
THE JOURNAL OF PHILOSOPHY

VOLUME LXXXI, NO. 9, SEPTEMBER 1984

REFERENCE EXPLAINED AWAY

Anaphoric Reference and Indirect Description

THIS paper presents a construction by means of which I believe it is possible to account for the use of expressions containing 'refers' and its cognates in natural languages. Reference comes in two varieties: word-*world* or extralinguistic reference of the sort invoked when we say that the phrase 'the first left-handed President' refers to a certain concrete individual (namely James Garfield), and word-*word* intralinguistic or anaphoric reference instanced by pronouns such as 'she' referring to the word 'Mary' in the sentence "Mary tired of surgical training, so she became an anesthesiologist." Intralinguistic reference of this sort has not been of much interest to philosophers, for it has seemed natural to assimilate pronouns to bound variables, and so to expect to explain anaphoric reference as grammatically guaranteed coreference. This coreference is in turn thought of as understood in terms of the primary, word-world sense of reference to the same extralinguistic item.

I will argue that an analysis of anaphoric mechanisms provides the resources for a purely *intralinguistic* account of the use of the English sentences by means of which philosophers attempt to make assertions about *extralinguistic* referential relations. More specifically, although we can and must distinguish between our words and what those words refer to or have as their referents, the truth of claims about what we are referring to by various utterances gives us no reason to believe that there is a *relation* of reference between expressions and the objects we use them to talk about. Following the lead of Wilfrid Sellars,¹ I will argue that 'refers' should

¹For instance "Truth and 'Correspondence'," this JOURNAL, LIX, 2 (Jan. 18, 1962): 29-56, and *Science and Metaphysics* (London: Routledge & Kegan Paul, 1963; New York: Humanities, 1968).

not be semantically interpreted by or as a relation and, *a fortiori*, not a word-world relation. Instead, 'refers' will be explained as a complex anaphoric pronoun-forming operator, by analogy (in the category of terms) to the analysis of '. . . is true' offered by the presentential theory of truth. To show this we first investigate the anaphoric roles that expressions can play. This leads to the specification of a new part of speech, *indirect descriptions*. Next, a formal test is offered for identifying expressions playing the anaphoric role of indirect descriptions, and 'refer' is explained as an operator that forms such descriptions. A paraphrase strategy is offered by means of which reference claims ostensibly of other forms can be wrestled into forms in which 'refer' appears only inside indirect descriptions. Finally, it is argued that doing this should be seen as explaining reference away.

I

In his seminal article "Reference and Context" Charles Chastain² suggests a novel approach to the understanding of singular-term reference. The basic concept he employs is that of an *anaphoric chain*, a notion best approached by example. Consider the discourse:

- (1) #A man in a brown suit approached me on the street yesterday and offered to buy my briefcase. When I declined to sell it, the man doubled his offer. Since he wanted the case so badly, I sold it to him.#

Two anaphoric chains are intertwined here, one corresponding to the buyer, and one to the briefcase:

- (2) A man in a brown suit . . . the man . . . he . . . him

and

- (3) my briefcase . . . it . . . the case . . . it

The reference of later elements in such chains (e.g., 'it' and 'the man') is secured only by the relation these elements stand in to the singular terms that initiate the chains in which they appear. This is the word-word (in fact, token-token³) relation of *anaphoric reference* or *anaphoric dependence*. The presence of an anaphoric chain in a discourse signals that not all its singular terms have reference

²In Keith Gunderson, ed., *Minnesota Studies in the Philosophy of Science*, vol. vii: Language, Mind, and Knowledge (Minneapolis: U of Minnesota Press, 1975), pp. 194-269.

³To be exact, it is tokenings which stand in anaphoric relations, although I haven't been fussy about this distinction in the text. Written expressions are particularly vulnerable to re-use of tokens, as when someone makes a sign with an arrow saving "He is a sinner," and walks around pointing it at passersby.

independently. Rather, some elements are related to their referents only in a derivative manner, in virtue of their anaphoric links to other expressions.

Examining the kinds of expression that can initiate and continue such chains enables Chastain to make two important related observations. The first concerns the significance of indefinite descriptions. Since Russell's discussions early in the century, indefinite descriptions have been treated as if they were not singular referring expressions at all, but were rather to be understood by means of a quantificational paraphrase. The presence of an indefinite description often does signal existential quantification rather than singular reference, but Chastain points out that the role of indefinite descriptions in anaphoric chains indicates that these expressions can also have a purely referential function. As in the example above, an indefinite description can initiate an anaphoric chain which may then be continued by pronouns or definite descriptions. And it seems clear that, in the context in which it occurs above, 'a man' purports to refer to a unique individual, namely the man in the brown suit who approached me on the street yesterday and eventually purchased my briefcase.⁴

This observation leads to Chastain's second point, which is that the reason that apparently nonquantificational uses of indefinite descriptions have not been thought of as straightforwardly referential is that they do not behave enough like proper names, the paradigm of singular terms. Except under deviant circumstances, if a proper name is used somewhere in a discourse invoking a particular referent, then other tokens of that same type which appear elsewhere in the discourse will be coreferential with it, in a sense which can be explained in terms of intersubstitution. In

- (4) #Leibniz has been called a pluralist, and he has been called a monist. But no one has ever thought of that philosopher as a materialist.#

the sense is not altered if we replace all the other elements of the anaphoric chain by the initiating expression to which they anaphorically refer. An inelegant redundancy is the only cost of replacing 'he' and 'that philosopher' by 'Leibniz'.

⁴Not all elements of anaphoric term chains need be understood as singular referring terms. Chastain says that quantificational, modal, and hypothetical contexts are "referentially segregating" and that the syntactically singular expressions that occur inside them should not in general be understood as singular referring terms. The claims of this essay don't require special treatment of such segregated occurrences.

