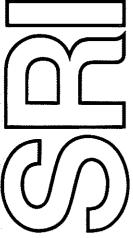# THE DESCRIPTIVE APPROACH TO REFERENCE: WHY IT IS DIFFICULT TO WORK WITH, AND WHY WE HAVE TO

Technical Note 435

April 20, 1988

By: Amichai Kronfeld

Artificial Intelligence Center
Computer and Information Sciences Division

APPROVED FOR PUBLIC RELEASE: DISTRIBUTION UNLIMITED

The Descriptive Approach to Reference:
Why it is Difficult to Work with,
and Why We Have to*

Amichai Kronfeld

April 20, 1988

Contents

1

1 The problem of reference

Suppose I tell you that my cat is hungry. There are four distinct entities, each participating in, and contributing to this simple linguistic event. First, there is the *sentence* uttered, namely, the declarative "My cat is hungry." Second, there is the *proposition* expressed by the sentence (or rather, by its utterance). Third, there is the *belief* I am trying to convey. Finally, there is the *speech act* that ties everything together, namely, the speech act of asserting that my cat is hungry.

Each entity in this narrative is categorically different from the rest: the utterance of a declarative sentence is a string of sounds (or signs), a proposition is a theoretical abstraction, a belief is a mental state, and an assertion is an act. These four entities can be generalized further. Instead of limiting ourselves to declarative sentences, we can consider sentences of all types. Instead of just beliefs, we can take into account other *propositional attitudes* such as desires, fears, intentions, hopes, and the like. Assertions, of course, are only one type of illocutionary act among many, and the notion of a proposition, usually associated with declarative sentences and beliefs exclusively, can be generalized by introducing the notion of *propositional content*, which is the content of both illocutionary acts and propositional attitudes [23,21]. Thus, within every linguistic event we can isolate the sentence uttered, the propositional content, the propositional attitude, and the speech act.

There are obvious relations among these four entities. If we consider the case of the hungry cat again, it is clear that the choice of that particular sentence was no accident. The content of my belief about the cat is a proposition that is true if and only if my cat is hungry, and the best way to express this proposition is by asserting it, which is done, in turn, by the literal utterance of the sentence "My cat is hungry." This sentence is a perfect candidate for performing, expressing, and conveying the appropriate speech act, proposition, and belief, respectively. It would have been nice if every utterance of a sentence were associated with a unique speech act, propositional content, and propositional attitude, but this, of course, would be too good to be true. Unfortunately, people do not always say what they mean, nor do they always mean what they say. One speech act is performed indirectly through another, different sentences express the same proposition, and different propositions can be expressed by the same sentence. Nevertheless, theoretically interesting relations among the four entities obviously do hold; it follows, therefore, that the study of each is relevant to the study of language and the mind.

Roughly, the problem of reference is to determine how thoughts and sen-

tences can be about objects. In searching for a more precise formulation, we may start by identifying four elements that can be found in every linguistic event in which reference is made to a particular object, and that correspond to the four entities discussed above. If we take up the case of the hungry cat again, there is, first, the *referring expression* "my cat." Then there is the *constituent* of the proposition introduced by the referring expression. The third element is the *mental representation* of the cat. Finally, there is the *speech act of referring* itself that ties all these elements together. These four elements obviously correspond to expressions, propositions, beliefs, and speech acts, respectively, and each highlights a different aspect of the general problem of reference. These four questions are presented in Figure 1.

Referring expressions. As an illustration of the kind of issue associated with the first question, let us consider the philosophical debate about the way proper names designate. Frege [5] believed that all names must have a *sense* that mediates between them and the objects they stand for. Searle [22] rejects Frege's claim, but insists that each name must be "backed by" a set of *identifying descriptions*. Kripke [9], on the other hand, sees names as lacking any intrinsic sense or descriptive content. According to him, names are related to objects through a special sort of causal chain stretching from the moment a name is given, to any particular use of that name. These different accounts attempt to provide a partial answer to the first question.

Propositions. To illustrate the point of the second question, we must distinguish between *general* and *singular* propositions. Let us suppose that the statement "The queen of England is ill" expresses the proposition *there is one and only one thing with the property of being the queen of England, and it is ill.* Such a proposition is called "general" because all references to particular things has been eliminated, and all that we have instead (apart

Expressions: How are referring expressions related to objects?

Propositions: What propositions are expressed by sentences containing referring expressions?

Beliefs: What is the role of mental representation in beliefs about objects?

Speech acts: What is the correct analysis of the speech act of referring?

Figure 1: Main Aspects of the Reference Problem.

from predicates) is a quantifier ("there is"), and a bound variable (represented by "thing" and "it"). A statement such as "7 is prime," on the other hand, is said to express the proposition *7 is prime*, which is called "singular." Note the vast difference between "7" and "The queen of England" as far as their contribution to the logical structure of the proposition is concerned: the referring expression "7" simply introduces the integer 7 into the proposition, but whatever is introduced by the phrase "The queen of England" in the foregoing example is surely to be distinguished from Her Majesty herself. Deciding whether the use of a referring expression results in a singular or general proposition is a central issue as far as the second question is concerned.

