NOTES ON THE THEORY OF REFERENCE

1

When the cleavage between meaning and reference is properly heeded,[1] the problems of what is loosely called semantics become separated into two provinces so fundamentally distinct as not to deserve a joint appellation at all. They may be called the *theory of meaning* and the *theory of reference*. 'Semantics' would be a good name for the theory of meaning, were it not for the fact that some of the best work in so-called semantics, notably Tarski's, belongs to the theory of reference. The main concepts in the theory of meaning, apart from meaning itself, are *synonymy* (or sameness of meaning), *significance* (or possession of meaning), and *analyticity* (or truth by virtue of meaning). Another is *entailment*, or analyticity of the conditional. The main concepts in the theory of reference are *naming*, *truth*, *denotation* (or truth-of), and *extension*. Another is the notion of *values* of variables.

Boundaries between fields are not barriers. Given any two fields, it is conceivable that a concept might be compounded of concepts from both fields. But if this were to happen in the case of the theory of meaning and the theory of reference, we should probably reckon the hybrid concept to the theory of meaning—simply because the theory of meaning is in a worse state than the theory of reference, and is consequently the more serious of the two presuppositions.

As applied to discourse in an explicitly quantificational form of language, the notion of ontological commitment belongs to the

[1] See above, pp. 9, 20.

theory of reference. For to say that a given existential quantification presupposes objects of a given kind is to say simply that the open sentence which follows the quantifier is true of some objects of that kind and none not of that kind. In so far as we undertake to speak of ontological commitment on the part of discourse not in an explicitly quantificational form of language, on the other hand, and to rest our case on a supposed synonymy between the given statements and their translations in a quantificational language, we are of course involved in the theory of meaning.

Given a theory, one philosophically interesting aspect of it into which we can inquire is its ontology. But we can also inquire into its *ideology* (to give a good sense to a bad word): what ideas can be expressed in it? The ontology of a theory stands in no simple correspondence to its ideology. Thus, consider the usual theory of real numbers. Its ontology exhausts the real numbers, but its ideology—the range of severally expressible ideas—embraces individual ideas of only certain of the real numbers. For it is known that no notation is adequate to the separate specification of each real number.[2] On the other hand, the ideology also embraces many such ideas as sum, root, rationality, algebraicity, and the like, which need not have any ontological correlates in the range of the variables of quantification of the theory.

Two theories can have the same ontology and different ideologies. Two theories of real numbers, for example, may agree ontologically in that each calls for all and only the real numbers as values of its variables, but they may still differ ideologically in that one theory is expressed in a language into which the sentence:

(1)   the real number x is a whole number

can be translated, while the other theory is not. Note the importance of this particular example; Tarski [1] has proved the completeness of a certain elementary theory T of real numbers, and we know from Gödel's proof [2] of the incompletability of

[2] See, for example, my [1], p. 273f.

the theory of whole numbers that Tarski's achievement would have been impossible if (1) were translatable into the notation of T.

It is instructive to observe that the ontology of a theory may embrace objects of some kind K even where kind K is not definable in the terms of the theory. For example, the ontology of T can be shown to embrace the whole real numbers despite the fact that (1) is not translatable into the notation of T.

I have described the ideology of a theory vaguely as asking what ideas are expressible in the language of the theory. Ideology thus seems to involve us in the idea of an idea. But this formulation may well be dropped, and with it the term 'ideology'. For such substantial work as would fall under ideology consists precisely of the theory of *definability*; and this theory, far from depending on the idea idea, stands clear of the theory of meaning altogether and falls squarely within the theory of reference. The word 'definition' has indeed commonly connoted synonymy,[3] which belongs to the theory of meaning; the mathematical literature on definability,[4] however, has to do with definability only in the following more innocuous sense. A general term *t* is said to be *definable* in any portion of language which includes a sentence *S* such that *S* has the variable '*x*' in it and is fulfilled by all and only those values of '*x*' of which *t* is true. Definability so construed rests only on sameness of reference—sameness of extension on the part of *t* and *S*. Definability of expressions of other categories than that of general terms may be explained in fairly parallel fashion. A typical theorem of the theory of definability in this sense, and hence of the theory of reference, is the above observation that 'whole' is not definable in T.

