

## Causality and the Paradox of Names

MICHAEL MCKINSEY

In the literature concerning the semantics of proper names, almost no attention has been paid to methodological questions, such as: What should be the formal structure of theories of reference for the various types of singular terms including names? What should be the subject matter of such theories, and what kinds of facts are relevant to determining their truth-values? In this paper, I wish to suggest some answers to these questions, answers that I believe provide an accurate conception of the form that a theory of names should take. Such a conception is valuable because it constitutes a constraint that any correct theory of names must satisfy, and so it provides a tool for evaluating particular theoretical proposals concerning names. Using this tool, I will argue that no *causal* theory of names can be correct and that the true theory must be a kind of description theory. The fact that the true theory of names takes the form it does has some surprising consequences regarding the semantic structure of natural languages and the concept of meaningfulness.<sup>1</sup>

### 1. THE REFERENTIAL PARADIGM

I will begin by considering *indexical* singular terms, since such terms provide fairly simple and clear models of what a semantic theory for a type of singular term should be about and what such a theory should look like. The semantic referent of an indexical term typically varies from context to context, even if the term has a single linguistic meaning. Consider the first person pronoun 'I', for instance. According to English convention, 'I' refers on a given occasion to whoever is its speaker on that occasion. So it is plausible to suppose that the word 'I' is governed in English by the rule:

- (1) For any  $\alpha$ , if  $\alpha$  is a token of 'I', then for any object  $x$ ,  $\alpha$  is to refer to  $x$  if and only if  $x$  is the speaker of  $\alpha$ .

Now the fact that a given utterance of 'I' in English refers to a given object would seem to be determined by the facts (a) that the speaker is in some sense following or invoking the correct rule for 'I' in English and (b) that the object in question satisfies the reference-condition contained in that rule.<sup>2</sup> In this way, the rule (1) determines the referent of 'I' relative to any given context. Moreover, it is plausible to suppose that the linguistic meaning of 'I' in English is completely specified by the fact that English contains the rule (1). In this manner, then, *the linguistic meaning of a term determines its referent*.

This doctrine contrasts with Frege's principle that the referent of a term is determined by its *sense* (Sinn).<sup>3</sup> On Frege's concept of sense, the proposition expressed by a sentence is functionally determined by the senses of its parts. As Frege knew, it follows from this fact that the sense of an indexical like 'I' varies from context to context because the propositional expressed by a single sentence containing the indexical varies from context to context.<sup>4</sup> But the linguistic meaning of an indexical does not vary in this way. So the linguistic meaning of such a term is not its sense. Rather, as David Kaplan has suggested, a term's linguistic meaning must be something that, together with a context, *determines* the term's sense in that context.<sup>5</sup> Since the sense of an indexical is determined by its linguistic meaning, and since the linguistic meaning of an indexical would seem to be given by a rule like (1) that determines reference, the most plausible view, contrary to Frege, is that the sense of an indexical in a context is *identical with* its referent in that context.

To say that a given term's sense in a context is identical with its referent in that context is to say that the proposition expressed in the context by a sentence containing the term is a function of the term's referent in the context. Terms of this sort I call "genuine terms" and propositions expressed in contexts by use of such terms I call "singular propositions." The example of indexicals like 'I' provides a useful paradigm of a semantic theory for a genuine term. To give a semantic theory for such a term amounts to stating the rule of reference analogous to (1) that speakers follow in using the term. I call this "the referential paradigm."

Now among genuine terms I count proper names. My reason for doing so is the same as the reason Kripke has given for thinking that names are "rigid designators," or terms that refer to the same object at every possible world.<sup>6</sup> Consider a particular use of any sentence containing an ordinary name, such as:

(2) Ben Franklin was bald.

Suppose that in our use of (2), the referent of 'Ben Franklin' is a certain man, say, the American patriot of that name. Then this use of (2) is true in the actual world just in case this man was bald. Kripke's point is that such a use of (2) is also true in any *other* possible situation if and only if *that very man* was bald in that situation. We might say that in any possible world in which our use of (2) is true, the state of affairs that makes it true is just that of *x's having been bald*, where *x* is the actual referent of the token of 'Ben Franklin' in question.<sup>7</sup>

Perhaps this consideration does not *prove* that the proposition expressed in a given context by (2) is a function of the referent of 'Ben Franklin' in that context.

Still, if one is casting about for likely candidates to play the role of the sense of 'Ben Franklin' in a particular use of (2), there seems to be no *other* choice that is consistent with that use's having the possible-world truth conditions that Kripke describes. For instance, as Kripke points out, a particular use of 'Ben Franklin' would not have the sense of any contingent *definite description* that Ben Franklin actually satisfies.<sup>8</sup> So I will take it that the proposition expressed in a context by (2) is functionally determined by the referent of 'Ben Franklin' in that context. And since similar considerations apply to all proper names, I will take all names to be genuine terms in the sense introduced above.<sup>9</sup>

Since names are genuine terms, they fall under the referential paradigm. However, in giving a theory of names, we are not interested in giving a theory of any *particular* name. So we are not interested in stating any particular semantic rules that speakers follow in using names. Instead, what is wanted in a theory of names is an adequate generalization concerning the *sort* of reference rule that speakers follow in using words as names.

Suppose that a certain language *L* contains the following reference rule concerning the word *N*:

- (3) For any  $\alpha$ , if  $\alpha$  is a token of *N*, then for any object *x*,  $\alpha$  is to refer to *x* if and only if *x* is the inventor of bifocals.

Would the fact that *L* contains this rule imply that tokens of *N* have the same *sense* in *L* as the English definite description 'the inventor of bifocals'? No, it would not. The presence of the rule (3) in *L* guarantees at most that tokens of *N* *have the same referent* as 'the inventor of bifocals' in fact has, since this is all that the rule (3) requires for its satisfaction. But the senses of such definite descriptions are notoriously *not* the same as their referents, and so the fact that *L* contains (3) does not imply that tokens of *N* have the same sense as 'the inventor of bifocals',

Kripke has suggested that a name's referent could in principle be "fixed" by use of a definite description, without thereby giving the name the *sense* of that description.<sup>10</sup> The hypothesis that languages might contain reference rules of the form (3) provides a clear and simple way of understanding the sort of possibility that Kripke has described. Whether or not any natural languages actually *do* contain such rules as (3) is an issue we shall discuss at length below.

In semantic treatments of modal languages containing singular terms, it is common to relativize the relation of reference to a possible world.<sup>11</sup> This is done primarily to handle definite descriptions, since to correctly state the possible-world truth conditions for sentences containing descriptions, we must allow the descriptions' referents to vary from world to world. Reference rules for genuine terms can be stated by use of a relativized reference relation, provided that in our statements we guarantee that the terms rigidly designate the same object at every possible world. For instance, letting ' $w^*$ ' represent the actual world, (3) would become:

- (3\*) For any  $\alpha$ , if  $\alpha$  is a token of *N*, then for any object *x* and possible world *w*,  $\alpha$  is to refer to *x* at *w* if and only if *x* exists in *w* and *x* is the inventor of bifocals in  $w^*$ .

However, I prefer to stick to the use of an unrelativized reference relation for genuine terms and to evaluate sentences containing such terms at different possible worlds simply on the basis of what, if anything, the terms in fact refer to. Since a genuine term's referent will only be an object that in fact exists, and since a term will have at most one referent, we get the same effect that we would get by requiring the term to refer to the same thing at every possible world.

We may if we wish keep a relativized reference relation for use in our treatment of nongenuine terms like definite descriptions. But in any case, it is clear that we will have to provide a semantic treatment for such terms that is quite different from the one just suggested for genuine terms. So I will assume that only genuine terms fall under reference rules that are formulated by use of the unrelativized reference relation.<sup>12</sup>

One advantage of using an unrelativized reference relation to formulate reference rules for genuine terms is that our formulations are simpler and more likely to express accurately the kind of rules that speakers of natural languages actually follow. For if we stated reference rules by use of a relativized reference relation, our statement would presuppose that the reference rules that people actually follow involve quantification over possible worlds, and this is a controversial assumption at best.