In the case of an anaphoric chain initiated by an indefinite description, on the other hand, such a substitution of terms alters the sense of the sentences in which the substituted terms appear. Consider:

- (5) #A Republican senator threatened to filibuster the Wilderness bill.
 The senator's staff persuaded him that this action was unwise,
 so he left the chamber.#

The anaphoric chain here is

- (6) A Republican senator . . . The senator . . . him . . . he

The sense of this discourse is completely altered if we substitute the initiating expression for each of the terms that anaphorically refers to it:

- (7) #A Republican senator threatened to filibuster the Wilderness bill.
 A Republican senator's staff persuaded a Republican senator
 that this action was unwise, so a Republican senator left the
 chamber.#

In this passage the indefinite descriptions do not purport to co-refer. Each initiates a potentially new anaphoric chain. To continue such a chain requires either the use of a pronoun, which always continues an existing chain, or the use of a definite description, which can either initiate or continue a chain. The fact that a chain beginning 'a SORTAL . . .' cannot be continued by repeating the initiating phrase as is, but can be continued with a definite description of the form '*the* SORTAL . . .' is called by linguists the requirement of a "definitization transformation." One may conclude from such special requirements either that indefinite descriptions are not referring expressions or that not all referring expressions must behave like proper names. Chastain's suggestion is that we explore the second alternative.⁵

II

Singular-term tokens can play various roles in anaphoric chains. Such a token may initiate an anaphoric chain, as 'A Republican senator' does in (5). Or it may continue an existing chain and so depend for its referent upon an anaphoric antecedent, as "the senator" does in (5). Besides distinguishing tokens as anaphoric initiators and dependents, we can sort them according to two distinctions regarding the term types they instantiate. Chastain's considerations concern-

⁵ As Chastain points out, at the least this approach requires us to amend the simple intersubstitutability model of coreference, and the correlative distinction between extensional and nonextensional contexts.

ing substitution show that in dealing with anaphoric chains one may not in general assume that cotypical term tokens are coreferential, even in the absence of overtly indexical elements. Let us call those expressions which [according to substitution tests such as that of comparing (5) and (7) above] do not vary in reference from token to token within the type, *type-substitution invariant*. An example would be proper names as conceived and idealized by the tradition. Expressions that are referentially variable from token to token within the type may be described as not cotypically intersubstitutable. Pronouns would be a paradigm.⁶

The third distinction it will be useful to make is that between lexically complex expressions and those which are lexically simple, though perhaps grammatically complex; that is, between phrases that are nouns and words that are nouns. Consider, for instance, two varieties of cotypically nonintersubstitutable anaphoric dependents: dependent definite descriptions and personal pronouns. The lexically simple pronoun 'he' is limited, in the information it can give about its anaphoric antecedent and the chain of which it is a part, to a small number of dimensions such as gender and number, specified in advance by the grammar of the language. Lexically complex anaphoric dependents, on the other hand, can use the full descriptive resources of the language to give anaphoric information. This open-endedness permits dependents such as 'the senator mentioned above, who opposed the Wilderness bill and was dissuaded by his staff from expressing his stand'. The same contrast of lexical complexity applies to anaphoric initiators; so we may compare the indefinite description 'A Republican senator' with the proper name 'Leibniz' in (5) and (4).

Deploying these three independent functional distinctions—between anaphoric initiating tokens and dependent tokens, between type-substitution-invariant types and cotypically nonintersubstitutable types, and between lexically complex types and lexically simple types—yields eight roles that tokens can be thought of as playing in anaphoric chains. Thus among anaphoric initiators that are type-substitution invariant there are those which are lexically simple, such as proper names like 'Leibniz' in (4), and

⁶ It would be less clumsy to call the expressions not all cotypical tokens of which are coreferential "token reflexive." I do not, because that phrase has an established usage and is not generally thought of as applying to expressions like 'the senator', even when such expressions are used as anaphoric dependents, which are among the paradigmatic cotypically nonintersubstitutable term occurrences. It may be that the special substitution conditions applying to anaphorically dependent expressions is more explanatorily fundamental than indexicality of canonical token-reflexives.

those which are lexically complex, such as 'the first U.S. President'. Among the cotypically nonintersubstitutable anaphoric initiators there are again the lexically simple, such as 'this', and lexically complex indefinite descriptions, such as 'A Republican senator' in (5). Among the anaphoric dependents that are not cotypically intersubstitutable one can similarly distinguish lexically simple pronouns such as 'it' from lexically complex dependent uses of definite descriptions such as 'the man' in (1). Finally, among the type-substitution-invariant anaphoric dependents one can distinguish some uses of lexically simple proper names, as in

#I met a man I'll call "Binkley." Binkley is a mechanic.#

from lexically complex dependents, which will be called *indirect definite descriptions*.

With the exception of the last category mentioned, this tripartite division just rearranges familiar facts about the linguistic behavior of standard kinds of singular terms. The categorization was presented, however, to introduce the notion of indirect definite descriptions, a kind of singular term whose existence has not generally been recognized. Indirect definite descriptions are accordingly characterized as lexically complex, type-substitution-invariant anaphoric dependents. From this specification it follows that expressions in this category are complex pronouns, as are ordinary anaphorically dependent definite descriptions. Unlike such descriptions, however, all cotypical tokens of expressions in this category are guaranteed to be coreferential with each other, since they all anaphorically depend upon and hence corefer with a single common antecedent token, and so with each other. The idea is that an indirect definite description is a pronoun which actually contains a description specifying the term occurrence that is its anaphoric antecedent. Cotypical tokens of an indirect-definite-description type will contain the same description, and so (except in special cases) specify the same antecedent. One immediate expressive advantage of a language containing locutions of this sort would be that identities employing anaphorically indirect descriptions could be used to assert that two term *tokens* were coreferential, even if the tokens were of cotypically nonintersubstitutable types (such as indefinite descriptions or pronouns) for which, as Chastain showed, standard substitutional accounts of coreference fail since they presuppose type-substitution invariance.

A useful picture of the functioning of these expressions, the picture that motivates our calling them *indirect* descriptions, is offered by the indirect addressing function offered in most basic computer architectures. Ordinarily, the central processor uses addresses to

pick out values, just as we use descriptions to pick out objects. But, in indirect mode, the CPU when given an address as an input does not return the value stored at that address as its output. Instead it treats that value as another address and returns the value stored in that second address as its output. Indirect descriptions are to be understood by analogy to this two-step process. First, a token is specified, perhaps by being described as to type and spatiotemporal location. But the token thus picked out is not the referent of the whole indirect description. For next, an indirectness operator is applied to that token specification to produce the indirect description, which only *anaphorically* refers to the specified token and so, as a whole, refers not to that token, but rather to whatever that token, its anaphoric antecedent, refers to, just as with ordinary pronouns. The flexibility of the von Neumann computer architecture is in large part due to its capacity to treat the same expression both as datum (that is, as a value) and as instruction (the address at which the value can be found). Indirect descriptions exploit the analogous use/mention amphibiousness made possible by anaphora, and in this fact, I want to claim, consists the expressive dividend *semantic* vocabulary pays in a language to which it is added.