In Essays II and III we dwelt on the sorry state of the theory of meaning. The theory of reference, actually, has also had its troubles, for it is the scene of the so-called semantic paradoxes.

2

3 See above, pp. 24ff.
4 Tarski [3]; Robinson; Myhill; Church and Quine. See also p. 80 above.

The best known of those paradoxes is the Epimenides, anciently rendered thus: Epimenides the Cretan says that Cretans always lie; hence his statement must, if true, be a lie. Here obviously we are involved in no real paradox, but only in the conclusion that Epimenides here lies and some Cretans sometimes do not. The situation can be developed into a paradox, however, by adopting three historical premisses: not only (a) that Epimenides was a Cretan and (b) that Epimenides said that Cretans never speak the truth, but also (c) that all *other* statements by Cretans were indeed false. Then Epimenides' statement becomes false if true, and true if false—an impossible situation.

It is instructive to contrast this paradox with the riddle of the barber. A man of Alcalá is said to have shaved all and only those men of Alcalá who did not shave themselves; and we find that he shaved himself if and only if he did not.[5] This is no real paradox, but only a *reductio ad absurdum* proof that there was no such man in Alcalá. On the other hand the Epimenides, as last refined, cannot be thus dismissed. For whereas it is evident that a self-contradictory condition was imposed on the barber, we cannot so unconcernedly acknowledge incompatibility of the three palpably independent conditions (a)-(c).

A variant of the Epimenides paradox, likewise ancient, is the *pseudomenon* of the Megarian school: 'I am lying'. A still simpler version may be put thus:

(2)   (2) is false.

Clearly (2), which reads as above, is false if true and true if false.

In an effort to escape the self-contradictory predicament of having to regard (2) as both true and false, one might protest that (2) is simply meaningless, on the ground that an attempt to expand the reference '(2)' in (2) into a specific quotation of an actual statement leads into an infinite regress. But this protest can be silenced by resorting to a more complex version, as follows:

5 A version of this was attributed by Russell ([4], pp. 354f) to an unnamed acquaintance.

(3) 'does not produce a true statement when appended to its own quotation' produces a true statement when appended to its own quotation.

The above statement is readily seen to say that its own denial is true.

Another so-called semantical paradox is Grelling's, which consists in asking whether the general term 'not true of itself' is true of itself; clearly it will be true of itself if and only if it is not. A third is Berry's, concerning the least number not specifiable in less than nineteen syllables. That number has just now been specified in eighteen syllables.[6]

These paradoxes seem to show that the most characteristic terms of the theory of reference, namely, 'true', 'true of', and 'naming' (or 'specifying'), must be banned from language as meaningless, on pain of contradiction. But this conclusion is hard to accept, for the three familiar terms in question seem to possess a peculiar clarity in view of these three paradigms:

(4) '_____' is true if and only if _____,
(5) '_____' is true of every _____ thing and nothing else.
(6) '_____' names _____ and nothing else.

(4) holds when any one statement is written in the two blanks; (5) holds when any one general term (in adjective form, or, omitting 'thing', in substantive form) is written in the two blanks; and (6) holds whenever any one name (which really names, that is, whose object exists) is written in the two blanks.

Strictly, the notions of the theory of reference, and likewise those of the theory of meaning (if they are countenanced at all), are relative always to a language; the language figures, albeit tacitly, as a parameter. Thus it will be recalled that the problem of construing 'analytic' was recognized as the problem of construing 'analytic in L' for variable 'L'.[7] Similarly, a statement, thought of as a string of letters or sounds, is never simply true, but true in language L for appropriate L. This is not a philo-

[6] See Whitehead and Russell, vol. 1, p. 61.
[7] See above, pp. 32ff.

sophical doctrine of the relativity of all fact to language; the point is much more superficial. The point is merely that a given string of letters or sounds could constitute at once a statement say of English and a statement (different in meaning, to borrow a phrase) of Frisian, and it might happen in its English meaning to be true and in its Frisian meaning to be false.[8] Properly, therefore, (4)–(6) should appear rather thus:

(7) '_____' is true-in-L if and only if _____,
(8) '_____' is true-in-L of every _____ thing and nothing else.
(9) '_____' names-in-L _____ and nothing else.

