

Rudiments of a Theory of Reference

CHRISTOPHER S. HILL*

The last fifteen years have witnessed a vigorous and extremely interesting debate concerning two competing views about the nature of semantic reference. One of these views is perhaps best described as a counterpart of the Redundancy Theory of Truth. Its most basic component is the thesis that pairs of sentences like (1) and (2) are more or less equivalent in point of assertional content.

- (1) "Snow" refers to this stuff.
- (2) This stuff is snow.

The other view claims that our term "refers" stands for an empirically manifest relation that has causal and explanatory significance, and by implication it asserts that there is a large gulf between the content of sentences like (1) and the content of sentences like (2). Both views have been presented and compared in a number of recent writings, including influential papers and books by Devitt [1,2], Field [6], Friedman [8], Leeds [12], Putnam [16], and Soames [18].

The present paper sketches a theory of reference which I believe to have the merits of each of these views and the flaws of neither. In Section 1 I try to fix ideas by describing the two views in greater detail and by presenting their main shortcomings. Sections 2 and 3 set the stage for the theory I wish to recommend by describing two semantic concepts that are employed by an imaginary linguistic community. Both concepts are simpler in several respects than any of the concepts we actually employ, but, as it turns out, there are some striking similarities between our thought and talk about reference and thought and talk about reference in the imaginary community. Building on these similarities, Section 4 states the central hypothesis of my theory of reference and presents some supporting arguments. Finally, in Section 5 I cite several features of our concept of reference that lie beyond the scope of my central hypothesis, and I argue briefly that it may be possible to extend the hypothesis in such a way as to accommodate them.

*Over the years I have learned a great deal about reference in conversations with Ivan Fox, Anil Gupta, and Hilary Putnam. I am also indebted to Gupta for advice and encouragement concerning an earlier version of this paper. Finally, I have been helped considerably by the comments of an anonymous referee for this *Journal*.

I According to the first view, which will hereafter be called the *Redundancy Theory of Reference*, it is possible to account for the most basic features of our use of “refers” by supposing that we are guided in its use by schemata like (3) and (4).

- (3) If t exists, then “ t ” refers to an object just in case the object is identical with t .
 (4) “ P ” refers to an object just in case the object is (a) P .

As can be seen from the grammatical forms of (3) and (4), “ t ” marks a position that is accessible only to singular terms, and “ P ” marks a position that is accessible only to monadic general terms. There are similar schemata that represent sentences involving terms from a number of other grammatical categories. All such schemata have come to be called *Disquotation Schemata*.

Advocates of the Redundancy Theory claim that instances of Disquotation Schemata are partially constitutive of the concept of reference, and that coming to appreciate this fact is the most fundamental step in acquiring the concept.

These initial insights can be developed in a number of different ways. One version of the Redundancy Theory claims that there is no single purpose or semantical role that underlies all uses of “refers”. This version explains simple contexts like (1) by appealing to Disquotation Schemata, but it takes a quite different line in explaining quantificational contexts like (5).

- (5) Every word in this list refers to a large North American city.

It asserts that sentences like (5) are equivalent to sentences that have substitutional quantifiers in place of “refers”. Thus, for example, it asserts that (5) is equivalent to something like (6):

- (6) $(\forall t)$ (if “ t ” is a word in this list, then t is a large North American city)

where “ $(\forall t)$ ” is a substitutional quantifier. (It is, of course, part of this view that substitutional quantifiers are capable of binding variables that stand within quotation marks. Quantifying into quotation marks is thought to be problematic only when the quantifier is objectual.)

There is another version of the Redundancy Theory which assigns a single role to “refers”. It claims that all English discourse concerning the reference of English terms can be explained by supposing that speakers of English are guided by definitions like the following ones:

- (7) x refers₁ to y if and only if $(\exists t)$ ($x = “t”$ and $y = t$ and t exists)
 (8) x refers₂ to y if and only if $(\exists P)$ ($x = “P”$ and y is (a) P)

where “ $(\exists t)$ ” and “ $(\exists P)$ ” are substitutional quantifiers. This claim meshes nicely with the initial insights of the Redundancy Theory: it is natural to maintain that the concept of reference is based on definitions like (7) and (8) while also maintaining that instances of Disquotation Schemata are the core of the concept, for it is natural to see definitions like (7) and (8) as summaries of instances of schemata like (3) and (4).¹

This brings us to the second of the two theories that I cited at the outset. According to this view, which will hereafter be called the *Empirical Relation Theory*, the concept of reference is more or less on a par with such concepts as

the concept of valence. The concept of reference stands for a component of the natural order: generalizations about reference can have the status of laws of nature, and reference can play a role in causal explanations.

Champions of the Empirical Relation Theory usually acknowledge that they are not yet in a position to capture all of the main features of reference in a set of general principles. Rather it is their practice to cite several sample principles, and to claim that others will follow after the scientific study of language has gained further momentum. Here are several generalizations that either have been or might be cited as examples by champions of the view:

- (9) The terms in a mature science typically refer.
- (10) If n is a proper name that can be used to refer to o at a given time t , then normally o was dubbed with n in a name-giving ceremony that took place at some time prior to t .
- (11) If an individual knows that n refers to o , then normally others can use sentences containing n to secure the individual's help in projects that involve o (provided that the individual has the desire and the opportunity to help).
- (12) If the members of a group know that n refers to a place p , then normally they can use sentences containing n to arrange to meet one another at p .

Champions of the Empirical Relation Theory claim explanatory significance for generalizations of this sort, and it is clear that they are right to do so. Thus, suppose that "Ben" is the name of a dog owned by my friend Michael. Suppose further that Michael telephones me to say that Ben has run away, and to ask me to look for him in my neighborhood. Suppose finally that at the time of the call I am ready, willing, and able to come to Michael's assistance. When these facts are combined with (11), we have a (partial) explanation of my setting forth to find Michael's dog, and of my grabbing the dog by the collar when I find him trotting down my street.