## 2. THE PROBLEM OF SEMANTIC AMBIGUITY

As I have characterized them so far, reference rules have the following general form:

- (4) For any  $\alpha$ , if  $\alpha$  is a token of  $W$ , then for any object  $x$ ,  $\alpha$  is to refer to  $x$  if and only if  $\varphi$ ,

where  $W$  refers to a word-type and  $\varphi$  is a formula containing  $x$  and perhaps  $\alpha$  as free variables.<sup>13</sup> But the phenomenon of semantic ambiguity raises a serious problem for the idea that reference rules have this form.

On one widely held view, *demonstrative* indexicals all have in common the feature that their tokens' semantic referents are determined either wholly or in part by what their *speakers* refer to, or demonstrate, by use of the tokens.<sup>14</sup> Given this view of demonstratives plus the idea that reference rules have the form (4), it is plausible to suppose that the feminine demonstrative 'she' is governed in English by the following rule:

- (5) For any  $\alpha$ , if  $\alpha$  is a token of 'she', then for any  $x$ ,  $\alpha$  is to refer to  $x$  if and only if  $x$  is the unique female to whom the speaker of  $\alpha$  refers with  $\alpha$ .

Now 'she' is also the name of a novel by H. Rider Haggard. So suppose on a given occasion I say out loud, "*She* is Haggard's most exciting book." Then my token refers to Haggard's novel, a genderless object. So my utterance of 'she' *violates* the rule (5), even though my utterance is in *English*. But if this is true, can (5) really be a rule of English? In my view, it can.

A term might have two meanings in a language, and in both meanings be a genuine term. There would then be two reference rules of the form (4) in the

language, both of which govern the same term. These rules will contain nonequivalent reference-conditions, and so in a sense the rules will be "inconsistent," for they can yield contradictory results when applied to the same token. But how is this possible? I call this "the problem of semantic ambiguity."

In my view, we should just accept as fact the ideas that natural languages contain conflicting rules and that we can follow a semantic rule in full awareness that the rule we are following is inconsistent with another rule of our language. After all, we often do this sort of thing in other spheres of activity. Thus, a piece of wood having a certain shape might be used in two different games. Each game might require the piece to be moved in ways that conflict with what the other game requires. In the context of each game, we find it easy to think of our moves with the piece as subject to a certain rule that we know is violated in the other game.

Similarly, when I use 'she' as a feminine demonstrative, I am thinking of my utterance as subject to the rule (5), even though I know that I could use 'she' as a name and violate this rule. In fact, I suggest, thinking of my utterance as subject to the rule (5)—or, in other words, following this rule—is precisely what using 'she' as a feminine demonstrative amounts to.

These facts strongly suggest that in many spheres of activity, including both playing games and speaking a language, we "follow the rules" of an activity in a given context by letting certain of these rules guide our behavior in that context, rather than by actualizing a permanent disposition to obey the rules in every context.

But the following objection might be raised. On my view, a game such as chess contains rules like:

(6) A bishop is to be moved only on the diagonal.

And also on my view, we could easily invent another game perhaps much like chess but containing rules that conflict with (6). However, it might be objected that the rules of chess don't really look like (6), because these rules are really of the form:

(7) *When playing chess*, a bishop is to be moved only on the diagonal.

Similarly, it might be said, the rule for 'she' is not (5), but instead is a rule whose antecedent is 'if  $\alpha$  is a token of "she" *that is used as a feminine demonstrative*'. On this suggestion, the rules of chess could not conflict with the rules of any other game, and the rule for using 'she' as a feminine demonstrative could not conflict with any other rule of English.

Now if the sentences that express the rules of chess all contain the phrase 'when playing chess' as a prefix, then every rule of chess mentions the game of chess itself. But this implies that chess cannot be defined in terms of its rules. For obviously any definition of chess as a game consisting of such-and-such rules, each of which mentions chess, would be viciously circular. But I think it is clear that chess, like any other game, *can* be defined in terms of its rules. Therefore, the rules of chess cannot be expressed by sentences like (7) that contain the phrase 'When playing chess' as a prefix.

Statement (7) does not express a rule of chess. On the one hand, it can be understood as stating an analytic truth *about* the rules of chess, namely:

- (8) It is a rule of playing chess that a bishop is to be moved only on the diagonal.

Understood as meaning (8), (7) correctly states *that* a certain rule—namely (6)—is a rule of chess. But it would be a fallacy to infer from this that (7) *itself* states a separate rule of chess. On the other hand (7) could be understood to mean:

- (9) When one is playing chess, one should move a bishop only on the diagonal.

But (9) is not a rule of chess either. It is instead most plausibly construed as a “hypothetical imperative” to the effect that if one wants to play chess correctly (that is, obey its rules), one should obey rule (6). On both ways of understanding (7), it is (6) and not (7) that expresses a rule of chess.

So there seems to be no plausible way of avoiding the fact that the rules of the different games we play can conflict, and hence there seems to be no inherent difficulty in my suggestion that the semantic rules of a language may also conflict. Moreover, there is a good reason for thinking that the semantic rules of English do in fact conflict. For to use ‘she’ as a feminine demonstrative is surely the same as using ‘she’ *with a certain linguistic meaning*, and this in turn is the same as the speaker’s following a certain semantic rule in using ‘she’. But then the sort of rule in question cannot itself involve the concept of a feminine demonstrative, and so the rule’s antecedent cannot be expressed by ‘if  $\alpha$  is a token of “she” that is used as a feminine demonstrative’. The only alternative would seem to be that the relevant rule must like (5) be simply a rule for using ‘she’, and if this is so, then the rule conflicts with other rules of English. Again, we should just accept as fact the idea that natural languages may contain conflicting semantic rules.

However, there is still a serious problem connected with this idea that we have not yet faced. Let us say that a rule  $R$  is an *inviolable* rule of a language  $L$  if and only if any utterance that violates  $R$  cannot be an utterance that is in  $L$ . Thus (5) as well as other reference rules of the form (4) and rules for ambiguous words in general cannot be inviolable rules of their languages. But then the question arises: What makes a rule of this kind a rule of a language at all?

A language, I suggest, may be identified with a finite set of syntactic and semantic rules. Intuitively, the rules in such a set are the rules that define the language in question. Let us call a rule  $R$  a *basic* rule of  $L$  if and only if  $R$  is a member of the finite set of rules that defines or comprises  $L$ . I would speculate that the inviolable rules of any language  $L$  are just those rules that are either basic rules of  $L$  or logical consequences of basic rules of  $L$ .

The rules of English permit ‘she’ to be used as a feminine demonstrative and prohibit other words from being so used. But again, to use ‘she’ as a feminine demonstrative is just to follow rule (5). This suggests that at least some of the rules of English must be *second-order* rules that permit the following of certain rules and prohibit the following of others. The rules granting permission to follow a rule like (5) would be “second-order” in the sense that they would be rules about rules. So let us say that a rule  $R$  is a *nonbasic* rule of a language  $L$  if and only if:  $R$  is not a

basic rule of  $L$ , but there is a rule  $Q$  such that  $Q$  is a second-order basic rule of  $L$ , and  $Q$  explicitly permits speakers to follow  $R$ .

I propose that (5) is a rule of English in the sense that it is a nonbasic rule of English. Similarly, any rule of a language that can be violated by an utterance in that language is, I suggest, either a nonbasic rule of the language or a logical consequence of such a rule. Below, when I speak simply of a "rule of" a language  $L$ , I shall mean a basic or nonbasic rule of  $L$ , or a logical consequence of a basic or nonbasic rule of  $L$ .

Reference rules like (5) provide examples of first-order, nonbasic semantic rules. But what do second-order semantic rules look like, and in what sense do they "explicitly permit" the following of some first-order rules and forbid the following of others? These are difficult questions that I cannot really do justice to here, but I will make a brief suggestion. The second-order rule of English permitting 'she' to be used as a feminine demonstrative would, I suggest, grant this permission in the sense that it would explicitly allow tokens of 'she' that are used in this way to have semantic referents. Roughly, the relevant second-order rule would be:

- (10) For any  $\alpha$ , if  $\alpha$  is a token of 'she' that is used as a feminine demonstrative, then the referent of  $\alpha$  is to be determined in accordance with the first-order rule being followed.

Since a token of 'she' is used as a feminine demonstrative if and only if its speaker is following rule (5), (10) says that tokens of 'she' that are produced by speakers following rule (5) may have their referents determined accordingly.