To be entitled to claim that there actually are expressions in natural languages which should be understood as playing the anaphoric role just abstractly described and to see what indirect descriptions have to do with specifically *semantic* vocabulary, we must look at some examples. Consider a discourse in which Joe N. says:

- (9) #I should have known better than to let the mechanic Binkley work on my car. That airhead misadjusted the valves . . . #

Suppose that, later, Jim, forgetting the name Joe used, says:

- (10) #For car repair, don't go to the mechanic Joe N. referred to as "that airhead."#

How are we to understand this latter remark, and in particular the singular term

- (11) the mechanic Joe N. referred to as "that airhead" ?

Clearly this term refers to Binkley, Joe N.'s hapless mechanic. But how is this reference secured? The most obvious way to interpret such a singular term is as a straightforward definite description, by analogy to 'the mechanic who worked on Joe N.'s car and misadjusted the valves'. In both cases some purportedly unique feature of Binkley is used to single him out—his relation to Joe N. either in

being referred to by him in a certain way or in having abused his car in a certain way.

But the anaphoric category of complex cotypically nonintersubstitutable dependents and the brief discussion of anaphorically indirect definite descriptions suggest that an alternative analysis may be more illuminating. For we can think of (11) as being an anaphoric dependent, having Joe N.'s original token of 'that airhead' as its anaphoric antecedent. If tokens of (11) are anaphorically dependent on the original token of 'that airhead', then they are co-referential with it and, hence, refer to Binkley the mechanic. On this account, (11) should be thought of as referring to Binkley in the way a token of 'he' would, if Jim could arrange to ensure that the antecedent of that token of 'he' were Joe N.'s token 'that airhead'. Pronouns, as simple cotypically nonintersubstitutable anaphoric dependents, can take such antecedents if the antecedent and dependent tokens are sufficiently close to each other in time, space, or audience attention. But for distant antecedents, one may not simply rely on the meager resources grammar gives us to work backward from a simple dependent token such as 'he', which even with contextual supplementation can give us only so much information about its antecedent. Here, according to the current suggestion, is where indirect definite descriptions enter. For these locutions are grammatically complex, like ordinary definite descriptions, and enable the use of the full descriptive resources of the language to specify the antecedent token to which they are anaphorically linked.

In our example the antecedent token is specified as that token whereby Joe N. referred to someone as "that airhead," the token directly picked out by the phrase "Joe N.'s utterance of 'that airhead'." Knowing what the individual was referred to *as* tells us the *type* of the antecedent token. Invoking Joe N. locates the particular token of that type which is in question. The presence of 'refer' marks the indirect-addressing feature, by which we are to understand the referent of the whole description to be not the term token picked out as anaphoric antecedent, but rather, as with simple pronouns, the referent of that antecedent token. Indirect definite descriptions like (11) should be understood as complex pronouns (anaphoric dependents), and 'refers' and its cognates should be understood as complex anaphoric pronoun-forming operators.⁷

⁷One difference that might be remarked between ordinary pronouns and those formed by indirect description concerns backwards anaphora, in which the anaphorically dependent occurrence precedes its antecedent in the discourse. Such cases are unusual, though by no means always deviant or strained, for ordinary depend-

The 'refer' cognates consist of all the sorts of expressions that we would normally think of as being used to assert semantic word-world relations. Thus the following examples ought to be understood according to the model of indirect or anaphoric descriptions: 'the philosopher N. mentioned yesterday', 'the restaurant he talked about at the committee meeting', 'the difficulty discussed above', 'the person denoted by the second name on the list', 'the criminal described by the police in the morning paper', 'the referent of the expression you are pointing to', and "the advisor Kissinger characterized publicly as 'almost a third-rate intellect'."⁸

III

Although there are important differences among these examples, all of them could be paraphrased so as explicitly to use some form of 'refer'. But even this rough characterization is only of use insofar as it is possible to say what is special about the functioning of 'refer' which would enable one, for instance, to tell whether some alien language possessed an expression playing an analogous role. Putting the question more generally, even if it turns out that we can properly account for the behavior of expressions like those in our examples according to the indirect-addressing model of anaphoric descriptions, how could we explain and justify enforcing such a radical distinction between the analyses of descriptions as apparently analogous as (11) and

- (12) the one Joe N. startled (insulted, deafened) by his remark about
airheads ?

Doesn't the most intuitive reading of (11) assimilate it to (12), treating both as ordinary definite descriptions of a man, who in each case happens to be picked out by his relation to some utterance by Joe N.? What difference between these cases *makes* the difference in virtue of which (11) should be treated as an indirect description

ents like 'he' and 'the man'. Indirect descriptions, in virtue of the explicit way they pick out the tokens they depend upon anaphorically, exhibit no such prejudice for the discursive past and, accordingly, often possess "antecedents" only in the broader sense of anaphorically inheriting content from another tokening.

⁸This last sort of example shows that the term token can be picked out in a variety of ways, in particular, by citing a predicate type to pick out a sentence token which is the characterizing and which contains the term token (perhaps a tokening of 'Richard Allen') which on the present account is the anaphoric antecedent of the indirect description in the example. We may also note that *indefinite* descriptions may similarly be constructed from indirect anaphoric sortals, as in "A woman the lawyer referred to as 'she who must be obeyed' explained the matter to us," which both initiates a new chain *and* characterizes the referent of that chain by anaphoric relation to some antecedent tokening of 'she who must be obeyed' by the lawyer in question. See the discussion of referential predication below.

that essentially involves an anaphoric link, whereas (12) should be treated as an ordinary description using a relation to an utterance to pick out an object? What is the crucial difference between being referred to by a certain token and being startled, insulted, or deafened by it?

