but not only will we ordinarily have no way of determining the levels of Nixon's utterances about Watergate, but also in unfavourable circumstances it may actually be impossible to assign levels consistently – suppose that among Nixon's utterances about Watergate is:

All of Jones' utterances about Watergate are false

then Jones' utterance has to be at a level one higher than all of Nixon's, and Nixon's at a level one higher than all of Jones'.

Tarski's approach, Kripke argues, fails to take adequate account of the 'risky' character of truth-ascriptions. Quite ordinary assertions about truth and falsity, he points out, are apt to turn out paradoxical if the empirical facts are unfavourable. Suppose e.g. that Nixon had said that all of Jones' utterances about Watergate are true; then

Jones' assertion that all of Nixon's assertions about Watergate are false would be false if true and true if false (cf. the 'postcard paradox' in §1). The moral, he suggests, is that one can scarcely expect the recalcitrant sentences to be distinguished by any syntactic or semantic feature, but must seek a rationale which allows that paradox may arise with respect to any truth-ascription if the facts turn out badly.[1]

Kripke's solution: groundedness

Kripke seeks to supply an explanation of the source of paradox which is more satisfactory in this respect, and then to build a formal theory on this basis. (My hunch is that this is the right way round to go about it.) His proposal depends upon the rejection of the idea – taken for granted by Tarski – that the truth-predicate must be totally defined, that is to say, that every suitably well-formed sentence must be either true or false. It thus has affinities both with Bochvar's proposal of a 3-valued logic, and with the no-item proposals discussed above. But Kripke stresses that his idea is *not* that paradoxical sentences have some non-classical truth-value, but that they have *no* truth-value.

The key idea in the explanation of how ordinary sentences are assigned truth-values – and how extraordinary sentences fail to get a value – is the concept of *groundedness*, first introduced by Herzberger 1970. Kripke explains the idea as follows:

Suppose one is trying to explain the word 'true' to someone who doesn't understand it. It could be introduced by means of the principle that one may assert that a sentence is true just when one is entitled to assert that sentence, and one may assert that a sentence is not true just when one is entitled to deny it. Now given that the learner is entitled to assert that:

Snow is white

this explanation tells him that he is entitled to assert that:

'Snow is white' is true.

Now he can extend his use of 'true' to other sentences, e.g. as 'Snow

[1] Kripke also makes the technical objection that Tarski's hierarchy has not been extended to transfinite levels, and that, furthermore, there are difficulties about so extending it.

is white' occurs in Tarski 1944, the explanation allows him to assert that:

> Some sentence in 'The semantic conception of truth' is true.

And he can also extend his use of 'true' to sentences which already contain 'true', e.g. to assert that:

> ''Snow is white' is true' is true

or:

> 'Some sentence in 'The semantic conception of truth' is true' is true.

The intuitive idea of *groundedness* is that a sentence is grounded just in case it will eventually get a truth-value in this process. Not all sentences *will* get a truth-value in this way; among the 'ungrounded' sentences that won't are:

> This sentence is true

and:

> This sentence is false.

This idea has affinities with the notion – expressed in Russell's V.C.P. and by Ryle and Mackie – that what's wrong with paradoxical sentences is a kind of vicious self-dependence. However, ungrounded sentences are allowed to be meaningful, whereas Russell's idea is that violation of the V.C.P. results in meaninglessness.

Formally, this idea is represented (I simplify considerably) in a hierarchy of interpreted languages where at any level the truth-predicate is the truth-predicate for the next lowest level. At the lowest level, the predicate '*T*' is completely undefined. (This corresponds to the initial stage in the intuitive account.) At the next level, the predicate '*T*' is assigned to wffs which don't themselves contain '*T*'. It is assumed that this assignment will be in accordance with Kleene's rules giving the assignment of values to compound wffs given the assignment – or lack of assignment – to their components: '$-p$' is true (false) if 'p' is false (true), undefined if 'p' is undefined; '$p \vee q$' is true if at least one disjunct is true (whether the other is true, false, or undefined), false if both disjuncts are false, otherwise undefined; '$(\exists x)\, Fx$' is true (false) if 'Fx' is true for some (false for every) assignment to x, otherwise undefined. (This corresponds to the first stage, in which the learner assigns 'true' to a

sentence if he is entitled to assert the sentence.) At each level the wffs assigned '*T*' and '*F*' at a previous level retain those values, but new wffs, for which '*T*' was previously undefined, are assigned values – '*T*' gets *more defined* as the process goes on. But the process doesn't go on indefinitely with new sentences getting values at each level; eventually – at a 'fixed point' – the process stops. Now the intuitive idea of groundedness can be formally defined: a wff is grounded if it has a truth-value at the smallest fixed point, otherwise ungrounded. The smallest or 'minimal' fixed point is the first point at which the set of true (false) sentences is the same as the set of true (false) sentences at the previous level. All paradoxical sentences are ungrounded, but not all ungrounded sentences are paradoxical; a paradoxical sentence is one that cannot consistently be assigned a truth-value at *any* fixed point. This supplies some explanation of why 'This sentence is true' seems to share some of the oddity of 'This sentence is false', and yet, unlike the Liar sentence, is consistent. A truth-value *can* be given to 'this sentence is true', but only *arbitrarily*; a truth-value *cannot* consistently be given to 'This sentence is false'. The picture also allows for the 'riskiness' of truth-ascriptions: for the paradoxical character of a sentence may be either intrinsic (as it would be with 'This sentence is false') or empirical (as it would be with 'The sentence quoted on p. 147 ll. 22–3 is false').