The general picture I would suggest goes roughly as follows. For each genuine term of English, there is at least one second-order rule like (10) permitting tokens of the term to have referents in accordance with a certain first-order reference rule of the form (4). To each such first-order rule for a term that is mentioned in the second-order rule of English, there corresponds *one* of the term's linguistic meanings in English. The fact that more than one first-order rule for a given term may be mentioned in the second-order rules of English allows the picture to account for semantic ambiguity. The list of second-order rules for genuine terms will of course be finite, and there will in addition be a rule to the effect that a token that is not used in any of the ways mentioned in the list has no referent.<sup>15</sup>

If I am right, the existence of semantic ambiguity in natural languages is best taken account of by a view according to which these languages contain nonbasic semantic rules. By definition, any such language also contains basic second-order rules like (10). So the existence of semantic ambiguity in natural languages provides good evidence that these languages contain second-order semantic rules. As we shall see below, the fact that there are second-order rules in natural languages provides a significant key to understanding names.

I said earlier that a theory of names should describe the kind of reference rule we follow when using words as names. Our discussion of 'she' strongly suggests that whether or not a word is used as a name depends upon the sort of rule the speaker is following. This gives us another way of describing the goal of a theory of names.

By providing an adequate generalization concerning the sort of reference rule that speakers follow when using words as names, an adequate theory of names would tell us *what it is for a word to be used as a name*.

Having described what the goal of a theory of names should be, I now wish to consider whether any form of causal theory of names is capable of reaching this goal.

### 3. CAUSAL THEORIES OF HAVING AN OBJECT IN MIND

One of the main motivations behind a certain kind of causal theory of names results from combining two ideas. The first idea is that the semantic referent of a particular name-token or utterance is in part determined by which object its speaker *has in mind* or *means* by the token. The second idea is that whether or not a speaker has a given object in mind in using a name is determined by whether or not the speaker's mental states are connected to that object by the right sort of causal relation. A view that results from combining these two ideas is best looked upon as a causal theory of having an object in mind that has been applied to the semantics of proper names.<sup>16</sup>

A causal theory of having an object in mind is a causal theory of what makes our mental states *about* or *of* certain objects and not others. I have criticized such theories elsewhere in some detail.<sup>17</sup> But here I wish to point out that even if a causal theory of having an object in mind were true, this fact would not suffice to show that any causal theory of *names* is true. For a causal theory of having an object in mind will not support a causal theory of names unless the concept of having an object in mind somehow generally figures in the semantic rules or conventions that people follow in using words as names. But as we shall see, it is far from clear that this concept does generally figure in these rules.

In his recent book *Designation*, Michael Devitt proposes a causal theory of having an object in mind and purports to use this theory as the basis of a causal theory of names.<sup>18</sup> The main concept of the book is that of designation, which Devitt explains in terms of a type of causal chain that he calls a "d-chain." The idea is that an utterance, or token, of a term designates an object if and only if the object "grounds," or is the ultimate source of, a d-chain that eventuates in the utterance. According to Devitt, his concept of a token's designating an object is a theoretical counterpart of the ordinary concept of a speaker's having an object in mind (or meaning an object) by a token (p. 33).

My question is: How is the concept of designation supposed to be relevant to the semantics of proper names? It is instructive to consider the way Devitt himself takes designation to be relevant to the semantics of *demonstratives*. His view is that the conventions for the various types of demonstrative terms all require that the terms' semantic referents be determined at least in part by what the terms designate. Since what is designated by a token corresponds intuitively to what the speaker refers to with (has in mind by) the token, Devitt's view is in essence the same as the view of demonstratives mentioned earlier, a view that I subscribe to. Devitt suggests, for instance, that the term 'this', 'that', and 'it' all mean (approximately) 'a

designated object', whereas 'he' and 'she' mean 'a designated male' and 'a designated female', respectively (p. 46). It is easy to put these suggestions in the form of reference rules. For instance, the relevant rules for 'this' and 'she' would be:

- (11) For any  $\alpha$ , if  $\alpha$  is a token of 'this', then for any  $x$ ,  $\alpha$  is to refer to  $x$  if and only if  $\alpha$  designates  $x$ ; and
- (12) For any  $\alpha$ , if  $\alpha$  is a token of 'she', then for any  $x$ ,  $\alpha$  is to refer to  $x$  if and only if  $\alpha$  designates  $x$  and  $x$  is female.

Notice that (12) is in effect equivalent to the rule (5) for 'she' that I introduced earlier.

Devitt proposes a causal theory of having an object in mind, and he describes how the latter concept figures in the various conventions or rules people follow in using demonstratives. By my criteria, Devitt has stated a causal theory of demonstratives, a theory that fits the referential paradigm. However, he nowhere states a comparable sort of theory for proper names. In sharp contrast to his treatment of demonstratives, Devitt never attempts to explain how the concept of designation figures in the conventions people follow in using words as names. In fact, nowhere in his book does he explicitly attempt to describe these conventions at all.

Thus, by my criteria, Devitt has not succeeded in stating a causal theory of names. Instead, he has given a causal theory of what he calls designation (having an object in mind) and then applied this theory to names to show the various ways that names can designate objects. But applying a theory of designation to names in this way is not at all the same as giving a theory of *names*. For, on Devitt's own view, ever so many different kinds of terms designate objects. Thus to merely describe how names designate objects in various ways is not to say anything that distinguishes names semantically from other kinds of terms, and so it is also not the same as giving a theory of names.<sup>19</sup>

There are of course various ways in which the concept of designation might conceivably figure in the rules people follow in using words as names. One obvious possibility is that these rules all have the following form:

- (13) For any  $\alpha$ , if  $\alpha$  is a token of  $N$ , then for any  $x$ ,  $\alpha$  is to refer to  $x$  if and only if  $\alpha$  designates  $x$ ,

where  $N$  refers to some name-type. But this proposal fails to distinguish names from other semantically different sorts of term. In particular, if this proposal were true, then whenever anyone uses a word as a name, he is following just the same sort of rule as the rule (11). But then the proposal has the false consequence that every name is semantically indistinguishable from the demonstrative 'this'. Moreover, since the proposal is that to use a word as a name is to follow a rule of the form (13), and since the rule (11) for 'this' is of that form, the proposal also has the absurd consequence that the demonstrative 'this' is itself a proper name.

Another sort of difficulty for the idea that reference rules for names have the form (13) is raised by examples of the following kind that Kripke has described.<sup>20</sup> Suppose that Smith always uses the word 'Reagan' as a name of the American

president, Ronald Reagan. But Smith mistakenly believes that his neighbor Jones—who perhaps looks just like Reagan—is the President, and because of this mistake Smith often refers to Jones with ‘Reagan’. On one such occasion, Smith says, “Look, Reagan is mowing his lawn for the third time this week.” Intuitively, as Kripke points out, Smith is using ‘Reagan’ *as a name of* Reagan but is *referring* with ‘Reagan’ to his neighbor Jones.<sup>21</sup> So if we assume that designation is a counterpart of speaker’s reference, this is a case in which a name-token semantically refers to one object (Reagan) and designates another (Jones).<sup>22</sup> But then the speaker cannot be following a rule of the form (13). Yet he is using ‘Reagan’ as a name. So again, the idea that the rules for names have the form (13) is false.<sup>23</sup>

At one point in his book (p. 151), Devitt mentions in passing a convention of *designating a certain object* with a name. So perhaps he would say that conventions for using names have the form:

- (14) For any  $\alpha$ , if  $\alpha$  is a token of  $N$ , then for any  $x$ ,  $\alpha$  is to designate  $x$  if and only if  $x = b$ ,

where  $b$  is a term referring to a certain object.<sup>24</sup> Now (14) is a rule governing designation, or speaker’s reference, as opposed to semantic reference. So it is difficult to understand how such a rule could determine a name-token’s *semantic* referent, especially since, as we’ve just seen, a name-token’s semantic referent can be an object that it *fails* to designate.<sup>25</sup> This problem suggests that perhaps Devitt did not have rules of the form (14) in mind, but rather rules of the form:

- (15) For any  $\alpha$ , if  $\alpha$  is a token of  $N$ , then for any  $x$ ,  $\alpha$  is to refer to  $x$  if and only if  $x = b$ ,

where again,  $b$  is a term referring to a certain object. But of course, the proposal that conventions for names are of this form does not help explain how the concept of designation is supposed to figure in these conventions.