The clearest manifestation of the difference in question concerns the *iteration* of pronoun-forming operators. Because the relation “. . . is an anaphoric dependent of . . .” is transitive, any operator that takes a term token and produces an expression that anaphorically depends upon it should be iterable without change of resulting reference. For tokens of the complex pronoun formed by applying the indirectness operator to (a token of) the result of applying that operator to an original initiating token should simply continue the anaphoric chain—dependents of dependents having the same original or ancestral antecedent.

Consider such iteration as applied to our examples (11) and (12) above. If we iterate the description-forming operators that produced these, with suitable variation of speakers, we get:

(13) the one John M. referred to as “the one Joe N. referred to as ‘that airhead’”

(14) the one John M. startled by his remark about the one Joe N. startled by his remark about airheads

If descriptions formed in the appropriate way from ‘refers’ are indirect descriptions, then, in virtue of the transparent iterability of anaphoric dependence, we should expect (13) to be coreferential with (intersubstitutable for) (11) and, hence, with Joe N.’s original token of type ‘that airhead’. And so they are, issues of speaker’s reference aside.⁹

But though (11) and (13) are coreferential *de jure*, the superficially analogous (12) and (14) would be coreferential only by accident and under special circumstances.¹⁰ These considerations can be formulated as the *iteration condition* (15) below, which is a necessary condition for understanding an operator PF as a pronoun-

⁹ Speaker’s reference will be discussed briefly below as an anaphoric phenomenon.

¹⁰ It may seem that the presence of a sortal restriction on indirect descriptions causes difficulties. The issue can be avoided as in (11) above, however, by using an *anaphorically dependent* sortal. ‘One’ anaphora has long been recognized by linguists as permitting anaphoric proforms as stand-ins for common nouns, as in “There were red pens as well as green ones on the table.” In fact philosophers have made up the expression ‘referent of *t*’ to mean “the one referred to by *t*.” So the fact that indirect descriptions are sortally restricted, as are ordinary descriptions or quantifications in natural language, adds no new difficulties to an anaphoric analysis of ‘refers’.

forming operator. If we agree to surround a term designation with angle brackets to form a designation $\langle \text{term} \rangle$ of the *type* of that term, and to surround such a type designation with subscripted slashes to form a designation $\langle \text{term} \rangle_i$ of a token of that type, then the iterability condition can be put as follows:

$$(15) \quad \text{PF}(\langle \text{term} \rangle_i) = \text{PF}(\langle \text{PF}(\langle \text{term} \rangle_i) \rangle_j)$$

It is clear that nothing can correctly be thought of as a pronoun-forming operator unless it meets this condition embodying the transitivity of anaphora. Our strategy will be to exploit such an iterability requirement, suitably qualified, as a *sufficient* condition for identifying operators that form indirect descriptions (which have been explained as type-substitution invariant, lexically complex pronouns).

Enough weight will be placed upon the strategy of transforming the iterability condition from a necessary to a sufficient condition for interpreting a syntactically relational expression as a complex pronoun-forming operator to make it worth stating precisely. The question the test is supposed to answer is this. Suppose we are given a construction that on the surface has the form

$$(16) \quad \text{the } x[\text{REF}(x, \langle \text{term} \rangle_i)]$$

The over-all expression appears to be a definite description that picks out a thing x by means of its relation REF to a token $\langle \text{term} \rangle_i$ of type $\langle \text{term} \rangle$. An example would be

$$(17) \quad \text{the man who was frightened when Bernadette uttered a token of type 'BOO'}$$

The iterability test tells us to consider (16) and

$$(18) \quad \text{the } y[\text{REF}(y, \langle \text{the } x[\text{REF}(x, \langle \text{term} \rangle_i)] \rangle_j)]$$

as in

$$(19) \quad \text{the man who was frightened when Bernadette uttered a token of type "the man who was frightened when Bernadette uttered a token of type 'BOO'"}$$

and (13) and (14).

The claim is that the syntactically relational expression REF here should be understood *not* as standing for a relation, as in an ordinary definite description, but as an anaphoric operator forming in-

direct descriptions, if and only if the following three conditions are met¹¹:

- (20) *If* (16) is a proper description (that is, in fact picks out one and only one object) and if (18) is a proper description, *then* they corefer (are intersubstitutable).
- (21) *If* the token by relation to which the individual in (18) is picked out *were not* of the same type as the expression (16), *then* (16) and (18) would *not* in general corefer.¹²
- (22) Accepting the identity statement by means of which the coreference of (16) and (18) is asserted is not accidental, in the sense that it is a condition of being taken to understand the expressions involved.

The first condition is required because expressions of type (18) need not always pick out unique objects even when the expression of type (16) does. There need be no one who is the man Bernadette frightened by uttering a token of type “the man Bernadette frightened by uttering a token of type ‘BOO’”, and similarly for genuinely anaphoric cases. Iterability is a relevant test only in the (in general counterfactual) situations where the appropriate individuals exist. The second condition is required in order to rule out cases where the same individual is picked out no matter what token one looks at—the case where one and only one man is frightened by *whatever* Bernadette says. The type of the intermediate antecedent of an anaphorically dependent expression is obviously essential to its having the reference that it has; so this condition represents a natural constraint. The third condition is required in order to rule out grammatically accidental coreference of (16) and (18), as might happen in a psychologically homogeneous population with the relation “is the first object one is reminded of on hearing the expression ⟨term⟩”.¹³ Together these three conditions ensure that any expression ‘REF’ which satisfies them may appropriately be understood as forming expressions of type (16) which ought to be under-

¹¹The identification of indirect descriptions by the iteration test provides us with a sense in which one term *type* can be anaphorically dependent on another type, as the iterated indirect description is on the type of its antecedent token. But this notion of type-anaphora is entirely derivative from the basic notion of token-anaphora, a derivation made possible by the existence of operators that form lexically complex, type-substitution-invariant pronouns.

¹²It may seem that talk here of “picking out”, “coreferring”, and so on begs important questions. But it will be shown below that such talk can be understood in a way that does not commit one to reference relations.