I observed above that the relaxation of the requirement that 'true' be fully defined, the admission of truth-value gaps, gave Kripke's idea some analogy, also, to proposals, like Bochvar's, that the semantic paradoxes be avoided by resort to a 3-valued logic. This raises the question, how Kripke avoids the criticisms made earlier of Bochvar's solution. Kripke himself stresses that he does not regard his use of Kleene's '3-valued' valuation rules as a challenge to classical logic. Whether the use of 3-valued matrices necessarily carries such a challenge, is a difficult question, on which I shall have more to say in ch. 11 §3; for the present I shall allow Kripke's claim that his proposals are compatible with logical conservatism. What, though, of the Strengthened Liar?

Kripke doesn't tackle this issue directly, but it is possible to work out what he would say about it. The notions of 'groundedness' and 'paradoxicality', he says, unlike the concept of truth, don't belong in his hierarchy of language levels. (Consider again the intuitive picture of a learner having the concept of truth explained to him. His instructions give him no way to assign a truth-value to an ungrounded sentence like 'This sentence is true'; but he cannot conclude that 'This sentence is true' is not true, for his instructions do tell him that he may deny that a sentence is true only if he is entitled to deny that sentence.) Now if 'paradoxical' belongs, not in the hierarchy of language levels, but in the metalanguage of that hierarchy, then Kripke can draw the teeth of the Strengthened Liar, 'This sentence is either false or paradoxical' in much the way that Tarski draws the teeth of the Liar. But this may occasion some dissatisfaction; for it *is* a little disappointing to find that the novelty of Kripke's approach to the Liar must be compromised by a neo-Tarskian dismissal of the Strengthened Liar. (*Is* it indifferent whether one is hung for a sheep or a lamb?)

It will be worthwhile to summarise the main points of comparison and contrast between Kripke's approach, Russell's theory of types, and Tarski's language hierarchy:

RUSSELL	TARSKI	KRIPKE
formal solution		
hierarchy of orders of propositions	hierarchy of languages (problems with transfinite levels)	hierarchy of language levels (with limit levels)
systematic ambiguity of 'true' and 'false'	distinct truth and falsity predicates at each level	single, univocal truth-predicate, with application extended up to minimal fixed point
'true' and 'false' completely defined	'true' and 'false' completely defined	'true' and 'false' only partially defined
'This sentence is false' meaningless	'This sentence is false-in-O' false-in-M	'This sentence is false' neither true nor false.
rationale		
V.C.P.	(language-relativisation of 'true')	groundedness

3 Paradox without 'false'; some remarks about the redundancy theory of truth; and the V.C.P. again

I shan't, I'm afraid, be able to offer, in conclusion, a novel solution to the paradoxes. The purpose of the present section is rather more modest: to redeem the promise (pp. 130, 134) to comment upon the consequences for the paradoxes of the redundancy theory of truth, with its resistance to the idea of truth as a metalinguistic predicate; a consequence of considerations which this investigation brings to light, however, will be some support for a proposal which, as I shall argue, has affinities with the V.C.P.

One of Tarski's reasons for refusing to countenance the treatment of quotation as a function, and hence, for denying that truth could be defined by generalising the (T) schema, to obtain '(p) ('p' is true iff p)', was, if you recall (p. 104 above) that with quotation functions paradox would ensue even without the use of the predicates 'true' and 'false'. (And Tarski's semantic openness requirement, of course, would be powerless to cope with paradox generated without semantic predicates.) Tarski's argument goes as follows:

> Let 'c' abbreviate 'the sentence numbered 1'.
> Now, consider the sentence:
> 1. $(p)\ (c = \text{'}p\text{'} \rightarrow -p)$
> It can be established empirically that:
> 2. $c = \text{'}(p)\ (c = \text{'}p\text{'} \rightarrow -p)\text{'}$
> and so, assuming that:
> 3. $(p)\ (q)\ (\text{'}p\text{'} = \text{'}q\text{'} \rightarrow p \equiv q)$

'by means of elementary logical laws we easily derive a contradiction' (1931 p. 162).[1] Notice that here one has a paradox that arises, not intrinsically in the nature of a single statement, but extrinsically, due, as Kripke would put it, to the facts' turning out badly. Tarski's diagnosis is that quotation functions are the root of the trouble, and must not be allowed. Some writers have, in response, suggested that, rather than quotation functions being disallowed altogether, certain restrictions should be imposed upon them; Binkley, for instance,