Unfortunately, both of these views about reference have serious shortcomings. Advocates of the Redundancy Theory are confronted with three main difficulties:

(i) The Theory fails to provide adequately for the fact that speakers of English can make fully meaningful claims about the reference of terms in other languages. In the hope of accommodating this fact, advocates of the Redundancy Theory have frequently tried to account for the concept of reference that can be applied to expressions in other languages by explaining it in terms of the concept of translation and the concept of reference that we use in talking about English terms. Thus, it has been suggested that a sentence of the form " t refers to o in L ", where " L " stands for a foreign language, is equivalent to something like " t as a term of L can be adequately translated by an English term s such that s refers to o in English". However, this suggestion seems wrong. For one thing, the suggestion is probably circular, for it is probably necessary to employ the concept of reference in analyzing the concept of an adequate translation. Moreover, it fails to accommodate contexts of the following form: " t refers to something in L ; but I can't specify the referent since there is no equivalent term in English". It is plain that such contexts exist. For example, where L is a language spoken by theoretical physicists who are much more advanced than their English-speaking counterparts, and x is a term that belongs to the upper-

most reaches of their theory, it would be true to say that x refers to something in L that cannot be specified in English. But the suggestion implies that all such claims are false.

(ii) The theory also fails to provide for the fact that speakers of English in the actual world can make true claims about the reference of English terms in various counterfactual situations. For example, the following sentence is true: "If we used 'Kripke' in the way that we actually use 'Hesperus', then 'Kripke' would refer to a planet." Definitions like (7) and (8) wrongly imply that sentences of this sort are false.²

(iii) Although the Redundancy Theory accommodates the fact that we can make true generalizations about reference, it does not do full justice to the intuition that some of these generalizations are laws of nature. Presumably, if a generalization about reference is a law of nature, then it should apply to *all* expressions that belong to a given category. However, as is shown by objection (i), concepts based on definitions like (7) and (8) are too narrow in scope to apply to all members of a category. It follows that generalizations involving such concepts must inevitably be limited in scope.

As for the Empirical Relation Theory, its chief failing is that it is unable to account for the role that substitution instances of Disquotation Schemata (hereafter called *Disquotation Sentences*) play in guiding our use of "refers". Disquotation Sentences are epistemically more basic than other propositions about reference: we take them as fixed points in checking to see whether other propositions about reference are acceptable. (Here, of course, I mean to except such other sentences as instances of the following schema: "If t exists, then ' t ' refers to t ." This schema is virtually equivalent to (3).) In particular, we are guided by Disquotation Sentences in assessing generalizations like (9)–(12). Unfortunately, the Empirical Relation Theory cannot explain this fact. It is unable to explain how we can know that an indefinitely large number of particular propositions about reference are true *before* we come to accept a body of general principles about reference.

The situation involving Disquotation Sentences appears to be much the same as the situation involving instances of (13).

(13) "S" is true if and only if S .

Since Tarski's classic papers (cf. [19] and [20]) instances of (13) have played the role of *data* in philosophical investigations of truth. Philosophers have recognized that an analysis of truth is unacceptable unless it implies all instances of (13), and that a general principle about truth is wrong unless it is compatible with them. In a phrase, philosophers have come to recognize the *epistemic primacy* of instances of (13). A theory of truth must explain this primacy.

The point is not simply that there are facts which the Empirical Relation Theory does not explain. If the problem were only this, then there would be room for hope that it could be solved by supplementing the theory in some appropriate way. The problem is rather that the Empirical Relation Theory appears to *impede* the explanation of certain facts. If the concept of reference stands for an empirically manifest relation, then it should be possible to account for acceptance of truths involving the concept in more or less the same way as we can account for acceptance of truths involving other concepts of empirically

manifest relations. But this seems not to be the case. Thus, consider the following propositions:

If 17,391,020,651 exists, then “17,391,020,651” refers to an object just in case the object is identical with 17,391,020,651.

“Quark” refers to an object just in case the object is a quark.

I accept both of these propositions. It cannot be the case, however, that this acceptance is based directly on perception; for I came to accept them without benefit of perceptual contact with any of the extralinguistic entities with which they are concerned. Nor is my acceptance based on a prior knowledge of general principles of reference, for, as we noticed earlier, one’s acceptance of propositions of the given type is more fundamental than one’s acceptance of any general principles of reference. (As we noticed earlier, we assess general principles of reference by checking them against the class of Disquotation Sentences.)

It appears that each of our two views about reference is able to do justice to a dimension of the concept of reference that is ignored or even distorted by the other. The Redundancy Theory recognizes the epistemic primacy of Disquotation Sentences, but it fails to do justice to the interlinguistic scope and modal import of the concept of reference, and it fails to account for the causal and explanatory force of the concept. The Empirical Relation Theory acknowledges and explains the three features of reference that call the Redundancy Theory into question, but it is unable to account for the epistemic primacy of Disquotation Sentences. In view of this configuration of offsetting strengths and weaknesses, it is natural to ask whether it is possible to obtain a theory of reference that enjoys the advantages of the Redundancy Theory and the Empirical Relation Theory by taking the average of their main features. I think that the answer is “yes”, and I want to turn now to the task of arguing that this answer is correct.

2 In order to formulate the theory I wish to recommend it is helpful to appeal to an imaginary community whose members use a version of English that is much simpler than ours. I will therefore begin by discussing the users of an imaginary language called *Senglish*. Senglish is a language that is very much like ours in grammatical structure and vocabulary; however, it differs from ours in that it lacks hopelessly vague terms, indexical terms, and ambiguous terms.

Speakers of Senglish have a concept of reference that is introduced by the following definitions:

(7) x refers₁ to y if and only if $(\exists t) (x = “t” \text{ and } y = t \text{ and } t \text{ exists})$

(8) x refers₂ to y if and only if $(\exists P) (x = “P” \text{ and } y \text{ is (a) } P)$

where of course “ $(\exists t)$ ” and “ $(\exists P)$ ” count as substitutional quantifiers. Thus, this Senglish concept literally *has* the properties that are *attributed* to the English concept of reference by the Redundancy Theory.