Beliefs. A crucial distinction for understanding the third question is that between *de dicto* and *de re* beliefs. A belief *de dicto* is a belief that a certain general proposition *(dictum)* is true. A belief *de re* is a belief about a particular thing *(res)* that it has a certain property.[1] My belief that there are spies is *de dicto* but not *de re* since it does not attribute the property of being a spy to anyone in particular . The belief I have about myself that I am left-handed is certainly *de re* since I am attributing left-handedness to a particular person (namely, myself). My belief that the president of the United States in 1987 is old seems to be both *de dicto* and *de re*. First, it is the belief that the proposition *the president of the United States in 1987 is old* is true, and hence it is *de dicto*. In addition, the content of the belief contains a representation of Ronald Reagan, and thus the belief is about the man himself. Hence, in holding this belief, I am attributing the property of being old to Reagan, and so the belief is *de re* as well.

   The problem expressed by the third question is that of characterizing the relation between *de dicto* and *de re* beliefs. There are two parts to this question: are *de re* beliefs a subclass of *de dicto* ones? Which *de dicto* beliefs are not also *de re*? We can express the same thing in terms of *individuating representations*. Let an individuating representation be a representation whose descriptive content is satisfied by one and only one object. The point of the third question is whether an individuating representation is *necessary* and *sufficient* for a belief containing it to be *de re*. Note that there are four possible ways to characterize the role of such representations in *de re* beliefs: they are either necessary or sufficient, or neither of these, or both.

---

[1]The *de dicto/de re* distinction applies to all propositional attitudes, but for stylistic convenience I am considering only beliefs here.

Referring. Finally, we get to the fourth question about referring as a speech act. The intuitive purpose of referring is to let the hearer know what is being discussed. The problem is how to provide a general account of how this is accomplished.

Each question brings a different perspective to the general problem of reference: the first question is a semantic one (in the sense of relating language to the world), the second has to do with logical form, the third belongs to cognitive psychology, and the fourth deals with the way language is actually *used*. Of course, any answer to one question has some immediate implications for the other three.

The formulation of these four questions makes it easier to characterize a general approach to the problem of reference that I call *the descriptive research program*.

2 The descriptive research program

If a rough statement of the problem of reference is "how can thoughts and sentences be about objects?" the descriptive solution would be, in a word, "descriptive content." But, as the formulation had to be distilled into four distinct questions, the descriptive solution must be spelled out in some detail as well. At the same time, I do not intend to present here a full account of the predominant theories. All I do here is reconstruct, as precisely and as succinctly as I can, the *general principles* that have guided major philosophers in their attempts to provide answers to the four questions corresponding to sentences, propositions, beliefs, and speech acts, respectively. These principles are enumerated in Figure 2.

Expressions (E): Reference is determined by meaning.

Propositions (P): Singular propositions cannot be believed, expressed, or understood.

Beliefs (B): Individuating representation is both necessary and sufficient for a belief to be *de re*.

Speech acts (S): Referring is performed by means of identifying descriptions.

Figure 2: Main Postulates of the Descriptive Program.

What do the theses in Figure 2 mean? Underlying each of them are two simple ideas. The first is that reference is a function of *internal* (i.e., mental) representation. In other words, to refer to an object — either in thought or in speech — is essentially to have or invoke a mental representation of that object. The second idea is that the relation between a sentence or a thought and the object they are about is the relation of *denotation*, which in turn is a function of *descriptive content*. That is, a sentence or a thought about an object contains a referring expression or a concept with a certain descriptive content and, if this content is true of or is satisfied by a unique object, the expression or the concept *denotes* that object. Thus, the crux of the descriptive program is that *reference is entirely a matter of associating a mental state with descriptive content*. What exactly is meant by descriptive content and how it can be associated with mental states depend on the particular theory within the descriptive program that one happens to hold, but the principle remains the same, and is presupposed by all of the theses in Figure 2. Let me illustrate how with respect to each individual thesis.

Referring Expressions. At the level of expressions, descriptive content is associated with *meaning*. The only way a referring expression and an object can be matched, according to the descriptive program, is through the relation of denotation; which object is denoted is determined by the descriptive content associated with the referring expression. Since the requirement for having descriptive content extends to *all* types of referring expression, one of the consequences of Thesis **E** is that proper names and demonstrative are, in one sense or another, disguised descriptions.