Moreover there is a further overwhelming difficulty that confronts both the proposal of (14) and that of (15). Rules of this form are supposed to be rules that concern *certain objects*. So I take it that in the instances of (14) and (15) expressing such rules, the instances of ‘ $b$ ’ are supposed to be *genuine terms*.<sup>26</sup> Consider any instance  $I$  of (14) or (15) containing a genuine term  $b$ . Since  $b$  is genuine, the rule expressed by  $I$  is a function of the referent of  $b$ . But then unless  $b$  has a referent,  $I$  expresses no rule at all. Hence no rule has the form (14) or (15), unless that rule concerns an existing individual.

Now it is clear that we do not *always* follow such rules in using words as names. For sometimes we use a word as a name without using it as a name of any existing object. (A child asserts “Santa Claus is going to bring me a laser gun for Christmas.”) In such a case the rule being followed is not a rule concerning any real individual, and so the rule is neither of the form (14) nor of the form (15). So neither of these forms of rule can provide the basis of a generally adequate account of what it is to use words as names.<sup>27</sup>

Furthermore, it is highly implausible to suppose that speakers *ever* follow

rules of this kind in using words as names. For any such rule essentially involves a certain individual. Hence the mere fact that one is following such a rule in using a word guarantees that one is using the word as a name of an existing individual. So if we ever followed such rules, then we could sometimes know whether we are using a word as a name of an existing individual, merely by finding out which rule we are following. But since it is implausible to suppose that we could ever gain such knowledge in this way, it is also implausible to suppose that we ever follow rules of this form.

There are no doubt other ways in which Devitt's concept of designation might conceivably figure in the reference rules for names. But rather than discuss more of these possibilities, let us instead turn to a more general consideration of how causal concepts of any kind might figure in these rules.

#### 4. A GENERAL ARGUMENT AGAINST CAUSAL THEORIES

On every sort of causal theory of names, the semantic referent of a particular use or token of a name is determined by a causal or historical chain of communication reaching back in time from the use to an initial point at which an object, the use's referent, acquires the name in some way. But when confronted with this picture, we should ask ourselves, How *could* the referent of a name-use be determined by a causal chain of this sort? The only possible answer would seem to be that a name-use's referent can be determined by such a causal chain only if the rule or convention being followed by the speaker *makes* it the case that the use's referent is so determined. Thus causal theories seem to all be committed at a minimum to the view that every (or at any rate *almost every*) reference rule for using a name has the form:

- (16) For any  $\alpha$ , if  $\alpha$  is a token of  $N$ , then for any  $x$ ,  $\alpha$  is to refer to  $x$  if and only if  $x = (\exists y)Cy\alpha$ ,

where each instance of  $\lceil Cy\alpha \rceil$  expresses a relation involving causality, and  $\lceil x = (\exists y)Cy\alpha \rceil$  is equivalent to  $\lceil (\exists y) ((z)(Cz\alpha \equiv y = z) \& x = y) \rceil$ .

Within the range given by this minimal assumption, there is room for considerable variety. But even so, the theories that make this assumption share one striking feature in common. For on all of these theories, every proper name turns out to be a species of *indexical*.<sup>28</sup> On all of these theories, an object is the referent of a token of a name only if the object uniquely bears a certain causal relation to that token. But if this is true, then the referent of a name in a given context is always determined in part by features of the particular token that is uttered in the context, and this is the defining mark of an indexical.<sup>29</sup> The relevant instances of  $\lceil x = (\exists y)Cy\alpha \rceil$  in the various conventions for names would each be something like ' $x$  is the unique object that grounds such-and-such a kind of causal chain that eventuates in  $\alpha$ '. Typically, one imagines, chains of the various relevant sorts could link distinct objects to distinct tokens of the same name. If so, then the conventions for names, like those for most indexicals, allow a name's referent to vary from context to context.

However, it seems to me that proper names are intuitively *not* indexicals, and that instead their referents are determined independently of context. Most philosophers writing on names, I believe, have shared this intuition, either explicitly or implicitly. For instance, in philosophical discussions of names, it is common to assume for the sake of simplicity that name-types have unique referents. Such an assumption is quite harmless, provided that each convention or rule for a name determines a unique referent for that name, for then the assumption amounts to just ignoring for simplicity's sake the possibility of a single name's falling under more than one convention. But if names were indexicals, the assumption would be far from harmless; it would be like assuming that the word "I" has a unique referent.

And when philosophers do remark the fact that a single name-type may have several distinct referents, they almost always classify this as a kind of *ambiguity*.<sup>30</sup> This classification assumes, correctly I think, that distinctness of a name's referents in different contexts is a sign that the name is being used with distinct linguistic meanings in those contexts, an assumption that is reasonable if each rule or convention for a name determines a unique referent for that name, but an assumption that commonly fails for indexicals.

It is worth noting that intuitions of the above kind have frequently been expressed by *causal theorists*, indicating a significant tension in their views between these intuitions and the picture they've proposed of how names' referents are determined.<sup>31</sup>

In addition to these intuitive grounds, there are strong theoretical reasons for thinking that names cannot be indexicals. We have seen that to use a word as a name is to follow a certain sort of semantic rule. Similarly, it would seem, a word *is* a name in a certain language just in case the language contains a semantic rule of the relevant sort governing that word. Now if a word is a name in a language by virtue of the language's containing a certain rule, and if one is following that rule in uttering the word, then surely one would be using the word as a name, and hence would be using the word *as a name of* whichever object is the referent of one's utterance.

But now consider the imaginary indexical term 'toof' that is governed in a certain language *L* by the following rule:

- (17) For any  $\alpha$ , if  $\alpha$  is a token of 'toof', then  $\alpha$  is to refer to an object  $x$  if and only if  $x$  is the unique object that is precisely two feet in front of the speaker of  $\alpha$ 's nose at the time he utters  $\alpha$ .

Suppose that a speaker *s* follows (17) on a given occasion in uttering a token  $\alpha$  of 'toof', and that  $x$  is the unique object two feet in front of *s*'s nose at the time of utterance. Thus  $x$  would be the referent of  $\alpha$ . But would *s* be using 'toof' *as a name of*  $x$  on this occasion? Obviously not. But why not? The reason, I suggest, is that the existence of the rule (17) in *L* would not be sufficient to make 'toof' a name in *L of* any particular one of the indefinitely huge number of objects that it might refer to in different contexts. And in order to be *used as* a name of an object, a word must *be* a name of that object.

We said that if a word is a name in a language by virtue of the language's

containing a certain rule, then the existence of that rule in that language would allow the word to be used as a name of whichever object is referred to by any particular use of the word. But as we've just seen, this in turn implies that the existence of such a rule in a language would be sufficient to make a word *the name of* whichever object is referred to by any particular use of the word.

Now as the example of 'toof' shows, the existence of a rule in a language can make a word a name of an object in that language only if the rule contains a condition that determines a unique referent for that name. But then, the sort of rule in question cannot be the sort of rule we follow in using indexicals. Hence, the existence in a language of a rule cannot make a word a name of any object, if that rule is the kind we follow in using indexicals. But again, the existence in a language of the kind of rule we follow in using words as names *can* make a word the name of an object. Therefore, the kind of rule we follow in using names is *not* the kind of rule we follow in using indexicals. Therefore, names are not indexicals.<sup>32</sup>

It will help to nail the point down if we consider an example of a kind of rule that would be sufficient to make a word the name of an object in a language. In section 1 we saw that a name's referent could in principle be fixed in a language by a context-independent definite description. For instance, Kripke has described the possibility that 'Neptune' might have been first introduced as a name of whatever planet uniquely causes such-and-such perturbations in the orbit of Uranus.<sup>33</sup> Thus the following *could* be a rule of a given language *L*:<sup>34</sup>

- (18) For any  $\alpha$ , if  $\alpha$  is a token of 'Neptune', then  $\alpha$  is to refer to an object  $x$  if and only if  $x$  is the unique planet that causes the perturbations in the orbit of Uranus.