But now it becomes necessary that L and the language in which (7)–(9) themselves are couched (namely, English) be the same, or at least that they overlap to the extent of any notations to which (in the role of '_____') we propose to apply (7)–(9). Otherwise we might even get falsehoods as instances of (7)–(9), in the rare event of a coincidence such as was imagined between Frisian and English; but usually we would get mere nonsense, of the type:

(10)   'Der Schnee ist weiss' is true-in-German if and only if der Schnee ist weiss.

The quotation at the beginning of (10) is indeed a good English word, constituting a name of a German statement; but the rest of (10) is a meaningless jumble of languages.

If, however, we were to pool German and English to form a composite language, German-English, then (10) could be said to be true in German-English. In general, if language L (for example, German) is contained in language L' (for example, German-English), so that L' is simply L or else L plus some supplementary vocabulary or grammatical constructions, and if the portions, at least, of English usage which figure in (7) above (apart from the blanks) are part of L', then the result of putting any one statement of L for the blanks in (7) is true in L'. Correspondingly for (8); if L is contained in L', and the constant

[8] The need to allow in theoretical semantics for such interlinguistic coincidences has been noted in another connection by Church [5].

matter in (8) is part of L', then the result of putting any one general term of L for the blanks in (8) is true in L'. Correspondingly for (9).

Now it turns out that the semantical paradoxes noted earlier do not arise if we take these two precautions: qualify (4)–(6) in the fashion (7)–(9), and banish such terms as 'true-in-L', 'true-in-L of', and 'names-in-L' from the language L itself. These terms, appropriate to the theory of reference of L, may continue to exist in a more inclusive language L' containing L; and the paradigms (7)–(9) may then continue to hold in L', without paradox, as long as the statements or terms which fill the blanks belong not merely to L' but specifically to L.

3

It must be noted that the paradigms (4)–(6) were not strictly definitions of the verbs 'is true', 'is true of', and 'names', nor are (7)–(9) definitions of the verbs 'is true-in-L', 'is true-in-L of', and 'names-in-L'. For the paradigms enable us to eliminate those verbs only from positions preceded by quotations; not from positions preceded, for example, by pronouns, or variables of quantification. Nevertheless, the paradigms resemble definitions in this fundamental respect: they leave no ambiguity as to the extensions, the ranges of applicability, of the verbs in question. In the case of (7) this is seen as follows. Supposing two different interpretations of 'true-in-L' compatible with (7), let us distinguish them by writing 'true$_1$-in-L' and 'true$_2$-in-L', and let (7)$_1$ and (7)$_2$ be (7) with these respective subscripts inserted. From (7)$_1$ and (7)$_2$ it follows logically that

'——' is true$_1$-in-L if and only if '——' is true$_2$-in-L,

no matter what statement of L we write for '——'. Thus truth$_1$-in-L and truth$_2$-in-L coincide. Similar reasoning works for (8) and (9).

Tarski, to whom the reflections on truth in the foregoing pages are largely due ([4], [6]), goes on to show further that 'true-in-L' is in fact genuinely definable in L' if certain general circumstances obtain. Let us suppose that L is a language of the

general form described on page 30 above, and that the whole vocabulary of predicates of L is fixed in a finished list. Suppose further that L' contains L and, in addition, some specifically linguistic terminology adequate to naming each individual symbol of L and to expressing concatenation of symbols. Suppose finally that L' possesses a normal complement of logical notation, including that of the theory of classes. Now Tarski shows how to formulate within the notation of L' a sentence '——x——' which fulfills:

——x—— if and only if ——

whenever a statement of L is put for '——' and a name of that statement is put for 'x'. In short, he shows that 'true-in-L', in a sense conforming to (7), is definable in L', in a sense of 'definable' conforming to the early pages of the present essay.[9] His actual construction will be passed over here.