Propositions. At the level of propositions, descriptive content is part of the notion of individuating representation, i.e. representations that denote particular objects. Since singular propositions do not contain any individuating representation of such an object, but rather the object itself, the point of Thesis **P** is to assert that some sort of individuating representation is necessary for believing, expressing, and understanding propositions about particular objects. Note that propositions are taken by philosophers to be the content of both beliefs and utterances; it is tacitly assumed that whatever is true of propositions *qua* content of beliefs, is also true of propositions *qua* content of utterances. In particular, Thesis **P** implies not only that an individuating representation is *meant* whenever a speaker expresses a proposition about a particular object, but that it is crucial for determining whether the utterance is true of false.

Beliefs. At the level of beliefs, the notion of an individuating represen-
tation provides the perspective from which (and only from which) objects
can be thought about. Insisting, as Thesis B does, that an individuating
representation is both necessary and sufficient for a belief to be *de re* means
that beliefs about particular objects always depend on individuating repre-
sentations, which in turn depend on descriptive content.

Speech Acts. The goal of the speech act of referring is to identify an
object for the hearer. This is done by using *identifying descriptions*, i.e.,
descriptions that are satisfied uniquely by the object to which the speaker
intends to refer. Of course, it is not necessary to utter a fully identifying
description each time referring is done, but the speaker's ability to provide
such a description is necessary for referring to a particular object. More-
over, within the descriptive approach, the literal utterance of an identifying
description is sufficient for referring to succeed, given normal input and out-
put conditions. Needless to say, a description is identifying by virtue of its
descriptive content.

Among philosophers of language and mind today, it is Searle who most
explicitly identifies himself as working within the descriptive program. But
the program's founding father is undoubtedly Frege, whose theory of sense
and reference underlies all the theses listed in Figure 2. The basic postulate
in Frege's theory is that every expression possesses both a sense and reference
(the reference may be nil). Roughly interpreted, the Fregean sense of a
referring expression is the mode in which the object is presented to the
agent, while the Fregean reference is the object itself. Thus, in Frege's
theory, two names of the same person share the same reference but have
different senses.

The Fregean postulate that every expression has both a sense and a
reference is accompanied by three principles: (1) reference is determined
by sense; (2) the sense of a complex expression is a function of the senses
of its parts; (3) a belief is a relation between an agent and a sense of a
sentence. These three principles entail Theses E, P, and B. Since the notion
of referring as an *act* was introduced by Strawson only in 1950, Thesis S
could not be a part of Frege's theory, nor was Frege much interested in
the communicative function of language. But, as Searle, who incorporated
Strawson's account of referring into his theory of speech acts, specifically
acknowledges, the speech act theory of reference is clearly within the Fregean
tradition [23, 77].

So far I have hardly mentioned Russell, whose theory of descriptions has
long dominated discussions of reference and, to a large extent, still continues

7

to do so. Russell's relation to the descriptive program is an interesting one. Strictly speaking, he is not part of it, since his theory of proper names is inconsistent with all of the descriptive program's theses. Nevertheless, to exclude Russell from the program altogether would, in my opinion, be quite misleading. To understand why, we need to review briefly the essential elements of Russell's theory. This is particularly important because Russell's theory plays an important role in providing the modifications that enable the descriptive program to overcome its difficulties and serve as a foundation for a computational model.

Russell's account of reference has both an epistemological and a semantic segment, closely related to each other. The epistemological segment describes how knowledge of objects is possible. The semantic segment describes how referring expressions are to be interpreted. Crucial to Russell's epistemology is the distinction between *knowledge by acquaintance* and *knowledge by description* [19]. One has knowledge by acquaintance of an object when one has a direct cognitive relation with it, that is, when one is directly aware of the object itself. One has knowledge by description of an object when one knows that there exists one and only one object having a certain property. For example, since I am directly aware of a pain in my left knee, Russell would say that I have knowledge by acquaintance of my pain. On the other hand, I am not acquainted with the 12th president of the United States, but I know that there was one and only one person in the world who was the 12th president of the United States. Thus, I have knowledge by description of him. It is important to note that Russell's theory of knowledge contains two postulates: (1) every proposition that we can understand must be composed of constituents with which we are acquainted; (2) we do not have knowledge by acquaintance of physical objects (in fact, the only objects with which we are acquainted, according to Russell, are *sense data*, and possibly the *self*[2]).

Russell's account of names and descriptions mirrors his epistemological distinction between knowledge by acquaintance and knowledge by descriptions. A proper name designates its object *directly*, but has no meaning apart from the designated object. Thus, when such a name is employed, the object *itself* is a constituent of the proposition, which, in turn, is a singular one. But since we are never acquainted with physical objects (other than our *selves*), they cannot participate in propositions we can understand; consequently, the proper names we use to designate physical objects (including, of course, names given to people) are not really proper names at all but abbreviated descriptions. In fact, as far as names of particulars are concerned,

---

[2] At one time Russell considered self-knowledge to be knowledge by acquaintance. Eventually he changed his mind (ibid., 211, n. 1).

only the deictic words "this" and "I" are true proper names (*logically* proper names, as Russell calls them), corresponding to sense data and the self, respectively — the only particulars with which we can actually be acquainted.