Since (7) and (8) are components of a definition, it is appropriate to see them as necessary truths, and it is also appropriate to hold that speakers of Senglish know them to be true a priori. Moreover, the same is true of all Senglish instances of (3) and (4). This is because all such instances can be derived from (7) and (8) when (7) and (8) are supplemented with a few other truths that are both necessary and a priori.³

Now as soon as the concept defined by (7) and (8) is in hand, it is possible for speakers of Senglish to state generalizations which connect the concept to concepts that stand for various projectible properties. Many of these generalizations turn out to be true, and they are called *laws of reference*. Examples include (9)–(12), for it turns out that the laws of reference for Senglish are quite similar to the laws of reference for English. Here are several additional examples:

- (14) Normally, if n is a proper name that refers to o , then speakers who have mastered the use of n can use sentences containing n to call one another's attention to their propositional attitudes about o .
- (15) Normally, if n is a proper name that refers to o , then there is a set of sentences α such that: (i) the members of α contain n , and (ii) facts involving o are among the causally necessary conditions of speaker's acceptance of some of the members of α .⁴
- (16) Normally, if n is a proper name that refers to o , then speakers who have an interest in o make an effort to learn and remember n .⁵
- (17) Normally, if n is a proper name that refers unambiguously to o , S is a synthetic sentence that contains n , i is an individual who has mastered the use of n , and i has had considerable perceptual contact with o , then, provided that i has no other information that is germane to S (such as well-confirmed beliefs that imply it), i will consult memories of o in trying to decide whether or not to accept S .
- (18) Normally, if P is a monadic general term that unambiguously refers to the members of α , and α is a natural kind of perceptible objects, then there is a group of speakers who have the ability to recognize members of α , and other speakers tend to defer to members of this group in determining whether to accept contingent sentences that contain P .⁶
- (19) Normally, if P is a monadic general term that unambiguously refers to members of α , α is a class of theoretical entities, i is an individual who has mastered the use of P , and S is a synthetic sentence containing P , then, provided that i has no other information that is germane to S (such as well-confirmed beliefs that imply it), i will consult observable events that i believes to be caused by members of α in trying to decide whether or not to accept S .
- (20) Normally, if P is a monadic general term that refers unambiguously to the members of a class α , there are sentences containing P such that speakers' responses to those sentences vary with the distribution of certain features within the boundaries of members of α (in the way in which, e.g., responses to tokens of "That rabbit is white" vary with the distribution of whiteness within the boundaries of ostended rabbits).⁷
- (21) Normally, if P is a monadic general term that unambiguously refers to the members of α , α is a class of perceptible objects, i is a cooperative individual who has mastered the use of P , and S is Quinean occasion sentence that is obtained from P , then i will assent to a token of S accompanied by an ostensive gesture if the gesture ostends a member of α .
- (22) Normally, if P is a monadic general term, α is a class of perceptible objects, S is a Quinean occasion sentence obtained from P , and individuals are disposed to assent to a token of S when a member of α is

ostended and to dissent from a token when a member of some other class is ostended, then either: (a) P refers to the members of α , (b) P refers to the members of some other class that individuals are unable to distinguish from members of α on the basis of unsupplemented observation, or (c) individuals tend to construe pointing gestures directed at members of α as cases of deferred ostension.

There are two points about laws like (9)–(12) and (14)–(22) that should be emphasized. First, they have a variety of epistemological characteristics. While some link the concept defined by (7) and (8) to concepts that pick out forms of behavior and other observable phenomena, others link it only to concepts that are largely or even entirely theoretical. Second, the examples given thus far are only a tiny fragment of the set of all laws of reference, and there are laws quite different than any of (9)–(12) and (14)–(22). (For example, there are laws of reference that express correlations between *semantic* facts involving numerals and certain *formal* facts involving numerals.)

The laws of reference provide a foundation for a second concept of reference. After the first concept has been acquired, and the existence of laws of reference has been noted, speakers of Senglish begin to appreciate that it would be useful to have a concept of reference that would enable them to say the same kinds of things about foreign terms as the first concept enables them to say about domestic terms. Accordingly they introduce a second concept by a definition that comes to this:

- (23) A set of ordered pairs R is a reference relation for a language L in a possible world w iff the left field of R consists of terms that belong to L in w , and it is true in w that R has the projectible properties that the Senglish laws of reference attribute to the relation (i.e., to the set of ordered pairs) that is defined by (7) and (8).

Since this second concept owes its existence to laws that can only be seen to hold on the basis of experience, it can appropriately be called the a posteriori Senglish concept of reference. The first concept can appropriately be called the a priori Senglish concept.

3 These concepts have several features that should be considered before we go on to look at other matters.

First, (23) contains two terms that seem *prima facie* to carry a lot of semantic weight, namely, “define” and “attribute”. It might be held that their presence trivializes the project of defining reference, or that reference is somehow invoked by one or the other of them, and that (23) therefore introduces a circularity into the Senglish conceptual scheme. However, these worries are groundless. When the terms are used with the senses that they have in (23), they can be eliminated in favor of expressions that are ultimately reducible to primitives with no semantic significance.

To see this, notice that it is possible to obtain an equivalent definition by replacing “ R has the properties in w that the Senglish laws of reference attribute

to the relation defined by (7) and (8)" with " R satisfies the Senglish laws of reference in w under the interpretation that correlates 'refers' with R and all other terms with their normal referents". Here "normal referent" is to be seen as an expression that is explained by (7) and (8), and "satisfies" is to be seen as a term that is reducible via a Tarskian definition to a set of expressions that contains no semantic terms beyond "normal referent".

Assuming that the laws of reference are finitely axiomatizable, there is a second way of expressing the sense of (23) without using "define" and "attribute". First, speakers of Senglish would need to change the laws of reference by replacing every open sentence of the form " v_1 refers to v_2 " with an open sentence of the form " v_1 bears R to v_2 as used by speakers of language v_3 in possible world v_4 ", where " v_3 " and " v_4 " are individual variables that do not occur anywhere in the original laws of reference, and " R " is a variable that ranges over quaternary relations (considered as relations in intension). To complete this step, it would be necessary to prefix universal quantifiers that bind " v_3 " and " v_4 " to each of the laws. Second, they would have to form a single open sentence that captured the content of the sentences that resulted from the first transformation. This would be done by simply taking the conjunction of the axioms that summarize the laws of reference. When they had completed these steps, speakers of Senglish would be in a position to replace (23) with a new definition. Thus, if " $A(R)$ " was the conjunction of the axioms, they could define the a priori concept of reference as follows:

x refers to y as used by speakers of L in a possible world w iff x is a term of L in w and there exists a quaternary relation R (i.e., a relation in intension) such that $A(R)$ and x bears R to y and L in w .