¹³Joe Camp pointed out the need to deal with these relations which “piggy-back” psychological relations on semantic ones.

stood as really having the form of complex anaphoric dependents, that is, as indirect definite descriptions like (11) rather than as ordinary direct descriptions like (12). Frege placed great theoretical weight on the intersubstitutability of the terms ‘*t*’ and “the Bedeutung (or referent) of ‘*t*’”,¹⁴ and this same essential feature of referring lies at the center of the present account.

IV

The account so far has described the anaphoric category of indirect descriptions as a form of pronoun, has offered a formal test discriminating operators that generate expression types of this category, and has pointed out that ‘refers’ as it appears in contexts like

(23) the one Kissinger referred to as “almost a third-rate intellect”

can be understood as such a complex pronoun-forming operator. But there are other important uses of ‘refers’ and its cognates. The most fundamental of these are: “Tarskian” contexts, such as

(24) ‘Rabbits’ refers to (denotes) rabbits.

denials of reference, such as

(25) (The expression) ‘the present King of France’ does not refer (or refers to no one).

mere reference claims, as in

(26) During his talk the speaker referred to Napoleon.

and referential predications, as in

(27) The speaker talked about shadowy figures from the intelligence community.

The strategy is to approach these locutions in two stages. First, each such usage is paraphrased into a form in which the ‘refers’ cognate appears *only* inside an indirect description. Then that description is explained as functioning as a complex pronoun, according to the story already told. The present concern is thus with the paraphrase in terms of indirect descriptions.

The generalization of the token-based account of indirect descriptions required for Tarskian contexts is really a simplification to a special case. For statements about what a term refers to or denotes presuppose that the term type in question is type-substitution in-

¹⁴ See the discussion in the author’s “Frege’s Technical Concepts,” forthcoming in *Synthese*.

variant—that all cotypical tokens corefer. So we can parse

(28) The term 'Leibniz' denotes Leibniz.

as an identity involving an indirect description¹⁵:

(29) The one denoted by the term 'Leibniz' is (=) Leibniz.

Given the presupposition of type-substitution invariance of the type 'Leibniz' in the original claim, this indirect description is equivalent to

(30) the one denoted by any token of 'Leibniz'

which we may then straightforwardly understand as a complex pronoun, anaphorically dependent upon an antecedent which may be any token of the specified type, e.g., that which appears on the other side of the identity sign. The felt triviality of such reference claims is thus explained.¹⁶ Of course, type-substitution-invariant reference claims need not be epistemically trivial if different languages or different term types are involved, as in

(31) In our world (the expression) 'the first Postmaster General' refers to Benjamin Franklin [or: the inventor of bifocals].

which we understand as using anaphoric relations to claim

(32) In our world, the one referred to as "the first Postmaster General" is (=) Benjamin Franklin.

which is not a trivial assertion.

Such an account respects the different modal status of (28), which is contingently true, and "Leibniz is Leibniz", which is necessarily true. For the possibility that (28) is not true can be understood in terms of the paraphrase (29) as the existence of a possible world *w* such that the one referred to as 'Leibniz' in *w* is not Leibniz (that is, the one *we* refer to in our own world as "Leibniz"). The explicit relativization of the indirect description to a possible world *w* simply tells us which world its antecedent tokens are to be found in. The candidate antecedents of "the one referred to as 'Leibniz' in *w*" are tokenings of 'Leibniz' uttered in *w*. The anaph-

¹⁵ Compare Frege's similarly motivated paraphrase of "Jupiter has four moons" into "The number of Jupiter's moons is (=) four" in section 57 of the *Grundlagen*.

¹⁶ If one sets up Tarskian truth conditions using denotation claims for semantic categories besides terms, these can be accommodated as well by this scheme (although it is no part of the present project to discuss other parts of speech). Here one thinks of reading "'Red' refers to (denotes) red things" as "the ones referred to as (denoted by) 'red' are red things", and "'magnetic' applies to (has in its extension) magnetic things" as "the ones 'magnetic' applies to are magnetic things".

oric approach thus has room for what has been thought of as the contingency of word-world semantic relations, although it does not appeal to such relations.

Coreference claims represent a simple variation on Tarskian contexts, and can be interpreted in much the same way. Thus to say that the expressions $\langle \text{type}_1 \rangle$ and $\langle \text{type}_2 \rangle$ corefer is just to assert an identity between the corresponding indirect definite descriptions, i.e., to say that the one referred to by $\langle \text{type}_1 \rangle$ is the one referred to by $\langle \text{type}_2 \rangle$. Asserting such an identity is licensing as commitment—preserving the intersubstitution of expressions of those types. In the case of cotypically nonintersubstitutable expressions such as demonstratives, asserting an identity authorizes substitution of anaphoric dependents of the token on the left for anaphoric dependents of the token on the right of the identity, and vice versa.

It may be worth noticing that, if attention is restricted to type-substitution-invariant term types, the iteration test introduced above can be simplified correspondingly, by omission of token specifications. The necessary condition for operators PF to form complex type-substitution-invariant pronouns is then:

$$(33) \quad \text{PF}(\langle \text{PF}(\langle \text{type} \rangle) \rangle) = \text{PF}(\langle \text{type} \rangle)$$

and the corresponding condition suggested as sufficient for REF to be an indirect description-forming operator is

$$(34) \quad \text{the } y[\text{REF}(y, \langle \text{the } x[\text{REF}(x, \langle \text{type} \rangle) \rangle])] = \text{the } x[\text{REF}(x, \langle \text{type} \rangle)]$$

Simple negations of statements of reference, as in

$$(35) \quad (\text{The expression}) \text{ 'the shortest man in the room' does not refer to John.}$$

raise no new issues, for we have already explained the underlying identity that is being negated. But claims such as (25), that an expression does not refer to *anything*, deserve special mention. The obvious way of extending to these cases the previous strategy of paraphrasing what look like assertions of reference relations as identities involving indirect descriptions is to quantify into the identity and read the result as a negative existential statement. That is, statements of the form

$$(36) \quad \langle \text{type} \rangle \text{ does not refer.}$$

are to be read as

$$(37) \quad \text{The one referred to as } \langle \text{type} \rangle \text{ does not exist.}$$

where this last is to be understood just as we understand negative

existentials involving ordinary direct descriptions, as in

(38) The present King of France does not exist.