A description, on the other hand, does not designate an object directly. Rather it is capable of *denoting* an object that satisfies its descriptive content. The contribution made by a definite description to a proposition is rather complex, but, as was the case with proper names, the analysis reflects the epistemological distinction discussed earlier.[3] We have seen that to have knowledge by description is to know that one and only one object possesses a certain property. Thus, if a definite description is part of a statement, the contribution of the description to the content of that statement is a logical structure asserting that one and only one object possesses a certain property. Such a logical structure is a propositional function. As a concrete example, consider the statement "The president is old." The contribution of the definite description to the proposition expressed by this statement is taken to be

x is the president and no one else is.

The predicate "is old" adds another propositional function, namely, x *is old*, while the proposition expressed by the statement as a whole simply asserts that, for some value x, the complex propositional function

x is the president and no one else is, and x is old

is true. In standard first-order logic, this is expressed as

$$(\exists x)(\mathbf{Pres}(x) \ \& \ (\forall y)(\mathbf{Pres}(y) \rightarrow x = y) \ \& \ \mathbf{OLD}(x)).$$

Now it should be obvious why it would be quite misleading to exclude Russell's theory from the descriptive program. His account of logically proper names is indeed inconsistent with the descriptive theses, but, according to Russell, the only logically proper names for particulars are "this" and "I."[4] Since all other referring expressions are descriptions (abbreviated or not), what should really matter to us is whether Russell's theory of *descriptions* is part of the descriptive program. Let us consider each of the four theses in turn. For reasons that are essentially technical, Russell claims that descriptions are not meaningful in isolation, only within the context of

---

[3] I am disregarding indefinite descriptions here.

[4] If self-knowledge is indeed knowledge by acquaintance (as I think it is), then, *for each person*, that person's name is also a logically proper name.

a sentence (ibid., 215). Yet, he definitely recognizes that a description denotes by virtue of its descriptive content, which is a function of the meanings of the words that appear in that description. Hence, if reference to physical objects in Russell's theory exists by virtue of denotation, and denotation is a function of meaning, then Thesis E still holds. Moreover, since we are never acquainted with physical objects, we can never believe, express, or understand singular propositions about them. Hence, as far as physical objects are concerned, we have Thesis P and, since all *de re* beliefs about such objects are analyzed in terms of denoting descriptions, which in turn provide individuating representations, thesis B is true as well. As for thesis S, Russell, like Frege, was not much interested in language as a system of communication, but he would have agreed that the purpose of using referring expressions in conversation is to identify an object for the hearer. I don't think he would have been very impressed by this fact, but he would surely not have denied it.

This discussion of Russell's theory concludes my characterization of the descriptive program. I now turn to arguments against it.

3 The descriptive program: objections

The descriptive research program, including Russell's theory of descriptions, has dominated the debate about reference for most of this century. However, in the last two decades or so, a new approach to the problem of reference has been developed. The new approach, to be sure, is not entirely separable from earlier theories. Nevertheless, it marks a radical departure from the descriptive program as a whole and, as a philosophical movement, it advocates an entirely new program. The new theory flatly rejects all four theses mentioned in Figure 2, but offers an alternative: if, according to the old program, *describing* the object is the way to reach it, the new approach takes *pointing* to be the central mechanism of reference. If the core of the old program is descriptive content, which mediates between object and mind, the core of the new one is the notion of a causal chain leading from the object directly to the agent. The most significant difference between the two, therefore, is this: according to the new approach, reference is determined by facts *outside* to the mind, in contrast with the descriptive program, which seeks to explain reference in terms of properties of mental states.[5]

---

[5]The central architects of the new approach to reference are Donnellan [4,3], Kaplan [7,8], Kripke [9], Perry [15,16], and Putnam [17].

3.1 The referential/attributive distinction

The shift from the descriptive program to the new approach was initiated by Donnellan [4], when he introduced a distinction between the *referential* and *attributive* uses of definite descriptions. Since the referential/attributive distinction (*Donnellan's distinction* for short) plays an important role in arguments against the descriptive program, it should be presented first. Here is how Donnellan himself first describes it.