In short, speakers of Senglish could obtain a set of open sentences that implicitly defined the a posteriori concept by simply replacing a term in the laws of reference with a variable that ranged over relations, and they could then convert this implicit definition into an explicit one by expressing the laws as a conjunction of axioms and binding the new variable with an existential quantifier. This would enable them to do without "define" and "attribute".⁸

This brings us to the second fact that should be considered. Although it is extremely convenient to use substitutional quantifiers in defining a priori reference, it seems that, strictly speaking, it is not necessary to do so. If Senglish had lacked substitutional quantifiers, it would have been possible to gain the effect of (7) and (8) by first defining two concepts of primitive a priori reference in terms of disjunctive lists, and by then defining two concepts of derived a priori reference by recursion from the primitive concepts. Thus, where " n_1 ," " n_2 ," . . . , and " n_m " represent all of the proper names of Senglish, and " P_1 ," " P_2 ," . . . , and " P_j " represent all of the logically simple monadic general terms, primitive a priori reference could have been defined as follows.

x primitively refers₁ to y iff either x is identical with " n_1 " and y is identical with n_1 or . . . or x is identical with " n_m " and y is identical with n_m .⁹

x primitively refers₂ to y iff either x is identical with " P_1 " and y is (a) P_1 or . . . or x is identical with " P_j " and y is (a) P_j .

The first of these biconditionals is in effect what we get when we restrict (7) to singular terms that are logically simple, and the second is what we get when we restrict (8) to simple monadic general terms. As for derived a priori reference, limitations of space and knowledge make it impossible for me to provide a detailed account of it here. However, perhaps it is possible to convey the idea I have in mind by stating several of the simplest components of a definition. Here are four such:

If x is a logically simple singular term, then x derivatively refers₁ to y iff x primitively refers₁ to y .

If x is a logically simple monadic general term, then x derivatively refers₂ to y iff x primitively refers₂ to y .

If x is a monadic general term that is obtained by inserting “or” between two other monadic terms z and w , then x derivatively refers₂ to y iff either z derivatively refers₂ to y or w derivatively refers₂ to y .

If x is a monadic general term that is obtained by prefixing “non-” to a monadic term z , then x derivatively refers₂ to y iff z does not derivatively refer₂ to y .

In addition to these clauses, an adequate definition would contain clauses for all other ways of forming logically complex referring expressions from logically simple ones. The definition would be extremely long and extremely complex, and our current ignorance of the syntactic structures of natural languages makes it impossible to envision all of its details. On the other hand, since there seem to be good reasons for holding that the class of referring expressions of a natural language is recursively specifiable, we apparently have the right to hold that it is in principle possible to construct a recursive definition of the right sort.¹⁰

Third, despite the fact that Senglish contains a number of dyadic general terms and also a number of general terms of higher adicities, neither the a priori concept of reference nor the a posteriori concept applies to such terms. However, it would be easy to fill these lacunae in Senglish semantics. The scope of the a priori concept could easily be extended by supplementing (7) and (8) with new definitions of the same type. Each new definition would bring new laws of reference in its wake.

Fourth, the a posteriori concept introduced by (23) is best seen as a concept of *primitive* reference, that is, as a concept that applies only to terms that are components of the lexicon, terms that are logically simple. Unlike logically simple terms, which cannot be said to belong to a language unless they are put to use in communication from time to time, logically complex terms are not always put to use in communication. This follows from the fact that every language contains an infinite number of logically complex terms. Further, the laws of reference fail to pin down the reference of terms that are never used. They state necessary (or nearly necessary) conditions of reference and sufficient (or nearly sufficient) conditions for many terms, but such conditions always involve features of the use of terms or features of psychological states that underlie use. Hence, they cannot be said to state conditions of reference for terms that never get used. And by the same token, since the foregoing definition is based on the

laws, it cannot provide information about what it means to say that an unused term refers to something.

If they wanted an a posteriori concept that applied to complex terms as well as to simple ones, speakers of Senglish could erect a recursive definition of a new a posteriori concept on the foundation provided by the primitive a posteriori concept. That is to say, they could construct a recursive definition with the following properties: the basis clause would be something like (23), and each of the recursion clauses would correspond to one of the standard ways of forming more complex terms from simpler ones.

Fifth, the definition of the a posteriori concept implies that the words of a language will refer to different things in different possible worlds. Thus, according to the definition, reference depends upon the way in which the words of a language are *used*. Since the terms of the language are used in different ways in different worlds, reference can vary from world to world.

Sixth, (23) implies that the laws of reference hold no less for the a posteriori concept than for the a priori concept. In other words, it implies that there are two sets of laws of reference. The sets are just alike except that one contains a predicate that expresses the former concept and the other contains a predicate that expresses the latter concept.

4 This brings us to the theory of reference that I wish to recommend. According to the central hypothesis of this theory (which is hereafter called the *Similarity Hypothesis*), speakers of English (and other actual natural languages) have two concepts of reference, and the concepts are similar in content and in logical properties to the two Senglish concepts that are discussed in Section 2.

Like many other hypotheses about portions of our conceptual scheme, the Similarity Hypothesis does not claim that the structures it postulates are explicitly represented at the level of conscious awareness. Thus, the Hypothesis is compatible with the fact that we are at best tacitly or marginally aware that we have two concepts of reference, and also with the fact that we do not explicitly formulate definitions like (7), (8), and (23) in everyday life. The Hypothesis is not to be tested by trying to gain introspective access to our semantic concepts, but rather by determining whether it provides the best explanation of the main features of the thought and talk in which our semantic concepts are manifested.

We have already taken note of four features of our thought and talk that have caused problems for other theories of reference. They are as follows: (i) we are able to make true statements about the reference of terms in other languages; (ii) we are able to make true statements about the reference of our own terms in various counterfactual situations; (iii) we are able to use generalizations about reference in giving causal explanations of psychological states and behavior; and (iv) Disquotation Sentences are epistemically more fundamental than any general principles of reference. We can test the Similarity Hypothesis by checking to see whether it provides an adequate basis for explaining these four features.

It is clear that the Hypothesis can handle (i)–(iii) quite nicely: the Hypothesis asserts that we have an a posteriori concept of reference with roughly the same content as the Senglish concept that is based on (23), and it is clear that the latter concept can be put to use in making statements about terms in lan-

guages other than Senglish, in making statements about the properties of Senglish terms in counterfactual situations, and in making statements that figure in causal explanations of Senglish psychological states and Senglish behavior.