Now suppose we've just discovered that (18) is a rule of *L*. Would we say that 'Neptune' is a *name* in *L*? Yes, obviously. And assuming that a certain planet uniquely satisfies the relevant description, would 'Neptune' be a name *of* that planet in *L*? Again, the answer is obviously Yes.

Thus merely by knowing that a word's referent is fixed in a language by a context-independent description, we know automatically that the word is a name in that language. This fact supports my suggestion that for a word to be a name in a language is for the language to contain a reference rule of a certain sort. It also confirms my contention that the sort of rule in question is not the sort of rule that governs indexicals.

The evidence we've cited seems to lead inevitably to the conclusion that *to be a name just is to be a context-independent genuine term*.

This view is theoretically satisfying because it not only allows us to distinguish names from both descriptions and indexicals, but it also allows us to see the respects in which names are significantly similar to both of these other types of term. Like indexicals but unlike descriptions, names are genuine terms. Like many descriptions but unlike indexicals, names are context-independent. The fact that names are significantly like both descriptions and indexicals explains why philosophers have been prone to assimilate names to one or the other of these two types of term. The

fact that names are significantly different from both descriptions and indexicals shows that such assimilations are mistaken.

If names are not indexicals, then it is false to suppose that any reference rule for using a name is ever of the form (16). Thus, no causal theory of names according to which the conventions or rules for names are of this form is true. Yet as we've seen, the picture provided by causal theorists of how names' referents are determined is most naturally interpreted as suggesting that the conventions for names are of the form (16).

In fact, if the conventions for using names are not of this form, then there are very few alternatives left to a causal theorist. We saw in section 3 that the rules people follow in using names cannot involve actual individuals, and thus we eliminated the possibility that these rules are of the form (15), where  $b$  is a genuine term referring to a certain object. But if the reference rules for names must make names context-independent, and yet these rules are not of the form (15), then there is only one sort of form left that these rules could take. *They must be rules of the form (4) that are expressible by means of context-independent definite descriptions.*

So the only sort of view left for a causal theorist to take is a view according to which the various reference rules that people follow in using words as names are each of the form:

- (19) For any  $\alpha$ , if  $\alpha$  is a token of  $N$ , then for any  $x$ ,  $\alpha$  is to refer to  $x$  if and only if  $x = (\exists y)Cy$ ,

where in each such rule the instance of ' $Cy$ ' expresses a property involving causality. However, I should think that most causal theorists would be unable to accept this idea with equanimity. For causal theorists recommend their sort of view as a preferable alternative to description theories of names. In opposition to description theories, the proponents of causal theories assert that the referents of most uses of names are not determined by any properties or descriptions that are associated with the names by their speakers and that typically a name-use would have the referent it in fact has, even if the descriptions associated with the use were satisfied by objects other than the referent or by nothing at all.<sup>35</sup>

But surely, if whenever a speaker is using a word as a name he or she is following a reference rule of the form (19), then the speaker associates with that use the property mentioned in the rule's reference condition, and an object must uniquely satisfy this property to be the referent of the name-use in question. Of course this "association of a property with his or her use" amounts to no more than is implied by the fact that the speaker is "following the rule" in question, and it is difficult to say precisely what is implied by this sort of fact. I do not think that a speaker need know, or be explicitly aware of, the rule being followed, in order to be following it. Nor need the speaker be able to formulate the rule being followed.

But the fact that a speaker is following a rule of the form (19) would at least have to be reflected in the truth of various counterfactuals concerning the speaker's mental dispositions. Thus if in uttering a token of 'Reagan' Smith is following the rule that tokens of 'Reagan' are to refer to the president, then Smith must be

disposed, among other things, to think that his utterance is not about a real man, if he comes to believe that the president never existed and is, say, a figment of his imagination. It is in this manner that a speaker who follows a rule of the form (19) would be “associating” the relevant property with his use of a name.

So I think it is clear that if to use a word as a name one must follow a rule of the form (19), then the correct theory of names must be a description theory, and causal theories that deny the truth of every description theory must be false. Of course, if some description theory based on rules of this form were correct, then an element of causality would be involved in the various rules that determine reference for names. But I should think it would be of small comfort to a causal theorist to know that the correct theory of names is a description theory of this form.

## 5. THE PARADOX OF NAMES

Our discussion so far shows that whatever semantic rules we follow in using words as names must be expressible by use of context-independent definite descriptions, and this is a serious difficulty for causal theories. Yet the idea that names’ referents are determined by associated descriptions is itself not without difficulties. For Donnellan and Kripke have described examples that make a convincing case for the conclusion that an object can be the referent of a name, even though the object uniquely satisfies none of the properties that are commonly associated with the name.<sup>36</sup>

One of these examples is Kripke’s Gödel-Schmidt case. Practically the only thing most people have heard about Gödel is that he discovered the incompleteness of arithmetic. So it is quite plausible to suppose that this is the only property commonly associated with the name ‘Gödel’ that Gödel in fact uniquely satisfies. But as Kripke points out, people’s uses of the name ‘Gödel’ would still succeed in referring to Gödel even if it had not been Gödel but an unknown Viennese named ‘Schmidt’ who actually discovered incompleteness.<sup>37</sup> Moreover, no matter how many further properties might be commonly attributed to Gödel and associated with his name—properties, for example, involving additional achievements for which Gödel is famous—it is quite clear that people’s uses of ‘Gödel’ would refer to Gödel even if he had possessed none of these properties, even if someone else had done all those things for which Gödel is famous.

So an object can be the referent of a name, or a particular use of a name, even though the object uniquely satisfies none of the properties that are commonly, or publicly, associated with the name. This in fact seems generally true of the names we use. But then it surely must be false to suppose that names are governed in public languages by reference rules that are expressible by use of definite descriptions. For if a name were governed by such a rule in a public language, English say, then speakers of English would commonly follow this rule in using the name, and so there *would* be a property that is commonly associated with the name by speakers of English, a property that the name’s referent would have to uniquely satisfy. But again, there in general seem to be *no* such properties.

We saw earlier that the reference rules we follow in using names are not the kind of rules that govern indexicals and that they are also not rules that are expressible by use of genuine terms. The only alternative left is that these rules are expressible by use of context-independent definite descriptions. Yet we've just seen that names are in fact not governed in public languages by rules that are expressible with descriptions. It seems to follow that names are governed by no semantic rules or conventions at all! But if this is so, then it seems impossible that names should even *have* referents. This is the paradox of names.

It is important to be clear about what examples like the Gödel-Schmidt case do and do not show. I have said that these examples show that an object can be the referent of a name or use of a name, even though the object uniquely satisfies none of the properties that are *commonly associated* with the name. Causal theorists usually make the additional claim that these examples show that an object can be the referent of a use of a name, even though the object satisfies none of the properties that *the speaker associates* with the name. However, as I have argued elsewhere, this additional claim is not supported by the features of the examples.<sup>38</sup>

A typical use of the name 'Gödel' by a speaker who associates the property of having discovered incompleteness with this name would no doubt refer to Gödel even if Schmidt and not Gödel had discovered incompleteness. But this would not be a case in which Gödel, though the referent of the use, fails to uniquely satisfy *every* property that the speaker associates with the name 'Gödel'. For among the many other properties that a typical user of 'Gödel' would associate with the name are such properties as being one to whom the speaker has heard others refer with 'Gödel'; being a man of whom the speaker has heard that he discovered incompleteness; and so on. Now these are not properties that are commonly associated with the name 'Gödel', for they are properties that involve the particular speaker in question. But nevertheless, they are properties that Gödel would uniquely satisfy in the example. So it is open to a description theorist to maintain that the speaker's use of 'Gödel' would intuitively refer to Gödel in the example simply because the use's referent would be typically determined by one or more of these other properties that *the speaker* associates with the name.

Causal theorists sometimes object that description theorists cannot appeal to such "buck-passing" descriptions as 'the one to whom I have heard others refer with "Gödel"' without involving their view in some kind of circularity.<sup>39</sup> But in my opinion, no clear account of this alleged circularity has yet been given. Of course, if the description theorist's view were that a name's referent is *always* determined by buck-passing descriptions, then his view would in a sense be circular. But this just shows that a description theorist should agree that the success of any name-use based on borrowed reference depends ultimately on the existence of speakers who are in a position to make *independent* references to the object in question.