> To illustrate [the referential/attributive] distinction, in the case of a single sentence, consider the sentence, "Smith's murderer is insane." Suppose first that we come upon poor Smith foully murdered. From the brutal manner of the killing and the fact that Smith was the most lovable person in the world, we might exclaim, "Smith's murderer is insane." I will assume, to make it a simpler case, that in a quite ordinary sense we do not know who murdered Smith (though this is not in the end essential to the case). This, I shall say, is an attributive use of the definite description.
>
> The contrast with such a use of the sentence is one of those situation in which we expect and intend our audience to realize whom we have in mind when we speak of Smith's murderer and, most importantly, to know that it is this person about whom we are going to say something.
>
> For example, suppose that Jones has been charged with Smith's murder and has been placed on trial. Imagine that there is a discussion of Jones's odd behavior at his trial. We might sum up our impression of his behavior by saying, "Smith's murderer is insane." If someone asks to whom we are referring, by using this description, the answer here is "Jones." This, I shall say, is a referential use of a definite description. (Ibid., 198)

The intuitive presentation of Donnellan's distinction seems simple enough. The important features to note are as follows. In the referential use, the intended referent can be identified even though no single entity fits the description used or, alternatively, more than one does. The intended referent may also be identified even if something else altogether fits the description. In Donnellan's example, the speaker would have referred successfully to Jones in the trial even if Jones had not been the murderer. In the attributive use, on the other hand, if nothing fits the description, no entity can be said to have been picked out and referred to.

Consequently, if nothing fits the description in the referential use, the speech act may still be successful. If the speech act is an assertion, the speaker may still say something true with respect to his intended referent. If the speech act is a command regarding the intended referent, the command can still be obeyed — and so on for other speech acts as well. This is not the case, however, in the attributive use; if nothing fits the description, the assertion cannot be true of anything, the order cannot be obeyed, etc.

Thus, a description in referential usage is just a tool for identifying the referent; other descriptions that can perform the same task may also be employed. Those that are used attributively, on the other hand, can be deemed "essential." In a sense, they are indeed irreplaceable. The immediate, intuitive reason for this is that, in the referential usage, the speaker can be said to have a particular object in mind. There is a particular entity to be identified. In attributive usage, however, the speaker is referring to whoever or whatever fits the description and there is no particular entity to be identified apart from the description employed.

To the surprise of many, Donnellan's distinction turned out to play an important role in arguing against each of the four major theses of the descriptive program enumerated in Figure 2, page 5. Consequently, Donnellan's distinction will be used here as a methodological tool for the systematic presentation of general objections to the descriptive approach. This does not mean that all objections to the descriptive program are sustained or nullified by Donnellan's distinction. Kripke [9], for example, has attacked significant parts of the descriptive program very forcefully, while at the same time rejecting some of the alleged consequences of Donnellan's distinction [10]. But my purpose here is not to discuss arguments for or against particular elements of the descriptive program. My interest in the descriptive approach lies in its value as a research program, and Donnellan's distinction is an excellent tool for characterizing the general problems that the descriptive program has to resolve.

3.2 Away with meanings

At the level of referring expressions, the central postulate of the descriptive program is thesis E: reference is determined by meaning. The basic idea underlying this thesis is simply this: the only way to establish the required relation between a referring expression and an object is by having a descriptive content that is associated with both of them. This descriptive content is, in a loose sense, the meaning of the expression. But it seems that, in the referential uses of definite descriptions, the descriptive content plays no direct role in establishing the relation between the expression and the object.

12

The expression "Smith's murderer" in its referential use refers (in Donnellan's example) to Jones, but the descriptive content seems irrelevant. The expression will refer to Jones whether he is the murderer or not.

But the challenge to Thesis E transcends the claim that, in some cases, reference is established independently of meaning. Underlying this thesis is a certain view of semantics that can be characterized by three general principles. The first is that meaning determines truth conditions. According to this principle, when one knows the meaning of a sentence, one also knows under what circumstances the sentence is true. A second principle is the Fregean dictum, already mentioned in a different form, which affirms that the meaning of a sentence as a whole is a function of the meaning of its parts. A third principle is that the meanings of individual words are determined by the general conventions of the language, no matter what the speaker wants them to mean. Combining these three principles, we arrive at the notion of semantic independence: with the exception of words that are by nature context-dependent (such as pronouns, demonstratives, and so on), the role of expressions in determining the truth conditions of what is said is independent of the speaker's intentions in using those expressions.

Such a view, however, seems to be inconsistent with Donnellan's distinction. For it seems that two literal utterances of the same sentence may differ in their truth conditions, depending on whether the definite description is used referentially or attributively. Compare, in Donnellan's example, the two utterances of "Smith's murderer is insane." In the first case, in which the description is used attributively, the speaker means *Smith's murderer, whoever he is*, and the statement will be true if and only if the one and only murderer of Smith is insane. In the second case, when the description is used referentially, the speaker means *Jones*, and whatever is said is true if and only if *Jones* is insane. Thus, in a limited sense, it seems that Humpty Dumpty was right after all: you can utter the words "Smith's murderer," and it is entirely up to you whether you meant *whoever murdered Smith*, or *Jones*. In other words, it seems that one and the same expression — "Smith's murderer" — plays a different role in forming the truth conditions of whatever is said, depending on whether the speaker intends to refer to a specific individual or just to the murderer of Smith, whoever he might be.[6]

---

[6]One way to maintain the independence of semantics in the face of Donnellan's distinction is to claim that definite descriptions are simply semantically ambiguous. But this is really a desperate move, and there is a much better way to solve the problem [11].