As for (iv), recall from Section 2 that Senglish contains two sets of Disquotation Sentences that are instances of (3) and (4), and that the members of these sets follow immediately from the definitions that introduce the Senglish a priori concept of reference. Second, notice that if the Similarity Hypothesis is true, then speakers of English have an a priori concept of reference with the same properties as the Senglish a priori concept. In view of these two facts, we can see that the Similarity Hypothesis predicts that speakers of English will be able to recognize the truth of Disquotation Sentences before they recognize the truth of general principles of reference.¹¹

It appears, then, that the Similarity Hypothesis is able to explain (i)–(iv), and that it is to be preferred both to the Redundancy Theory and the Empirical Relation Theory. But is it the best explanation of the data? Is it to be preferred to all alternative accounts?

I cannot consider the whole space of alternatives here, but I would like to compare the Similarity Hypothesis to a fourth theory of reference that I call the *Tacit Knowledge Theory*. It claims that it is impossible to learn to use words unless one has the concept of reference, and that one must therefore be in possession of the concept of reference before one acquires a language. The Theory also claims that general principles like (9)–(12) and (14)–(22) are constitutive of the concept of reference. Thus, according to the theory, we must in some sense know general principles like (9)–(12) and (14)–(22) before we set out to acquire a language. The Theory glosses this claim by adding that the language learner has only *tacit* or *implicit* knowledge of general principles of reference. (Perhaps this gloss can be explained in turn by saying that the language learner relies on *innate dispositions to draw inferences* about the reference of particular words from information about the ways in which words are used, where the inferences are the same as the ones that would be drawn if the learner was consciously guided by a certain set of general laws of reference.)

It seems *prima facie* that if the Tacit Knowledge Theory is correct, then we cannot be said to know facts about the reference of particular words before we know general principles of reference. That is to say, the theory appears to call the epistemic primacy of Disquotation Sentences into question. However, it is open to advocates of the Theory to attempt to accommodate the epistemic primacy of Disquotation Sentences by saying that recognition of their truth is prior to *conscious awareness* of the laws of reference. They can claim that we are led to see the truth of Disquotation Sentences (and of other propositions about the reference of particular words) by tacit knowledge of the general principles of reference, and also that we become consciously aware of the truth of general principles by reflecting on information about the reference of particular words (such as the fact that all Disquotation Sentences are true).

There are two reasons for preferring the Similarity Hypothesis to the Tacit Knowledge Theory.

In the first place, the latter does not adequately explain the special status of Disquotation Sentences within the class of truths about the reference of particular words. Consider (24) and (25):

- (24) If Tully exists, then "Cicero" refers to x if and only if x is identical with Tully.
- (25) If Cicero exists, then "Cicero" refers to x if and only if x is identical with Cicero.

It seems obvious that (25) is epistemically more basic than (24). One can be skeptical about (24), or can at least, even if one is in fact quite sure that (24) is true, understand such skepticism. But skepticism concerning (25) would be silly: it would be a self-refuting skepticism, a skepticism that no one could take seriously. The Similarity Hypothesis acknowledges and explains this difference between (24) and (25), but the Tacit Knowledge Theory does not.

Second, it is highly questionable that one must have the concept of reference in order to learn how to use the terms of a language. There is an alternative view that can be expressed as follows: first one learns to use a term by learning the *assertibility conditions* that are associated with sentences containing the term; and second, the assertibility conditions associated with a sentence can be fully described without using the concept of reference or any of the other concepts that belong to truth conditional semantics. Several illuminating and defensible versions of this second view are presented in the literature (cf. Field [7], Harman [9] and [10]).

I find the second view to be more plausible than the Tacit Knowledge Theory. To be sure, it seems quite likely that we rely on observed correlations between utterances of words and facts involving extralinguistic entities in learning the assertibility conditions of the sentences of a language. However, the hypothesis that we move directly from information about such correlations to conclusions about assertibility conditions is much simpler than the hypothesis that we traverse two inferential paths, one leading from information about correlations to intermediate conclusions about reference, and the other leading from intermediate conclusions to final conclusions about assertibility conditions. Moreover, in many cases it seems that we must master the use of a term in order to have a concept of the entity to which the term refers. (Think, for example, of concepts that stand for numbers and other abstract objects.) In a case of this sort it is simply false that our mastery of the use of a term depends upon information about correlations between utterances of the term and facts involving an appropriate extralinguistic entity; for in a case of this sort our awareness of facts involving the entity must await our mastery of the use.

The last two paragraphs answer a question that has been lurking in the background for some time. The laws of reference in Section 2 are restricted either explicitly or implicitly to individuals who have mastered the use of Senglish terms. What is the force of this restriction? According to the position taken in the last two paragraphs, its force can be explained in terms of the concepts that belong to assertibility conditions semantics. In order to master the use of a Senglish term, it suffices to learn the assertibility conditions that are associated with Senglish sentences containing the term.

Since I have considered only three alternatives to the Similarity Hypothesis, and since my discussion of the third alternative has been rather superficial, I cannot claim to have shown that it provides the best explanation of (i)–(iv). I wish to claim only that I have provided evidence that it deserves to be taken seriously. It seems to be worthy of further attention and development.

5 Although there are a number of striking similarities between the portion of our domestic conceptual scheme that is concerned with reference and the corresponding portion of the Senglish conceptual scheme, there are five important features of the former that are not possessed by the latter. In the present section I will attempt to develop and refine the Similarity Hypothesis in such a way as to accommodate one of these five features. Moreover, while a full discussion of the four remaining features lies beyond the scope of the present paper, I will urge briefly that an advocate of the Similarity Hypothesis may well be able to deal with the other features in a satisfactory way.

We have a concept of reference with a modal dimension that is foreign to both of the Senglish concepts. There are two senses in which a term that is used by a linguistic community in a possible world w_1 can be said to refer to an object in another possible world w_2 . (i) A term that is used in w_1 can be said to refer to an object in w_2 in virtue of semantic properties that it has in w_1 . Thus, for example, it is one of the semantic properties of the English term "red" that it expresses the property *being red* in the actual world. In virtue of this fact, "red" can be said to refer to objects that have the property in other possible worlds. (ii) A term that is used in w_1 can be said to refer to an object in w_2 in virtue of semantic properties that it has in w_2 . To continue the example, since "red" is used in some other possible worlds to express the property *being blue*, it can be said to refer to objects that are blue in other worlds.