Perhaps a causal theorist would claim that description theories must inevitably give a false account of *unborrowed* name-reference, because if we trace any chain of reference-borrowing involving a name back to its initial point, the description theorist's account of how the name's referent is initially determined by descriptions

that are *not* buck-passing will always be open to counterexamples of the Gödel-Schmidt sort. But this claim would be unjustified. A typical buck-passing use of 'Gödel' would surely be traceable eventually to one or more of Gödel's close relatives, colleagues, or other acquaintances. But can we *really* describe an example that shows that, say, the uses of 'Gödel' by one of Gödel's closest colleagues would have referred to him, even had he uniquely satisfied *none* of the no doubt enormous number of properties that are not buck-passing that the colleague associated with his name? It surely seems unlikely that any such example is forthcoming.

I have yet to see a decisive objection to the point that in the Gödel-Schmidt case and others like it, the referents of the names involved would intuitively be determined by buck-passing properties. So it is consistent with these cases to suppose that the referents of names are in general determined by properties that are associated with the names by their speakers. This fact provides the basis of my solution to the paradox of names.

## 6. THE PRIVATE-RULE THEORY

We have seen that to use a word as a name is to follow a certain sort of reference rule of the general form (4). We have also seen that the rules in question cannot be expressed by genuine terms and cannot be the kind of rule that governs indexicals. Thus the rules we follow in using names *must* be rules of the form (4) that are expressible by context-independent descriptions. In fact, I suggest, one uses a word as a name if and only if one follows such a rule, a rule of the form:

- (2) For any  $\alpha$ , if  $\alpha$  is a token of  $N$ , then for any  $x$ ,  $\alpha$  is to refer to  $x$  if and only if  $x = (\exists y)Fy$ ,

where each instance of ' $Fy$ ' expresses a property.<sup>40</sup> This suggestion does justice to the conclusion reached earlier that to be a name is just to be a context-independent genuine term.

But again, examples like the Gödel-Schmidt case show that names are not governed in public languages by rules of the form (20). This may sound inconsistent with the principle that to use a word as a name is to follow a rule of the form (20). But in fact there is no inconsistency. What follows instead of a contradiction is the conclusion that *the semantic rules we follow in using words as names are not rules of the public languages we are speaking in using the words.*

Now I admit that this consequence seems implausible at first glance. It is certainly not a possibility that strikes one as immediately obvious. But nevertheless, I maintain that our unlikely-seeming consequence *must* be true. For if it is not, then we have to deny one of the above set of compelling assumptions that together imply it, and the paradox of names is unresolved.

To make this solution to the paradox acceptable, we need to make it plausible that people could be meaningfully speaking a language in using a word, even though the semantic rule they are following is not a rule of the language they are speaking. This can be done by use of the concept of a second-order rule that was

introduced earlier. In section 2, I argued that the basic rules regarding singular terms in natural languages are second-order rules that permit the following of certain first-order reference rules and prohibit the following of others. The only kind of second-order rules considered before were rules that explicitly mention the first-order rules that they permit speakers to follow. But it is conceivable that a language could contain a second-order rule that permits speakers to follow any of a certain general *sort* of first-order rule, without mentioning any *particular* rule of that sort. This, I suggest, is the kind of second-order rule regarding names that occurs in natural languages.

Simply put, the second-order rule for names that occurs in natural languages is just the following: it is permitted to use any word as a proper name. More precisely, the rule is:

- (21) For any token  $\alpha$  of any word  $N$ , if  $\alpha$  is used as a proper name by a speaker, then the referent of  $\alpha$  is to be determined in accordance with the first-order rule being followed.

Here it is of course understood that to use a word as a proper name the speaker must be following a first-order rule of the form (20).

On the view I am proposing, (21) is the *only* semantic rule concerning names that occurs in natural languages like English. But (21) mentions no particular first-order reference rule involving any particular name. So the existence of (21) in English is not sufficient to turn any particular first-order rule of the form (20) into a non-basic rule of English. Thus when a speaker of English uses a word as a name, he or she is following a reference rule of the form (20), but the rule being followed is not a rule of English. Yet such a speaker would nevertheless be meaningfully speaking English, for in following a rule of the form (20), he or she would be doing something permitted by the rules of English, and the name-use in question would have its referent determined accordingly.

The basic idea is that proper names are like "wild cards" that the rules of English allow to be used in accordance with unspecified rules of a certain sort at the speaker's discretion.

The following is my official explication of what it is for a speaker to use a word as a proper name:

- (22)  $\alpha$  is used as a proper name by  $s = df.$  There are a property  $F$  and a word  $N$  such that: (i)  $\alpha$  is a token of  $N$ ; (ii)  $s$  utters  $\alpha$ ; (iii) in uttering  $\alpha$ ,  $s$  is actualizing a stable disposition to utter tokens of  $N$  in certain circumstances with the understanding that these tokens are subject to the rule that:

- ( $R$ ) For any  $\beta$ , if  $\beta$  is a token of  $N$ , then for any  $x$ ,  $\beta$  is to refer to  $x$  if and only if  $x = (\iota y)Fy$ ;

and (iv) is logically possible that, for some object  $x$ , both  $s$  utters  $\alpha$  and  $x = (\iota y)Fy$  even though  $\alpha$  does not refer to  $x$ .

This definition has two new features that require further explanation. Statement (22) conforms to the idea already proposed, that to use a word as a name is to follow a reference rule of the form (20). But in addition, (22) requires that a speaker who uses a word as a name must be exercising a stable disposition to follow the rule in question in certain circumstances. This requirement is necessary to avoid violation of the principle enunciated earlier that a word cannot be *used as* a name of an object unless it *is* a name of that object. For a word to be a name of an object, I suggest, it is sufficient that there be at least one person for whom the name is a name of that object. And further, I would say that a word is a name of an object for a person just in case the person has a stable disposition to in certain circumstances follow a rule of the form (20) concerning that word, where the object uniquely satisfies the property *F* mentioned in this rule. Given these assumptions, it follows from definition (22) that a person cannot use a token of a word as a name of an object unless the word is a name of that object.

The second feature of (22) requiring explanation is the presence of clause (iv). This clause is intended to rule out the occurrence of what we might call “blatantly question-begging” properties in rules of the form (20) that determine reference for uses of names. For instance, consider the property of being referred to by all tokens of ‘Socrates’ that are uttered by *s*. We surely would not want to allow such a property to occur in any rule of the form (20) that determines reference for use of *s* of ‘Socrates’. Clause (iv) prevents such properties from occurring in rules of the relevant kind.

On the other hand, (22) does allow what we might call “factually question-begging” properties to occur in rules of the form (20) that determine name-reference. For instance, (22) allows buck-passing descriptions like ‘the one Jones’s tokens of “Socrates” always refer to’ to occur in rules of the relevant sort. If Smith always follows such a rule in uttering ‘Socrates’, and it should happen that Jones passes the buck right back to Smith by always following a similar rule concerning ‘the one Smith’s tokens of “Socrates” always refer to’, then of course neither Smith’s nor Jones’s tokens of ‘Socrates’ would ever have referents. But the fact that this unfortunate sort of thing *can* happen is not a good reason to insist that buck-passing properties could *never* determine a name’s referent. For surely not all, and perhaps very few, buck-passing properties actually turn out to be factually question-begging in this way.

The theory of names that I am proposing has two main parts, each of which is independent of the other. The first part is the explication (22) of what it is to use a word as a name. The second part is the thesis that the only rule about names to be found in natural languages like English is the second-order rule (21). As we’ve seen, it is a consequence of this second thesis that the rules of the form (20) that we follow in using words as names are, as a matter of fact, not rules of the languages we are speaking. But I should stress that my theory allows there to be *possible* languages that contain rules of this form governing words that are names in those languages.

## 7. SOME FEATURES AND CONSEQUENCES OF THE PRIVATE-RULE THEORY

The private-rule theory is a description theory of names on which names' referents are determined by privately associated descriptions. Theories of this kind are of course not new to the literature. I take the chief novelty of the present paper to lie not in the type of theory proposed but in the new *argument* I've given in favor of this type of theory, an argument that if sound refutes other types of theories of names, such as causal theories. But the private-rule theory does have some novel features that both distinguish it from other forms of private-description theories and make it preferable to these other theories.