In certain formalized notations capable of treating their own grammar or capable of treating some subject matter in which a model of that grammar can be constructed, Tarski's method enables us to derive a form of the Epimenides paradox tantamount to (3). Gödel's theorem [2] of the incompletability of number theory, indeed, can be got by a *reductio ad absurdum* along these lines; such is my method in [1], ch. 7. Generally, if L is not to be involved in the Epimenides, 'truth-in-L' must be definable only in an L' which includes notation for a stronger logical theory (a stronger theory of classes, for example) than that available in L.[10]

Tarski's construction of truth is easily extended to other concepts of the theory of reference. It is a striking fact that these notions, despite the paradoxes which we associate with them, are so very much less foggy and mysterious than the notions

---

[9] It is sometimes overlooked that there is no need to claim, and that Tarski has not claimed, that the statements of the form (7) (or (8) or (9)) are analytic. This point has been repeatedly set right; cf. Lewy, White [1], Thomson.

[10] See Tarski [4], [5], [6]; also Quine [8]. But if L is especially weak in certain ways, this requirement lapses; witness Myhill's system, which lacks negation.

belonging to the theory of meaning. We have general paradigms (7)–(9) which, though they are not definitions, yet serve to endow 'true-in-L' and 'true-in-L of' and 'names-in-L' with every bit as much clarity, in any particular application, as is enjoyed by the particular expressions of L to which we apply them. Attribution of truth in particular to 'Snow is white', for example, is every bit as clear to us as attribution of whiteness to snow. In Tarski's technical construction, moreover, we have an explicit general routine for defining truth-in-L for individual languages L which conform to a certain standard pattern and are well specified in point of vocabulary. We have indeed no similar single definition of 'true-in-L' for variable 'L'; but what we do have suffices to endow 'true-in-L', even for variable 'L', with a high enough degree of intelligibility so that we are not likely to be averse to using the idiom. No term, of course, is definable except in other terms; and the urgency of the demand for definition is proportional to the obscurity of the term.

See how unfavorably the notion of analyticity-in-L, characteristic of the theory of meaning, compares with that of truth-in-L. For the former we have no clue comparable in value to (7). Nor have we any systematic routine for constructing definitions of 'analytic-in-L', even for the various individual choices of L; definition of 'analytic-in-L' for each L has seemed rather to be a project unto itself.[11] The most evident principle of unification, linking analyticity-in-L for one choice of L with analyticity-in-L for another choice of L, is the joint use of the syllables 'analytic'.

[11] See above, pp. 32–36.

VIII
REFERENCE AND MODALITY

1

One of the fundamental principles governing identity is that of *substitutivity*—or, as it might well be called, that of *indiscernibility of identicals*. It provides that, *given a true statement of identity, one of its two terms may be substituted for the other in any true statement and the result will be true*. It is easy to find cases contrary to this principle. For example, the statements:

(1)   Giorgione = Barbarelli,

(2)   Giorgione was so-called because of his size

are true; however, replacement of the name 'Giorgione' by the name 'Barbarelli' turns (2) into the falsehood:

Barbarelli was so-called because of his size.

Furthermore, the statements:

(3)   Cicero = Tully,

(4)   'Cicero' contains six letters

are true, but replacement of the first name by the second turns (4) false. Yet the basis of the principle of substitutivity appears quite solid; whatever can be said about the person Cicero (or Giorgione) should be equally true of the person Tully (or Barbarelli), this being the same person.

In the case of (4), this paradox resolves itself immediately. The fact is that (4) is not a statement about the person Cicero, but simply about the word 'Cicero'. The principle of substitutivity should not be extended to contexts in which the name