As we noticed in Section 3, definition (23) authorizes speakers of Senglish to say that their words refer to objects in other possible worlds. However, it only authorizes them to make claims of this form when "refers to objects in other worlds" is used in the second of the foregoing senses. Speakers of Senglish are unable to make such claims when "refers to objects in other worlds" is used in the first sense. In short, the Senglish definitions make no provision for *intermundane reference*.

In attempting to refine the Similarity Hypothesis so as to accommodate intermundane reference, I will first introduce a concept that can be used in discussing the intermundane reference of proper names. Let " A " stand for the set of ordered pairs $\langle x, y \rangle$ such that: (i) x is English proper name, and (ii) x refers₁ to y (where "refers₁" is understood to have the sense that it is given in definition (7) above). Further, let " B " stand for the set of all laws of reference that are satisfied by A . Using these two terms, it is possible to define a concept of intermundane reference as follows.

- (26) A set of ordered pairs R is a name-reference relation of a language L in a possible world w iff the left field of L consists of the words that are the proper names of L in w and it is true in w that R has the projectible properties that the members of B attribute to the relation A .
- (27) x name-refers to y at world w_2 as used by speakers of L in world w_1 iff the pair $\langle x, y \rangle$ is in the name-reference relation of L in w_1 and y exists in w_2 .

Of course (27) presupposes the view that a proper name refers to the same object in all possible worlds. If this view should turn out to be wrong, it would be necessary to account for the intermundane reference of names in some other way.

The task of accounting for the intermundane reference of monadic general

terms is a bit more complex. It has several parts. (i) It is necessary to begin by explaining what it means to say that an English monadic general term *connotes a property*. Let us suppose that the operator “being an object such that” can be used to convert monadic open sentences into names of properties. This assumption authorizes us to construct the following definition:

An English monadic open sentence x connotes the property y if and only if $(\exists S)$ (x is identical with “ S ” and y is identical with *being an object such that S*)

where “ $(\exists S)$ ” is a substitutional quantifier. Once it has been explained what it is for an English monadic open sentence to connote a property, it is possible to explain what it is for an English monadic general term to connote a property. The definition runs as follows:

(28) An English monadic general term x connotes the property y if and only if there is a monadic open sentence S such that: (a) x is a general term in the predicate of S , (b) there is no general term in the predicate of S of which x is a proper part, and (c) S connotes y .

(ii) The next step is to bring a certain fact into focus, namely, the fact that there is a set of laws that link the concept defined by (28) with other concepts. The members of this set (hereafter called “ C ”) are similar in many respects to the laws of reference. Here are two examples:

Normally, if it is true (a) that x is a monadic general term, (b) that y is a property, (c) that individuals who have mastered the use of x are disposed to use sentences containing x when they wish to induce others to take a certain attitude or to behave in a certain way toward things that they believe to have y , and (d) that they would also be so disposed in possible situations in which y is not believed to be coextensive with the properties with which it is believed to be coextensive in the actual world, then x connotes y .

Normally, if x is a monadic general term that unambiguously connotes y , y is a theoretical property, i is an individual who has mastered the use of x , and S is a synthetic sentence containing x , then, provided that i has no other information that is germane to S (such as well-confirmed beliefs that imply it), i will consult observable events that i believes to be caused by instances of y in trying to determine whether or not to accept S . Moreover, i would also consult such events in possible situations in which y is not believed to be coextensive with the properties with which it is believed to be coextensive in the actual world.

(iii) The third step is to explain what it means to say that a property is connoted by a monadic general term in some language other than English or by an English monadic general term in some possible world other than the actual world. It is possible to explain statements of these two kinds by the following definition.

(29) A set of ordered pairs R is the connotation relation of a language L in a possible world w iff the left field of R consists of the words that are the

monadic general terms of L in w and R has the projectible properties in w that the laws belonging to C attribute to the relation D .

(Here “ D ” is a term for the relation picked out by (28).)

(iv) Finally, it is necessary to state a definition like (30).

- (30) x refers₂ to y at world w_2 as used by speakers of L in world w_1 iff there exists a property z such that: (a) the pair $\langle x, z \rangle$ is in the connotation relation of L at w_1 , and (b) y has z in w_2 .

(30) gives us the desired concept of intermundane reference for monadic general terms.

In order to obtain a concept of intermundane reference that applies to descriptions, one must provide a series of recursive definitions that are based on concepts of intermundane reference for more basic linguistic categories. Thus, for example, before one can obtain a concept that applies to descriptions, one must supplement the foregoing definitions with a definition that applies to dyadic general terms.

We are now in a position to consider a version of the Similarity Hypothesis that is more precise and has greater explanatory power than the version with which we started. According to this version, we have two concepts of reference. One, which may be called the *a priori concept*, is based on (7) and also on the following definition.

- (31) x refers₂ to y iff x is a monadic general term and there exists a property z such that x connotes z and z is instantiated by y .

(The sense of “connotes” that is relevant to (31) is the sense that is introduced by (28).) Second, there is a concept (hereafter called the *a posteriori concept*) that is based largely on laws that involve the *a priori* concept of reference and the *a priori* concept of connotation. In other words, a large part of its content is given by definitions (27) and (30).

At the beginning of this section I mentioned that the portion of our domestic conceptual scheme that is concerned with reference has five important features that are not shared by the corresponding portion of the Senglish conceptual scheme. We have found that it is possible to account for one of these features by changing and supplementing our first version of the Similarity Hypothesis. It is time now to consider the four remaining features.

In the first place, unlike speakers of Senglish, we make use of indexical expressions and are able to ascribe reference to them. What are the features of our concept of reference that support such ascriptions?

It is not my intention to offer an answer to this question that is in any sense final or complete. I wish only to point out that it is possible for an advocate of the Similarity Hypothesis to supplement the Hypothesis with principles like these:

- (32) A token of “I” refers to i iff i is the individual who produced the token.
 (33) A token of “you” refers to i iff i is the intended audience of the individual who produced the token.