3.3 In praise of singular propositions

The central postulate of the descriptive program at the level of propositions is thesis P: one cannot believe, express, or understand singular propositions. In other words, what a referring expression brings into the proposition is never the object itself — which is just a philosopher's fancy way of saying that an individuating representation of the object is essential for reference to take place, at least as far as physical objects are concerned. This view, however, is challenged by the referential use of definite descriptions. I shall illustrate the argument with respect to Russell's theory of descriptions. The same argument can be applied to other descriptive theories as well.

A standard objection to Russell's theory of descriptions is that, in many uses of definite descriptions, application of the theory yields the wrong results. When a speaker says "The computer is down," it is clear that he does *not* mean that there is one and only one computer in the world and that it is down. The standard reply is that the referring expression uttered is an elliptical form of a uniquely denoting description. What the speaker really means is, say, that the VAX-750 in Room EK247 at SRI International is not operational at the moment. There is no problem in applying the Russellian analysis to such a complete form of the description.

However, when we look at some referential uses of definite descriptions, it seems that the standard reply won't do as it stands. Consider again an utterance of "The computer is down," in which the definite description "the computer" is used referentially. It is clear that the computer can be described more fully in many ways that are not equivalent — but which of these alternatives is the one actually meant by the speaker? In other words, which of them should be taken as part of the proposition that the speaker expressed?

First, it should be noted that in many cases the hearer may complete the description in a different way from the one originally intended by the speaker, and still it may not be clear that the hearer misunderstood what is said. For example, suppose a speaker uttered the above sentence, meaning "The computer in room EK247 is down," and that the hearer understands what is being said to mean "The computer used by the natural-language group at sri is down." Has the hearer misunderstood the speaker? It is by no means clear that he has: since the machine in EK247 *is* the one used by the Natural-Language Group, the hearer has identified the right computer and he now knows that it is down. Shouldn't this suffice?

Second, in many cases the speaker himself may not conceptualize a complete description at all; thus he would not know which complete description should be considered to be the one he really *meant*. A speaker may utter

14

"The computer is down" without intending that the hearer pinpoint any *particular* individuating property of the computer.

Third, in the radical case of the referential use in which the description misses the mark entirely (e.g., "Smith's murderer" used in reference to Jones, who is innocent), "completing" the description is simply impossible. Moreover, when asked after the utterance which description he *really* meant, the speaker may be at a loss to respond. He would probably be inclined to say that he meant "Smith's murderer," but now, after realizing that Jones is not the murderer, he is no longer sure of what he did mean. This is not to say that the speaker is not readily capable, if challenged, of replacing "Smith's murderer" with a new, more accurate description, but it would not be what he *meant* before. In any case, as these examples illustrate, it seems that the speaker would not care much *which* description he actually meant, or which was the accurate one, as long as the hearer was able to identify the right person or the right computer.

A natural conclusion to be drawn at this stage of the argument is that in referential usage, the only thing relevant to the proposition that the speaker is trying to express is the object per se, independently of any of its descriptions. In other words, in the referential case, it seems that the only thing relevant to the truth conditions of what the speaker is trying to express is the object itself; if this is indeed so, why not just take the object itself as a constituent of the proposition? Thus we arrive at what Kaplan has called the semantics of *direct reference*, according to which utterances containing referential uses of definite descriptions (as well as proper names and demonstratives) express singular propositions, with all such referring expressions constituting rigid designators, i.e., expressions that designate the same object in all the possible worlds in which it would exist. Such a semantics, of course, flies in the face of thesis P: singular propositions, according to this account, are frequently believed, expressed, and understood.

3.4 The status of de re beliefs

At the level of belief, the central postulate of the descriptive program is thesis B: individuating representations are both necessary and sufficient for a belief to be *de re*. The attempts to reject thesis B and Donnellan's distinction are related first by a simple extension of thesis P: if the semantics of direct reference is accepted and singular propositions can be the content of beliefs, then such beliefs *de re* contain no individuating representations. But the connection between Donnellan's distinction and propositional attitudes goes beyond this. In fact, some philosophers would say that Donnellan's distinction and the *de re/de dicto* dichotomy are very closely related to each

15

other [6]. The idea is simply this: when a speaker asserts that so-and-so is such-and such, the belief he expresses, if he uses the definite description referentially, is *de re*. If, on the other hand, the definite description is used attributively, the belief expressed is *de dicto*, but not *de re*. Conversely, when a speaker wishes to express a *de re* belief, the referring expression he utters will be used referentially. If the belief is *de dicto* (and is concerned with a particular object), the referring expression uttered will be used attributively.[7]