[1] Tarski does not give the derivation, but it would presumably go along the following lines. From 1, if $c = \text{'}(p)\ (c = \text{'}p\text{'} \rightarrow -p)\text{'}$, then $-(p)\ (c = \text{'}p\text{'} \rightarrow -p)$, so, given 2, $-(p)\ (c = \text{'}p\text{'} \rightarrow -p)$; hence by RAA, -1. If -1, then $(\exists p)\ (c = \text{'}p\text{'}\ \&\ p)$. Suppose for instance that $c = \text{'}q\text{'}\ \&\ q$; then $\text{'}q\text{'} = \text{'}(p)\ (c = \text{'}p\text{'} \rightarrow -p)\text{'}$ since both $= c$, hence, by 3, $q \equiv (p)\ (c = \text{'}p\text{'} \rightarrow -p)$. But q; so $(p)\ (c = \text{'}p\text{'} \rightarrow -p)$, i.e. 1. Hence, 1 & -1.

suggests (1970) a 'no-mixing' rule, which prevents one and the same quantifier from binding both variables inside, and variables outside, quotation marks, and hence disallows 1 above. But neither Tarski's diagnosis, nor this kind of response, can be quite right; for an analogous paradox can be derived without the use of quotation marks:

> Let '§' be an operator forming a term from a sentence; it could be read e.g. 'the statement that...'
> Let 'c' abbreviate 'the statement made by sentence numbered 1'.

Now, consider the sentence:

1. $(p)\ (c = \S p \rightarrow -p)$

It can be established empirically that:

2. $c = \S (p)\ (c = \S p \rightarrow -p)$

and a contradiction follows as before.[1] Now one might try, again, to impose restrictions on term-forming operators like '§'; for instance, following the example of Harman 1971, one might rule that if 'p' belongs to L, '§p' must belong, not to L, but to the metalanguage of L. But this kind of manoeuvre – quite apart from its disagreeably *ad hoc* character – again seems not to go to the heart of the problem; for an analogous paradox can be derived without the use of '§'; if one could let 'c' abbreviate the sentence numbered 1 (instead of 'the sentence numbered 1'; 'c' now abbreviates a sentence, not a term):

1. $(p)\ ((c \equiv p) \rightarrow -p)$

so that, in virtue of the abbreviation,

2. $(c \equiv (p)\ ((c \equiv p) \rightarrow -p))$

and once again a contradiction would be derivable.

[1] Some comments are called for about the moral to be drawn about quotation marks. Tarski holds (and Quine agrees) that the result of enclosing an expression in quotation marks is an expression denoting the enclosed expression, but *of which the enclosed expression is not genuinely a part*. The idea that quotation forms a sort of 'logical block', that 'dog' isn't part of "dog", leads to very curious consequences, and is quite counter-intuitive (cf. Anscombe 1957). So it is a relief to find that the failure of Tarski's diagnosis of the paradox leaves one free to treat quotation as a function; cf. Belnap and Grover 1973, Haack 1975 for more detailed discussion.

This shouldn't be too surprising. For the effect of a truth-predicate can, as investigation of the redundancy theory (pp. 130–4 above) showed, be achieved by using second-order (propositional) quantifiers; and adding negation gives the effect of 'false'. So the fact that a Liar-type paradox is derivable without explicit use of semantic predicates, provided propositional quantifiers and negation are available, was to have been expected.

But how are paradoxes of this kind to be avoided? Suppose the propositional quantifiers are interpreted – as I recommended in ch. 4 §3 – substitutionally. On a substitutional interpretation, a quantified formula, A, of the form $(v)\ \Phi\ (v)$, is true just in case all its substitution instances, $\Phi(s)$, are true. Since in the case under consideration the quantifier binds sentence letters, the substituends for v will be wffs, and may, therefore, themselves contain quantifiers. Now, the usual conditions of definitional adequacy require that only substituends which contain fewer quantifiers than A itself be allowed; otherwise ineliminability would result (see Marcus 1972, and cf. Grover 1973). This restriction is in no way *ad hoc*, since it is only a special case of quite ordinary conditions on definitions; but at the same time it is sufficient to block the paradoxical argument where the wff substituted for 'p' in '$(p)\ ((c \equiv p) \rightarrow -p)$' is '$(p)\ ((c \equiv p) \rightarrow -p)$'.

It wouldn't be altogether fanciful, I think, to see affinities between this idea and the theory of types, with its hierarchy of propositions ordered according to what propositional quantifiers occur in them; nor to see affinities between the motivation for the restriction on substituends for sentential variables, and the V.C.P. Russell's argument why a proposition 'about all propositions' can't itself be a member of that totality is that it 'creates' a new proposition not previously belonging to that totality, which is unconvincing since it assumes, what it is intended to prove, that the proposition about all propositions isn't already a member of that totality; Ryle and Mackie, however, urge, in favour of the V.C.P., that violations of it lead to a 'vicious self-dependence' which *results in ineliminability*. And, finally, the fact that paradoxes can be generated without semantic predicates might be thought to suggest that there might, after all, be *something* in Russell's hunch that the paradoxes weren't to be handled in distinct groups according as semantic or set-theoretic predicates occurred essentially in them, but were to be handled together, as all the result of one fallacy.