- (34) A token of “here” refers to p iff p is the place at which the token was produced.
- (35) A token of an expression consisting of “this” followed by a monadic general term refers to o iff o is an object such that: (a) the monadic term refers to o , and (b) o is made especially salient by the previous discourse, or by an indexical gesture, or by o 's relationships to other entities within the nonlinguistic context.

These principles are at best a fragment of a theory of indexicals. A full treatment would require additional principles concerning English indexicals, principles concerning foreign indexicals, and also a detailed account of the various forms of salience that are presupposed by principles like (35). It is fairly clear, however, that it would be possible to extend the list of principles without encountering any substantial difficulties. Further, there is no reason to think that the project of analyzing salience would require concepts or principles that are unavailable to an advocate of the Similarity Hypothesis.

Second, we differ from speakers of Senglish in that we are able to ascribe reference to singular terms that are in some sense vacuous. Even though there is no such thing as Achilles, we can say, truly, that “Achilles” refers to Achilles. How is this possible? Further, how can it be possible to say that “Achilles” refers to the slayer of Hector, or that “Αχιλλεὺς” refers in Greek to Achilles?

Here I wish to claim only that an advocate of the Similarity Hypothesis is not *ex officio* excluded from developing a position that has some plausibility. An advocate can, for example, avail himself or herself of the following view: if S is a true sentence containing a singular term for an imaginary object, then, normally, it is appropriate to see S as an abbreviation of a sentence like “It is true in fiction that S ” or “It is true in the world of make believe that S ”. This view, which has of course enjoyed considerable popularity among philosophers, encourages us to hold, among other things, that “‘Achilles’ refers to Achilles” is an abbreviation of “It is true in the world of Homer that ‘Achilles’ refers to Achilles”.

It is possible to apply the view to sentences that ascribe reference to vacuous terms without insisting on a uniform interpretation of the occurrences of “refers” that are found in such sentences. If the term to which reference is ascribed belongs to a foreign vocabulary, then “refers” must be seen as expressing our a posteriori concept of reference. On the other hand, if it belongs to our domestic vocabulary, then “refers” may be seen as expressing either our a posteriori concept or our a priori concept.

I do not wish to suggest that the “invisible operator” approach is the only way or even the best way of bringing the Similarity Hypothesis into line with the fact that we can ascribe reference to terms that are in some sense vacuous. There are several interesting alternatives, and one may turn out to be preferable to the invisible operator approach.

Third, we differ from speakers of Senglish in that we are able to ascribe reference to terms with semantic content, that is, to terms that contain “refers” and other semantic expressions. For example, in English it is fully meaningful to say “The referent of “George Washington”’ refers to George Washington”. It would be easy to supplement the Similarity Hypothesis in such a way as to

accommodate a fairly sizeable portion of our talk about the reference of terms with semantic content. However, our talk of this kind is extremely complicated, and it would be quite difficult to capture *all* of its intricacies and convolutions in a set of definitions. Thus, for example, a theory of reference for terms with semantic content must come to grips with all of the paradoxes of reference (e.g., the Berry Paradox). Perhaps a theory should also suggest ways of solving the problems that are posed by the other semantic paradoxes.

The options that are available at this point to an advocate of the Similarity Hypothesis are quite different than the options that are available to advocates of other pictures of reference. However, I see no reason to think that they are less promising. We are not in a position today to conclude that a theory of reference is wrong because it conflicts with a theory of the paradoxes that we know to be correct. Rather our position is just the opposite. It seems that our best hope of understanding the paradoxes is to examine theories that purport to capture the point or the content of the concept of reference.¹²

Fourth, we have thus far failed to come to grips with referential ambiguity. A proper name can refer to two or more individuals, and a monadic general term can have several disjoint extensions. The foregoing definitions fail to take these facts into account.

Fortunately, while referential ambiguity shows that we do not yet have an adequate version of the Similarity Hypothesis, it does not necessarily constitute an insurmountable obstacle. Thus, for example, it may well be true that if a term is characterized by an n -fold ambiguity at the level of reference, then it is also characterized by an n -fold ambiguity at the level of assertibility conditions (in the sense that some contingent sentences containing the term are governed by n distinct sets of assertibility conditions). If this proposition *is* true, then it is possible to accommodate referential ambiguity by relativizing reference to sets of assertibility conditions. Equally, we can accommodate it by maintaining that it is not a term taken alone that has reference, but rather a term *qua* indexed by a set of assertibility conditions.

I have already mentioned several reasons for thinking that the theory of reference presented in these pages is worthy of further attention. I would like to conclude by mentioning one additional reason, namely, that the theory seems to be philosophically fruitful. It has implications concerning a number of important problems in the philosophy of language, and these implications seem on the whole to agree with fairly vivid intuitions. (For example, it provides a basis for responding to certain counterintuitive skeptical views about the uniqueness of schemes of reference.¹⁴)

NOTES

1. Two comments about the relationship between instances of (3) and definition (7):
 - (i) It is not entirely clear that (7) is the best way of summarizing the sentences that are represented by (3). Thus, it entails that a sentence of the form “‘ t ’ refers to t ” is *false* unless “ t exists” is true, and this proposition might well be questioned. (It might be held that a sentence of the form “‘ t ’ refers to t ” is *neither true nor false* if “ t exists” is false.)

An advocate of the Redundancy Theory who views this consequence as unwelcome can avoid it by replacing (7) with the following simpler definition.

(7') x refers₁ to y iff $(\exists t) (x = "t"$ and $y = t)$.

This alternative suggestion leads to problems unless it is supposed that rules for manipulating quantifiers in English are governed by restrictions like the ones that are familiar from systems of Free Logic. But there is plenty of independent evidence that this assumption is correct.

Except for a few paragraphs at the very end of the paper, I will not discuss questions concerning nondesignating terms. Although (7) turns out to play a fairly important role in the paper, (7') would have served my main purposes equally well. Questions about the comparative merits of (7) and (7') are irrelevant to my main concerns.

(ii) It might seem that (7) should be replaced by the following definition:

(7'') $(\forall t)$ (If t exists, then " t " refers to an object iff the object is identical with t)

where " $(\forall t)$ " is a substitutional quantifier. (7'') is simpler than (7), and its relationship to instances of (3) is more straightforward.