I have discussed the rational for identifying Donnella's distinction with the *de dicto/de re* dichotomy elsewhere [12]. For now, I'd like to point out how such an identification undermines thesis B: if indeed [the expression of] *de re* beliefs and referential usage always go hand in hand, and if, in referential usage individuating representation is not relevant for what the speaker *means*, why should such representation be relevant to what he *believes?* In other words, if individuating representation is neither necessary nor sufficient for referential uses to be successful, it should be neither necessary nor sufficient for *de re* beliefs.[8]

3.5   Identification reconsidered

At the level of speech acts, the main thesis of the descriptive program is Thesis S: referring is performed by means of identifying descriptions. But when we use a description attributively, even though we are certainly trying to let the hearer know what we are talking about, we don't seem to *identify* anything for him. We don't seem to do that because, at least in a strong intuitive sense, we are not able to identify the object even for ourselves: when a speaker says "Smith's murderer, *whoever he is,* is insane," he is obviously talking about Smith's murderer. But, Donnellan would argue, there is no particular person whom he is able to identify as Smith's murderer, either for himself or for the hearer. Thus, on this account, thesis S is at best misleading, since it ignores an important, nonidentifying function of

---

[7]Kripke [10] thinks that the identification of Donnellan's distinction with the *de dicto/de re* dichotomy is entirely without merit, but in my opinion, his argument misses the point. Kripke argues that in *de re* reports of beliefs — i.e., in constructions such as "concerning Smith's murderer, Ralph believes that he is insane" — descriptions can be used *either* referentially or attributively. The same can be said with regard to *de dicto* reports of beliefs such as "Ralph believes that Smith's murderer is insane." This is true enough, but it does not affect the contention that, in *expressing* a *de re* belief (rather than reporting someone else's), a descriptions is always used referentially and, whenever a description is used attributively, the belief *expressed* is *de dicto*.

[8]Cf. Burge [2]. Burge's argument that *de re* beliefs are not a subclass of the *de dicto* type is similar to Donnellan's argument that, at least in some referential uses, no particular description can be picked out as the one *meant* by the speaker.

referring expressions. On the other hand, Donnellan argues [3], even though in the referential use we do identify an object for the hearer, we do not necessarily do it by means of identifying descriptions — at least not in the sense intended by Strawson and Searle. Thus, Donnellan's argument forces a problem on the speech act theorist: if (as Strawson and Searle maintain) the central concept in the analysis of referring is identification by means of identifying descriptions, then in the referential use we seem to have identification without identifying descriptions, and in the attributive use we seem to have identifying descriptions (in the technical sense of speech acts theory) without identification. On the other hand, if we want a unified account of referring that includes both referential and the attributive usage, we cannot take our intuitive understanding of the notion of identification for granted. We must specify precisely in what sense identification takes place in both types of usage.

Donnellan's rejection of the speech act thesis concludes this discussion of the descriptive program. What I have tried to do is to show how Donnellan's distinction plays a role in repudiation of the descriptive program as a whole. The question is whether such a total rejection is justified.

4 The descriptive program: motivation

So far I have characterized in some detail the descriptive program, as well as some of its problems, and have introduced the program's leading rival. At this point, a computational linguist or an artificial intelligence researcher might raise two important questions. First, given that what we are after is a computational model of referring, does it matter really which program we choose? Second, assuming that it does matter, why — given the objections that the descriptive program must respond to, and given the availability of an alternative — should we stay with the descriptive program?

In this section I answer these questions in a way that can be summarized as follows. For a computational model, it matters a great deal which approach is chosen because referring requires reasoning about the beliefs, desires, and intentions of other agents; this in turn requires an adequate account of the *content* of belief. It is precisely with respect to the content of belief (and other propositional attitudes) that the descriptive program and the new, or *causal*, approach differ radically. The descriptive program, no doubt, must undergo a fair amount of revision to eliminate its difficulties, but its main thrust stands unchallenged, in my opinion. Furthermore, upon closer scrutiny, it becomes clear that, as far as an accounting for the content

17

of beliefs is concerned, the causal approach has so far had really very little to offer.

In order to make sense out of the behavior of others, we attribute to them beliefs, desires, intentions, and other propositional attitudes. For example, if you ask me why my neighbor gets into his car every morning, the most reasonable answer I can give you is that he *wants* to get to work and that he *believes* the car can take him there. Moreover, very frequently we attempt to achieve our goals by trying to change the beliefs and desires of others. This would be quite impossible if we did not have some idea to begin with as to the kind of beliefs and desires they possess. Thus, an important part of our model of the world around us is a representation of what we take to be the beliefs, desires, and intentions of others.