That the indirect definite descriptions involved in denials of referentiality are anaphoric dependents makes no difference to our reading of the negative existentials; we have no trouble understanding remarks like

(39) I would be comforted by the benevolence of a supreme being, except that such a being does not exist.

in which 'such a being' is an anaphoric dependent.

Statements like (26) tell us that reference has taken place, but give no information about what the referring tokens or types were. Such remarks may be understood as asserting that there is *some* term tokening *t* in the speaker's discourse such that the item referred to by *t* is Napoleon. Statements like (27) are similar, except that a predication, rather than an identity, is what is asserted of the items referred to or talked about. (27) says that there were tokenings *t* and *t'* and perhaps more, such that the items referred to (or talked about) by *t* and *t'* have the property of being shadowy figures from the intelligence community. These are predications involving pronouns, intrinsically no more mysterious than sentences like "They are confused." Common nouns can be formed from indirect descriptions just as they can be from ordinary direct descriptions, and the present account extends straightforwardly to these expressions, as in

(40) All the animals the speaker mentioned tonight were quadrupeds.

This example indicates as well how *generalizations* about reference are to be approached anaphorically. Endorsement of (40) commits one to all the substitution instances of the form "If *t* is an animal the speaker mentioned tonight, then *t* is a quadruped." The antecedent of each such conditional is a referential predication, equivalent to "There is a term token /*s*/ such that the speaker uttered /*s*/ tonight and the item referred to by /*s*/ is an animal and the one referred to by /*s*/ is (=) *t*," which we already know how to interpret. As long as we can in this way understand each of the substitution instances to which a universally quantified claim commits us, we can understand the universal generalization itself. A full discussion of such cases would require an account of anaphora and quantification which goes beyond the scope of this essay. The complications arise in part from the recognition that in the general case the term substituent *t* which is repeated in the specifi-

cation above of the form of each sentential substitution instance of the quantification would need to be replaced by two term tokens, one of which is anaphorically dependent upon the other. The present point is that, although an account is not being offered of quantification in general, it is clear from the example that no new difficulties are added by the presence of indirect descriptions in the quantificational substitution instances, so that generalizations about reference can be understood if any sort of generalizations can.

Consideration of generalizations about what is referred to by various expressions makes salient another issue, which can only be dealt with in passing here, namely the susceptibility of an anaphoric account of reference claims, to the formulation of semantic paradoxes. In the presumably analogous case of truth, a naive substitutional understanding of quantification into truth claims commits one to interpreting paradoxical sentences such as the Liar. Of course generalization is not the only way in which such paradoxical expressions can arise, nor is the possibility of semantic paradox restricted to the category of sentences. It is possible to use 'refers' to formulate empirically paradoxical term tokens such as

- (41) the square root of 2 which is the result of multiplying -1 by the one referred to by the term token numbered '41'

where 'square root of 2' is a sortal comprising the positive and negative roots and 'one' is understood as a prosortal anaphorically dependent upon it. Interpreting such tokens as anaphorically indirect descriptions focuses attention on grounding conditions for anaphoric inheritance—a large and important topic. In "Inheritors and Paradox" Dorothy Grover¹⁷ elaborates an anaphoric approach to semantic paradoxes for the closely analogous anaphoric treatment of '. . . is true.' Grover finds that the natural condition on anaphoric grounding yields an interpretation coinciding in general with the sentences that Saul Kripke assigns a semantic value to at the minimal fixed point, the interpretation he takes to provide the most natural model for the intuitive concept of truth. Her remarks can be applied to the present construction by means of the crucial

¹⁷This JOURNAL LXXIV, 10 (October 1977): 590–604. Also in this paper Grover independently states a weaker version of the observation which is exploited in the present account: "Descriptive phrases such as 'just mentioned', 'been talking about', 'are referring to', which ostensibly describe discourse may often be used merely to locate an antecedent piece of discourse from which a referent is inherited" (594). See also Grover's "'This Is False' on the Prosentential Theory", *Analysis*, xxx.2, 170 (January 1976): 80–83.

analogy between this pronominal account of reference and the presentential account of truth which inspires it.

V

The task of this paper has been to make it plausible that the use of 'refers' and cognate locutions in natural languages can be understood by first paraphrasing contexts in which they occur into a form in which they appear only inside indirect descriptions and then understanding their role in those paraphrases as operators taking token (or type) specifications and forming from them lexically complex, type-substitution-invariant pronouns whose anaphoric antecedents are the specified tokens. As just indicated, such an account of 'refers' should be compared to the presentential account of 'true' due to Grover, Joseph Camp, and Nuel Belnap.¹⁸ For, following the lead of that account, '. . . is true' can be seen as an operator that forms *prosentences*.

The presentential approach points to two uses of pronouns, lazy uses in which the pronouns can be replaced without loss of content by (a transform of) their antecedents as in (4), and quantificational uses in which their antecedents determine a *class* of admissible substituends, as in "If any object gravitates, then it has mass." So too *prosentences*, which, by analogy to indirect descriptions, we can understand as formed by using '. . . is true', have both lazy and quantificational uses. If I say "Snow is white," and you say "That is true," your remark would have had just the same content if you had simply repeated mine, save for the disclaimer of originality always involved in anaphora. It is for these "lazy *prosentences*" that F. P. Ramsey's redundancy approach to truth was developed. "Everything the policeman said is true," can be rephrased as "If the policeman said it, then it is true," and 'it is true' may then be understood in turn as functioning as a *prosentence* of quantification, by analogy to the second role of pronouns. The innovation in the presentential account is to see the entire expression 'that is true' or 'it is true' as an anaphoric dependent in the grammatical category of sentences, instead of (or as well as) seeing 'that' and 'it' as anaphoric dependents in the grammatical category of singular terms. In this way, it is argued, the uses of '. . . is true' in natural languages can be accounted for.

¹⁸ Grover, Joseph L. Camp, and Nuel D. Belnap, "A Presentential Theory of Truth," *Philosophical Studies*, xxvii, 2 (February 1975): 73-125. The ensuing discussion presupposes the exacter analogy made possible by revising the presentential theory so as to see ". . . is true" as a *prosentence*-forming operator rather than as a syncategorematic fragment of an indissoluble *prosentence* such as 'that is true'. I elaborate such a revision in "Sentential Anaphora and Truth", ms.