The reason for preferring (7) is that (7'') fails to give sense to such contexts of "refers" as " x refers to y " and " a refers to y " (where " a " picks out a term by describing it rather than quoting it). It gives sense only to contexts of the form " t refers to y ".

2. To see that (7) implies that the given counterfactual is false, notice first that the counterfactual cannot be (nonvacuously) true unless the sentence "Kripke' refers to Hesperus" is true at some possible world w . Second, observe that (7) implies that "Kripke' refers to Hesperus" is true at w if and only if some sentence of the form "Kripke' is identical with ' t ' and t is identical with Hesperus and t exists" is true at w . And finally, observe that the first conjunct of a sentence of this form is true at w if and only if the second conjunct is false at w .
3. The additional truths are identity axioms of the standard sort and the following two principles:

For all t_1 and for all t_2 , if " t_1 " = " t_2 ", then for every object y , $y = t_1$ if and only if $y = t_2$.

For all P_1 and for all P_2 , if " P_1 " = " P_2 ", then for every object y , y is (a) P_1 if and only if y is (a) P_2 .

Here all four of the initially placed quantifiers are substitutional, and the two internal quantifiers are objectual.

4. See Evans [3].
5. See Evans [5], p. 379.
6. See Putnam [15].
7. See Evans [4].
8. Here I am indebted to Lewis [13].
9. This construction derives ultimately from Tarski, but I am also indebted to a paper by Field. See Tarski [20] and Field [6].
10. The main ideas in the second half of the paragraph derive from Putnam's first John Locke Lecture. See [16], pp. 9–17.

11. There is a curious feature of the Similarity Hypothesis that may be worth pointing out: it provides *two* explanations of the epistemic primacy of Disquotation Sentences.

If the Similarity Hypothesis is correct, then Disquotation Sentences are ambiguous. This is because Disquotation Sentences contain the term “refers,” and because the Similarity Hypothesis implies that we have two different concepts that can be expressed by this term.

According to the argument given above, the Hypothesis predicts that Disquotation Sentences will enjoy epistemic primacy when “refers” is used to express the a priori concept of reference. But it also predicts that Disquotation Sentences will enjoy epistemic primacy when “refers” is used to express a posteriori reference. This follows from two facts: first, Disquotation Sentences are implied by the English counterpart of (23) no less than by the English counterparts of (7) and (8); and second, in order to adopt a definition like (23), it is not necessary to know any particular laws of reference (one need only have reason to believe that it will eventually be possible to find *some* laws).

12. Although the range of options associated with the Similarity Hypothesis is in some respects narrower than the ranges associated with other pictures of reference, it is wider in certain other respects. This can be illustrated by a passage in a familiar paper on substitutional quantification by Marcus (see [14]). Marcus cited a definition of a truth predicate that can be expressed as follows:

x is true iff for some S , x is identical with “ S ” and S .

In discussing this definition, Marcus pointed out that it is necessary to impose certain restrictions on its quantifier (which is of course a substitutional quantifier) in order to satisfy the requirements of adequacy for definitions. Further, she in effect suggested that if the truth predicate we use in English was based on a definition of this sort, then the paradoxes of truth could be seen as a side effect of unwitting violations of these restrictions. In view of the evident similarities between Marcus’s definition and several of the key components of the Similarity Hypothesis, there is reason to think that this second observation could be combined with the Similarity Hypothesis to obtain a partial solution to the paradoxes of reference.

14. One can obtain a fairly strong objection to such views by combining the ideas in the present paper with the ideas in Hill [11].

REFERENCES

- [1] Devitt, M., *Designation*, Columbia University Press, New York, 1981.
- [2] Devitt, M., *Realism and Truth*, Princeton University Press, Princeton, New Jersey, 1981.
- [3] Evans, G., “The causal theory of names,” *Aristotelian Society*, Supplementary Volume, vol. 47 (1973), pp. 187–208.
- [4] Evans, G., “Identity and predication,” *The Journal of Philosophy*, vol. 72 (1975), pp. 343–363.
- [5] Evans, G., *The Varieties of Reference*, Oxford University Press, New York, 1982.
- [6] Field, H., “Tarski’s theory of truth,” *The Journal of Philosophy*, vol. 69 (1972), pp. 347–375.

- [7] Field, H., "Logic, meaning, and conceptual role," *The Journal of Philosophy*, vol. 74 (1977), pp. 379-408.
- [8] Friedman, M., "Truth and confirmation," *The Journal of Philosophy*, vol. 76 (1979), pp. 361-382.
- [9] Harman, G., "Meaning and semantics," pp. 1-16 in *Semantics and Philosophy*, eds., M. Munitz and P. Unger, New York University Press, New York, 1972.
- [10] Harman, G., "Conceptual role semantics," *Notre Dame Journal of Formal Logic*, vol. 23 (1982), pp. 242-257.
- [11] Hill, C., "Animadversions on the inscrutability thesis," *Pacific Philosophical Quarterly*, vol. 66 (1985), pp. 304-313.
- [12] Leeds, S., "Theories of reference and truth," *Erkenntnis*, vol. 13 (1978), pp. 111-129.
- [13] Lewis, D., "How to define theoretical terms," pp. 78-95 in his *Philosophical Papers*, vol. 1, Oxford University Press, New York, 1983.
- [14] Marcus, R., "Quantification and ontology," *Nous*, vol. 6 (1972), pp. 240-250.
- [15] Putnam, H., "The meaning of meaning," in his *Language and Reality* (Collected Papers, vol. 2), Cambridge University Press, Cambridge, 1978.
- [16] Putnam, H., *Meaning and the Moral Sciences*, Routledge Kegan Paul, Boston, Massachusetts, 1978.
- [17] Putnam, H., *Reason, Truth and History*, Cambridge University Press, Cambridge, 1981.
- [18] Soames, S., "What is a theory of truth?," *The Journal of Philosophy*, vol. 81 (1984), pp. 411-429.
- [19] Tarski, A., "The semantic conception of truth," pp. 52-84 in *Readings in Philosophical Analysis*, eds., H. Feigl and W. Sellars, Appleton Century Crofts, New York, 1949.
- [20] Tarski, A., "The concept of truth in formalized languages," pp. 152-278 in his *Logic, Semantics, and Metamathematics*, Oxford University Press, New York, 1956.

Department of Philosophy
University of Arkansas
Fayetteville, Arkansas 72701