The main intuitive difficulty that faces any private-description theory is this: how can a description that is idiosyncratically associated with a name by a given speaker determine the semantic referent of that speaker's uses of the name in a public language? The private-rule theory gives a novel solution to this problem via the hypothesis that natural languages contain a second-order rule permitting speakers to invent and follow their own rules for proper names. I should stress that this second-order rule is *not* the same as a permissive rule that allows a speaker's use of a name to semantically refer to whatever the speaker intends it to refer to on a given occasion. A rule of this latter sort is the kind of rule that governs demonstratives like 'this', and as we've seen, the hypothesis that names are governed by such rules would not allow us to properly distinguish names from demonstratives. Instead of permitting speakers' uses of names to refer to whatever they like, the second-order rule in English for names in effect directs speakers to supplement English with new rules of the form (20). When a person does this and then follows such a rule, the person's name-use does not just refer to whatever he likes. Instead, it is *as if* the person were speaking a language containing the new rule in question, and his linguistic behavior is accordingly constrained by this rule.

A related difficulty that often plagues private-description theories is that of adequately distinguishing speaker's reference from semantic reference in the case of names. Consider again the example of Smith, who always uses 'Reagan' as a name of the president, but who by mistake often uses 'Reagan' to refer to his neighbor Jones. Of the descriptions associated by Smith with 'Reagan', some are true of the semantic referent, Reagan, and some are true of Smith's neighbor Jones. The problem for a description theory is to provide a method of distinguishing those descriptions that determine a name's semantic referent from those that determine what the speaker is referring to. The private-rule theory helps to provide such a method. When Smith says such things as "Look, Reagan is mowing his lawn again," Smith's primary intention is to say something about his neighbor. This is why it is intuitively correct to say that Smith is referring to Jones with his use of 'Reagan'.<sup>41</sup> But it is consistent with Smith's having this primary intention that when he utters 'Reagan', he is *following a semantic rule* of the form (20) that requires his utterances of 'Reagan' to refer to the president, and not Jones. This explains how it is possible for Smith to

use 'Reagan' as a name of the president while at the same time he is referring with 'Reagan' to his neighbor.

One of the most important consequences of the private-rule theory is the distinction it implies between a word's being meaningfully used in a language and the word's having a particular meaning in that language. As we explained earlier, the private-rule theory allows a word to be meaningfully used as a name, even though the rule being followed is not a rule of the speaker's language. The theory also says that natural languages like English do not contain rules of the form (20) for proper names. And since a name has a particular linguistic meaning in a language only if the language contains a rule of the form (20) governing that name, the theory implies that as a matter of fact, proper names have no linguistic meanings in English or other natural languages. Another way of putting this is to say that *proper names are not words of natural languages*.

The idea that names do not have particular meanings in their speakers' languages has not been discussed very much, but a few philosophers have expressed intuitions that support the idea. These intuitions confirm the private-rule theory. For instance, Strawson has pointed out that "ignorance of a man's name is not ignorance of the language."<sup>42</sup> Vendler has noted that names do not require translation into another language and that, accordingly, dictionaries do not list proper names.<sup>43</sup> And Ziff has given the following argument to support the idea that names are not words of their speakers' languages:

If I say 'Are you familiar with Hsieh Ho's view on art?' I am speaking English: I am not speaking a combination of English and Chinese. Yet if 'Hsieh' and 'Ho' are words then they can only be words of Chinese . . . and I must speak a combination of English and Chinese, which is absurd.<sup>44</sup>

Ziff is clearly right that having adopted a name used by speakers of another language, we may use it in sentences of pure English. But is he right when he claims that 'Hsieh Ho' could only be a word of Chinese? Why isn't it true that when we adopt the use of this name, it then *becomes* a word of English as well as Chinese?

If 'Hsieh Ho' becomes a word of English when we adopt its use as a name, then when we adopt this use, we are changing English. For we will have added a word to English, and to do this we will have had to add a new semantic rule to the rules of English, namely, a rule for 'Hsieh Ho'. But it seems clearly false that we have changed English in any way when we begin using 'Hsieh Ho' as a proper name. This seems true in general: when we introduce a new name for an object, we are not changing English. We are instead doing something that is already permitted and anticipated by the rules of English. Since the private-rule theory predicts that we do not change our language when we adopt new uses of names, the intuition that this is so is evidence in favor of the theory.

In order to explain how proper names work, I have had to argue for hypotheses about the semantic structure of natural languages that make such languages look

far different than they are usually described. I have had to explain how words that are literally without meanings can yet have referents and be used to express propositions. And I have had to explain how it is possible to follow a semantic rule and thereby be speaking a language, even though the rule is not a rule of that language. Kaplan has aptly remarked that proper names "are like bicycles. Everyone easily learns to ride, but no one can explain how he does it."<sup>45</sup> One further small piece of evidence in favor of my theory is that if it is true, then it is easy to see why proper names have proven so difficult to understand.

### Notes

1. The research on this paper, at various stages of its development, was supported by a Summer Stipend Award from the National Endowment for the Humanities and by a grant from the American Council of Learned Societies under a program funded by the National Endowment for the Humanities. Material in this paper forms part of a larger work currently in preparation on the semantics of singular terms and the intentionality of mental states. Many of the ideas and views expressed here, especially those in section 1, will receive much more extensive elaboration and defense in the larger work.

I am especially indebted to Lawrence Powers, whose valuable criticism of my earlier views led me to write this paper and who made useful comments on an earlier draft. I am also grateful to Lawrence Lombard and John Tienson for helpful comments and discussions.

2. I use the expression "to follow a rule" in such a sense that it means more than just acting in accordance with a rule. To follow a rule in the sense I mean is something like thinking of one's behavior as subject to the rule, or letting one's behavior be guided by the rule. I do not have an analysis of the concept, but I think its intuitive content is clear enough to justify my use of it here.

3. See Gottlob Frege, "On Sense and Reference," in *Translations from the Writings of Gottlob Frege*, edited by P. Geach and M. Black (Oxford, 1966).

4. See Frege's "The Thought: A Logical Inquiry," translated by A. M. Quinton and M. Quinton, *Mind* 65(1956):289-311.

5. David Kaplan, "On the Logic of Demonstratives," in *Contemporary Perspectives in the Philosophy of Language*, edited by P. A. French, T. E. Uehling, and H. K. Wettstein (Minneapolis, 1979), 403. What I am calling "linguistic meaning" and "sense," Kaplan calls "character" and "content," respectively.

6. Saul Kripke, "Naming and Necessity," in *Semantics of Natural Language*, edited by D. Davidson and G. Harman (Dordrecht, 1972), 270.

7. Gareth Evans puts the point this way in "Reference and Contingency," *Monist* 62 (1979), 173.

8. A description is contingent if and only if: it is possible for an object to satisfy the description, and necessary that any object that does satisfy it could exist without doing so.

9. It is worth noting that Kripke's modal intuitions about names apply equally well to indexicals, thus providing more evidence that indexicals are genuine terms.

10. "Naming and Necessity," 276-77.

11. See, for example, R. Thomason and R. Stalnaker, "Modality and Reference," *Nous* 2(1968):359-72.

12. For a fuller statement and defense of this approach, see Evans, "Reference and Contingency," 167-70.

13. Here I am assuming that any reference-condition  $\varphi$  that occurs in a rule of the form (4) for an indexical term can always be expressed in terms of a two-place relation that the referent of any token of the term uniquely bears to *that token itself*. I eliminate the need for variables ranging over other contextual features such as the speaker, time, and place by using the token to fix these features. Thus, a token  $\alpha$  of 'I' refers to  $x$  just in case  $x$  is the unique speaker of  $\alpha$ ; a

token  $\alpha$  of 'now' refers to  $x$  just in case  $x$  is the unique time at which  $\alpha$  is uttered; and so on.

14. For an example of this view of demonstratives, see Tyler Burge, "Demonstrative Constructions, Reference, and Truth," *Journal of Philosophy* 71(1974):205-23. On the distinction between speaker's reference and semantic reference, see Saul Kripke, "Speaker's Reference and Semantic Reference," *Midwest Studies in Philosophy* 2(1977):255-76.