The present theory of 'refers' simply does for terms what the presentational theory did for sentences (although we have not discussed quantification). Each theory explains the use of a bit of *semantic* vocabulary in terms of the formation of *anaphoric* proforms. The uses of the sentences by means of which we meant to be making claims about what objects in the world our talk refers to, and what it is in the world in virtue of which what we say is or is not true, are explained in terms of the purely *intra*linguistic relation of anaphora.¹⁹

Indirect descriptions formed from 'refer' *both* mention a term expression (in picking out anaphoric antecedents) *and* use that expression. The effect of applying an indirect description-forming operator to a mentioned term is that of *turning* the mentioned occurrence into a used occurrence. Thought of in this way, 'refers' is an *anaphoric disquotation operator* in the same sense that 'true' is.²⁰ Further argument would be required to maintain that the reference claims so formed deserve to play a basic explanatory role in giving an account of the *use* of term expressions. For it would seem that disquotation operators simply presuppose the features of the use of unquoted expressions which the linguistic theorist takes as explanatory target.

There is one potentially significant sort of locution which does not receive an analysis on this approach. For although accounts are offered of what someone referred to by an utterance and of what the reference of the utterance was, nothing is said about the *relation* of reference. The anaphoric approach will not tell us how to understand sentences such as

(42) Reference is a physical, causal relation.

The reason is clear. On the anaphoric account although '... refers to--' plays a syntactically relational role, its semantic role is anaphoric and pronominal rather than relational. Philosophers have misconstrued the plain man's use of 'refers' and hypostatized a relation of reference as the semantic interpretant of the apparently rela-

¹⁹As an example of the sort of philosophical question raised but not answered by the possibility of such an anaphoric account of semantic vocabulary, consider the distinction between substitutional quantification and the referential or objectual variety. Although the technical differences are clear between reading a quantifier in terms of an algebraic model via an interpretation function and by looking at linguistic substituends, the standard Quinean account of what turns philosophically on this technical difference seems to be undercut by the possibility of an anaphoric understanding of the interpretation statements involving 'denotes' or 'refers'.

²⁰The specially anaphoric nature of the disquotation involved is most apparent in cases where the indirect description picks up the speaker's reference of its mentioned antecedent.

tional expressions, and have then asked for a theory of it. Such a mistake is of a piece with the search for the objects corresponding to each expression that syntactically plays the role of a term—e.g., ‘someone’ and ‘everyone’. Our ordinary remarks about what some individual or expression refers to or has as referent are perfectly in order, and the anaphoric account tells us how to understand them. But reference is a philosopher’s reification and a fiction, generated by a grammatical misunderstanding. The present account distinguishes sharply between expressions and their referents (what is referred to by them), in that very different uses are associated with the expressions “the expression ‘Leibniz’” and “the referent of the expression ‘Leibniz’.” No doubt, as with any other two items in the causal order, there are many relations that can correctly be said to obtain between a term token as it stands and what it refers to. But the present considerations show that our talk about referring and referents gives us no reason whatsoever to conclude that some one of these could be singled out as *the reference relation*, that unique semantically significant word-world relation in virtue of which the non-expression is the referent of the expression. Various word-world relations may play important explanatory roles in philosophical projects, but to think of any one of these as what is referred to as “the reference relation” is to be bewitched by surface syntactic form.

In order to see what is and what is not being claimed for this analysis of ‘refers’ and to see the significance of the replacement of a relational account of reference with an anaphorically disquotational account of referring, it will help to redescribe and clarify the explanatory role that the concept of indirect descriptions is supposed to play in that analysis. Consider first a language-in-use such as English, and produce its nonsemantic fragment by extruding from it all the sentences containing semantic vocabulary (‘refers’ and its cognates for the present account, and ‘true’ as well if we appeal to a prosentential theory). Suppose that a theorist has succeeded in capturing the social practices governing the use of this *nonsemantic* fragment, at least in the sense that the theory attributes to each interlocutor a stock of sentences that this interlocutor is prepared or committed to assert and a set of inferences he is bound to endorse, and describes how assertion and inference proceed given these attributions.²¹ Then the anaphoric account of ‘refers’ developed here permits the extension of such an account of the

²¹ See for instance my “Asserting” *Noûs*, xvii, 4 (November 1983): 637–650.

use of the *nonsemantic* fragment of the language to an account of the use of the whole language.

First, an artificial operator REF stipulated to meet the iteration test can be introduced, by means of which new singular terms can be formed on the model of (16). These new terms can then be understood, following the discussion of anaphoric categories, as indirect descriptions. Understanding them in this way determines their role in assertion and inference. For one is committed to endorsing “the one referred to by $\langle \text{term}_1 \rangle$ is $\langle \text{term}_2 \rangle$ ” just in case one is already committed to the nonsemantic substitution-licensing identity “ $\langle \text{term}_1 \rangle = \langle \text{term}_2 \rangle$,”²² and similarly for referential predications. Next, using the paraphrase strategy outlined earlier, the sentences of the semantic fragment of the original language can be translated into the sentences of this artificial extension of its *nonsemantic* part. The claim is then that such a translation will correctly predict the inferential endorsements and assertional commitments in the semantic fragment, on the basis of only those endorsements and commitments involving *nonsemantic* vocabulary. In this sense the semantic fragment of the language is reduced to its nonsemantic base by the anaphoric analysis.

It is at this point that a theorist seeking a better account of the pragmatic dimension of his base language might introduce a second artificial operator SP_{REF} expressing speaker’s reference. Such an operator would be defined in terms of the basic semantic/anaphoric notion of reference as captured by REF. Although SP_{REF} would not be required to pass the iteration test [the equivalence of (13) to (11) clearly breaks down if we let ‘refers’ include speaker’s reference], the use of indirect descriptions such as “the one Joe speaker-referred to as ‘that fat politician’” will still be determined by the use of the nonsemantic base language, in virtue of an account of SP_{REF} in terms of REF. The idea is that it is incoherent to take someone as having speaker-referred to an object unless one also takes it that that individual *could have* referred to it “semantically.” Had I known there might be a question about whether there was champagne or ginger ale in the glass, I would have used the description “the man in the corner with bubbly liquid in his glass” instead of a more daring description. And it is these safe or minimal descriptions standing behind speaker’s references which should be seen as the ultimate anaphoric antecedents of indirect descriptions such as “the one speaker-referred to as ‘the man in the corner with champagne in his glass’”, which themselves depend

²²Type-token niceties are brushed over here.

upon such covert or merely virtual tokenings [just as 'that airhead' anaphorically depends upon Joe's tokening of 'the mechanic Binkley' in (9), and so would refer to Binkley even if he were not in fact an airhead]. So parsing indirect descriptions formed using *SP REF* just requires an extra step in determining the anaphoric antecedent, but otherwise raises no difficulties for the translation scheme into artificially extended nonsemantic English.