15. On this picture, notice, English contains *no* basic first-order reference rules for genuine terms. This is because in English, *any* term that is not a name *could* be used as a name, so that any first-order rule for a term that is not a name may be violated. We shall see below in section 6 why English contains no basic first-order rules for names themselves. Of course, my view allows there to be *possible* languages containing basic first-order rules for terms, provided that these terms are *unambiguous* in the languages in question.

16. Views of this kind have been suggested by David Kaplan, "Quantifying In," *Synthese* 19(1968):178-214; Keith Donnellan, "Speaking of Nothing," *Philosophical Review* 83(1974):3-31; Gareth Evans, "The Causal Theory of Names," in *Naming, Necessity, and Natural Kinds*, edited by S. Schwartz (Ithaca, 1977), 192-215; and Michael Devitt, "Singular Terms," *Journal of Philosophy* 71(1974):183-205.

17. See my paper, "Names and Intentionality," *Philosophical Review* 87(1978):171-200.

18. New York, 1981.

19. For further discussion, see my "Review of *Designation*, by Michael Devitt," *Canadian Philosophical Reviews* 3(1983):112-16.

20. See "Speaker's Reference and Semantic Reference," 263.

21. Here I am disagreeing with the theory of names that I proposed in "Names and Intentionality," 190-97. That theory implies that a name-token semantically refers to an object only if its speaker refers to that object with the token. But I now believe that this implication is false. Lawrence Powers changed my mind by pointing out to me that on each occasion when Smith refers to Jones with 'Reagan', he is doing so *just because 'Reagan' is his name for the president*. So it is intuitively correct to say that on each such occasion, Smith is still using 'Reagan' as a name of the president, and thus his use semantically refers to the president, even though *he* refers to Jones.

22. Here it matters whether we take designation to be a counterpart of speaker's reference or of having an object in mind. These latter two concepts are not precisely the same, since one can refer to only one object with a name-token, but one can have more than one object in mind in using the token. (I ignore this distinction in the text for the sake of simplicity.) In the present case, Smith has *both* Reagan and Jones in mind in using 'Reagan'. So if we take designation to be a counterpart of "having in mind," the proposal of (13) is still false, for then Smith's token of 'Reagan' would designate Jones even though the token does not semantically refer to him.

23. Devitt disagrees with Kripke's view (and mine) that in cases like this, the speaker is referring to an object that is not the name's semantic referent. He would prefer to say that the speaker *partially* refers to both Jones and Reagan. (See Devitt's paper, "Donnellan's Distinction," *Midwest Studies in Philosophy* 6(1981):511-24.) But this view still yields the result that the proposal of (13) is false, for on this view, Smith's token of 'Reagan' would semantically refer to Reagan without (fully) designating him.

24. Evans also suggests that the conventions for names are of this sort in "The Causal Theory of Names," 209.

25. I have argued at length that reference rules do not govern speaker's reference, or any other kind of psychological state, in my paper "Psychologism in Semantics," *Canadian Journal of Philosophy* 13(1983):1-25.

26. One could of course give interpretations of the proposals of (14) and (15) on which  $b$  is not assumed to be genuine but is instead, say, allowed to be a definite description. I doubt that Devitt had a proposal of this kind in mind, but in any case we'll take theories based on such rules into consideration below.

27. My earlier theory of names in "Names and Intentionality" also suffered from this kind of defect.

28. I ignore rules of the form (16) in which the occurrence of  $\alpha$  in  $\ulcorner (1y)Cy\alpha \urcorner$  is *vacuous*, as when for example the instance of  $\ulcorner (1y)Cy\alpha \urcorner$  is of the form  $\ulcorner (1y)(Fy \ \& \ \alpha = \alpha) \urcorner$ .

29. See note 13 above.

30. See, for example, John Searle, "Proper Names and Descriptions," in *The Encyclopedia of Philosophy*, vol. 6, edited by P. Edwards (New York, 1967), 490.

31. For instance, I mentioned above Devitt's intuition that the convention for a name is a convention that the name is to designate *a certain object*, and this of course is a convention that would make the name in question context-independent rather than indexical. And at one point in his book, Devitt remarks: ". . . for two tokens of an ambiguous physical name type are intuitively of different semantic types if they designate different objects. This is not the case with demonstratives and definite descriptions . . ." (*Designation*, 136.) Another causal theorist, Kripke, has also suggested that a name-type with several distinct referents should be thought of as homonymous. See the preface to *Naming and Necessity* (Oxford, 1980), 7-10.

32. The above argument is not conclusive as it stands because it is possible for there to be indexicals that could not refer to more than one object. One might claim that names are indexicals of this sort and thereby avoid my argument. Still, a similar point can be made about this (very unusual) sort of indexical. Suppose we have an indexical  $W$  that is governed by a rule of the following form in a language  $L$ :

- (R) For any  $\alpha$ , if  $\alpha$  is a token of  $W$ , then  $\alpha$  is to refer to an object  $x$  if and only if  $x = (1y)(y = (1z)Fz \ \& \ Cy\alpha)$ .

Though  $W$  could not refer to more than one object (since any object it ever refers to is  $(1z)Fz$ ), it could refer to an object  $x$  in one context and yet refer to *no* object in another context (since  $x$  and the token uttered might fail to satisfy  $\ulcorner Cy\alpha \urcorner$ ). But no *name* could have this sort of property. For if  $W$  is a name of  $x$  in  $L$  by virtue of  $L$ 's containing a certain rule, and one is following this very rule and thereby using  $W$  as a name, then surely one would have to be using  $W$  as a name *of*  $x$ , and so one's token would *have* to refer to  $x$ . So indexicals of this special sort cannot be names either.

33. "Naming and Necessity," 347-48, 33.

34. Causal theorists usually qualify this suggestion by saying that after a name has been introduced in this way, *future* uses of the name do not *in fact* have their referents determined by the introducing description, but rather by their causal relation to the introduction ceremony. But no causal theorist would, I think, wish to deny that a name *could*, at least in principle, have its referent permanently fixed in a language by a rule like (18). See the discussions by Donnellan in "The Contingent *A Priori* and Rigid Designators," *Midwest Studies in Philosophy* 2(1977): 12-27; and by Evans, "Reference and Contingency."

35. See Kripke, "Naming and Necessity," 293-303; and Devitt, *Designation*, 13-23.

36. Kripke, *ibid.*; and Donnellan, "Proper Names and Identifying Descriptions," *Synthese* 21(1970):335-58.

37. "Naming and Necessity," 294.

38. See my paper, "Kripke's Objections to Description Theories of Names," *Canadian Journal of Philosophy* 8(1978):485-97.

39. See Kripke, "Naming and Necessity," 297-300; and Devitt, "Brian Loar on Singular Terms," *Philosophical Studies* 37(1980):271-80. I replied to Kripke's objections about circularity in my "Kripke's Objections," *ibid.*; Loar replies to Devitt's similar objections in his "Names and Descriptions: A Reply to Michael Devitt," *Philosophical Studies* 38(1980):85-89.

40. I will add some qualifications to this proposal below in definition (22). Notice that the proposal is consistent with the possibility that the properties in the rules that determine names' referents often, or even typically, involve an element of causality. I myself am inclined to believe that this is true because it seems that names' referents are typically determined by buck-passing properties, and these properties usually seem to involve an element of causality. ("The one I have *heard* others refer to with 'Gödel'," for example.) If speakers do typically rely on buck-passing properties in their uses of names, this is no doubt because they generally desire to

use names in conformity with other speakers' uses to facilitate communication. I take this to be the grain of truth that lies behind causal theories. Of course, as I point out in the text, no description theory can imply that names' referents are *always* determined by buck-passing properties, on pain of circularity.

41. On the connection between referring to an object and primarily intending to say something about the object, see my "Names and Intentionality," 191-93. See also my "Causes and Intentions: A Reply," *Philosophical Review* 90(1981):408-23.

42. P. F. Strawson, "On Referring," in *Readings in the Philosophy of Language*, edited by J. Rosenberg and C. Travis (Englewood Cliffs, N. J., 1971), 190.

43. Zeno Vendler, "Singular Terms," in *Semantics*, edited by D. Steinberg and L. Jacobovits (Cambridge, 1971), 117.

44. Paul Ziff, *Semantic Analysis* (Ithaca, N. Y., 1960), 86.

45. David Kaplan, "Dthat," in French, Uehling, and Wettstein, *Contemporary Perspectives*, 385.