Suppose that semantic talk can in this way be reduced to nonsemantic talk by anaphoric links between the two kinds of talk secured by indirect descriptions, without invoking a relation of reference (or a property of truth). It might be thought that all such an intralinguistic reduction establishes is that *if* one can give an account of the use of the nonsemantic portion of the language without invoking a reference relation, *then* one can extend it to an account of semantic talk about referring without needing to *add* a notion of reference to account for our apparent talk about such a relation. The possibility may then seem open that a notion of reference is still required to account for the use of the nonsemantic fragment appealed to by disquotation. Such a response would be a mistake. For what is envisaged is that, in order to explain how nonsemantic sentences of the object language are used, the linguistic theorist must make claims in the metalanguage which specify what object-language expressions refer to. But the anaphoric account will apply equally to metalanguage statements of the form

(43) $\langle \text{term} \rangle$ in OL refers to electrons (ML).

No reference relation need be invoked to understand the role of these metalanguage semantic claims any more than for the semantic fragment of the object language. Understood by their paraphrases in terms of indirect descriptions, claims like (43) add the expressive capacity to pick up object-language antecedents with metalanguage anaphoric dependents, and so permit the absorption of the language under investigation into the theorist's use-language. The effect of adding to ML expressions of the form *the one referred to in OL by $\langle \text{term} \rangle$* is to permit the expression in ML of identities and predications involving OL expressions, including cotypically nonintersubstitutable ones or those used to speaker-refer.

Has reference been explained away? There are four reasons for which one might think that it has not. First, it might appear that an intralinguistic account of referring is committed to a sort of linguistic idealism which cuts language off from the world. But the present account distinguishes, e.g., the expression 'Leibniz' from the one referred to by that expression, and does not treat the latter

as in the linguistic order. Expressions do refer to nonlinguistic items (it is true that the one referred to by (Leibniz) is not a word and was a person). It is claimed only that it is a mistake to think that there is some relation of reference being invoked when we say this. Second, one might worry that reference had been smuggled into the account of indirect descriptions. For that account required that the antecedent token be described, picked out, or otherwise referred to before an indirectness operator is applied to produce a pronoun dependent on that antecedent token. Even though the things referred to are always term tokenings and hence in the linguistic order as well as in the causal order, haven't referential relations been presupposed? They have not. For the anaphoric account tells us just how to understand claims about *what is referred to as "Joe's utterance yesterday of the term 'that airhead' "*. The starred expression is just another indirect description. This response points to the third worry, and to a genuine qualification. For the notion of anaphoric dependence upon which the whole account rests has not been explained, but only introduced by example. Unless anaphora can be explained without invoking a relation of reference, the attempt to use it to explain away reference will be circular. I believe that such an account can be given in terms of inheritance by the dependent token of identificatory and, hence, substitution-inferential commitment from the antecedent token; but justifying such a claim is outside the scope of this essay. So reference is explained away only taking for granted a notion of anaphora.

The final reason one might have for denying that the dispensability of a relation of reference has been established by the foregoing considerations is sociological. If there is no relation of reference, surely some account is required of what has misled us into looking for such a relation as the core of an explanation of linguistic practice. Put another way, the story told here about the expressive function of indirect descriptions (and hence of claims about what refers to what) seems to preclude their use in *explanations* of nonsemantic linguistic practice, on grounds of circularity. For, understood anaphorically, metalanguage reference claims will presuppose the object-language uses which are the antecedents of those metalanguage indirect descriptions and, hence, cannot be used to explain those uses. Yet it is just to achieve such explanations of the nonsemantic practices of the language that theorists have always cared about the semantic relations here explained away anaphorically. It is in this context that the anaphoric account presents us with a closing dilemma. *Either* it must be possible to

offer an account of the project of explicitly codifying in a theory the assertional and inferential practices that constitute the use of a language in such a way that it becomes clear why the special expressive resources of anaphoric proform-forming operators in the theorist's language are crucial to that project, *or* it must be possible to explain what it is about the functioning of this anaphoric vocabulary which has *mised* theorists into attempting to use that vocabulary to express the quite different notions that really *are* crucial to that explanatory project.

ROBERT BRANDOM

University of Pittsburgh

THE PLAUSIBILITY OF RATIONALISM*

THE rationalist-empiricist debate that preoccupied so many linguists, philosophers, and psychologists during the late sixties ran out of steam in the mid-seventies when both parties eventually tired of unsuccessful attempts to convert each other. In this paper I wish to reopen the discussion of issues raised in that debate, because I believe that we are now able to formulate these issues with sufficient clarity to be able to assess the plausibility of rationalism. Specifically, I believe that we can now provide strong arguments for the plausibility of the so-called "innateness hypothesis." Such vindication seems especially timely, since there is a growing number of psycholinguists who believe that according semantics a substantial role in syntax acquisition makes unnecessary the rationalist assumption that the learner has rich innate knowledge about the class of possible natural languages. In defending the innateness hypothesis, I distinguish the commitment to rich innate structure from the concomitant rationalist commitment to an intentional and, more specifically, propositional-attitude account of language acquisition. It is, I argue, an open empirical question whether such accounts of language acquisition can satisfy the usual criteria for explanatory adequacy; the correct account may be neither empiricist nor rationalist. I begin by stating what seems to me to be at issue between rationalists and empiricists.

* An earlier version of this paper was presented at the annual meeting of the Society for Philosophy and Psychology, March, 1980. In revising the paper I have benefited from the criticisms and suggestions offered by several people, especially Ned Block, Noam Chomsky, William Demopoulos, Robert May, and Stephen Stich.