The role of propositional attitudes and their representations is particularly obvious when we try to model our linguistic behavior. The central idea underlying computational speech-act theory is that the performance of a speech act is the result of the speaker's planning to affect the hearer's conception of the speaker's intentions, beliefs, and desires. Thus, a computational model of any speech act, including that of referring, must be able to show how reasoning about someone else's propositional attitudes is possible; to do that, such a model must be able to represent that someone else's beliefs. But the notion of a belief makes no sense without a fairly precise notion of what the *content* of a belief is. It is the *content* of a particular belief that sets it apart from all others, and it makes no sense to represent that belief without somehow representing its content. Hence, a model of referring must begin with some notion of the contents of beliefs about particular objects, i.e., *de re* beliefs. Note, incidentally, that the term 'content' is used here to mean *whatever* individuates beliefs. That is, whatever the content of a belief, two beliefs are the same if and only if they have the same content.

Now, if the notion regarding the content of a *de re* belief is important for a model of referring, we cannot remain neutral with respect to choosing between the two research programs. According to the descriptive program, as we have seen, an essential component in the content of a *de re* belief is a representational entity that mediates between the belief and the object it is about. According to the causal approach, on the other hand, the relation between the belief and the object it is about is *direct*, and — at least according to the causal approach in its original form — the content of a *de re* belief is a *singular proposition* consisting of the object itself plus a propositional function. So, even at the level of formal representations of beliefs, the choice between the two programs is a significant decision. But much more is obviously at stake. The contrast between the two accounts of content is

18

a reflection of the far more significant difference between the two programs. In the descriptive program, reference is determined by mental states: when the agent *apprehends* (Frege's term) the representational entity, reference is assured. In the causal approach, by contrast, reference is determined (at least in part) by the objective situation, quite independently of what the agent himself may think or believe.

The difference between the two positions has far-reaching implications not only for a general theory of mind, but for particular computational models as well. Imagine, for example, the initial steps in designing an intelligent system that is supposed to recognize, manipulate, and discuss objects in its environment. Such a system would have to have knowledge of objects, as well as the ability to represent and reason about the various beliefs, desires and expectations of other intelligent agents with which it is supposed to interact. If the system designer were to take the descriptive approach, he would assume that (1) internal representations of objects are essential for such a system; (2) a successful act of referring involves invocation of an internal representation in the hearer; and (3) these internal representations must be rich enough in descriptive content to provide a mechanism for reference to succeed. The main task of the designer, therefore, would be to look for appropriate data structures that play the same role as internal representations in various computational activities. Perhaps the most obvious choice at this stage of AI development is data structures that encode formulas of first-order logic, although this is by no means the only possible way.

On the other hand, if the designer takes a causal approach to reference, he would seek computational techniques that emphasize objective correlations between a world state in which, say, the Empire State Building exists, and a machine state that can be said to contain the information that the Empire State Building exists. A practical way to implement such an account of *de re* beliefs may be borrowed, perhaps, from the situated-automata approach to knowledge representation. According to that approach, a machine state S is said to contain the [true] information that P if and only if P is a world condition guaranteed to hold whenever the initial state of the machine is coupled to a world state W and an input string that is fed into the machine while the latter is in its initial state causes the machine to be in state S [18]. It may be that the situated-automata approach to knowledge representation can be extended to representations of *belief* — in particular, *de re* belief.[9]

---

[9]There is no doubt that the characterization of a state in which an agent *knows* something must take into account not only the agent's internal state but also how the world is. After all, for an agent to know that P, P must be true. Thus, nothing prevents the descriptive theorist from adopting the situated-automata approach for knowledge repre-

The difference between the two competing research programs is therefore computationally significant, and my contention is that a rational choice ought to favor the descriptive program. Obviously, I do not presume to "refute" the causal approach — I do not believe that research programs can be "refuted" — but I think that the causal approach suffers from a major weakness that, in my opinion, far outweighs the difficulties faced by the descriptive program. This weakness is related to the notion of the content of a belief. But before stating my case against the causal approach, I need to say at the outset why I do not regard the difficulties of the descriptive program to be as devastating as is commonly thought.

The description program can be interpreted as a general framework for two distinct (albeit related) projects. The first project is the attempt to provide a theory of meaning for singular terms. The second project is the attempt to construct a theory of mind. The arguments against the descriptive program are most powerful when the focus is on meaning: for example, it is widely accepted that Kripke's criticism of the descriptive theory of proper names [9] is correct. But there is a tendency to extrapolate from problems of [linguistic] meaning to alleged inadequacies of the descriptive program as characterizing the general principles that underlie a theory of mind. Now, in trying to provide a model of referring, we are much more interested, for obvious reasons, in understanding the mind than we are in a theory of meaning. The latter per se cannot help us with questions of rationality, planning, actions, intentions, knowledge representation, and reasoning, — topics of relevance to a computational account of communication (as well as to other cognitive models). Hence, as far as a model of referring is concerned, the main objections to the descriptive program are valid only if problems arising from a descriptive theory of meaning are indeed carried over to a theory of mind. I think that this is not the case.

Let us turn now to the problems that must be overcome by the causal approach if it is to be a real alternative to the descriptive program. I shall start by stating a premise that I take to be trivially true. I call it the *trivial principle*:

> **Trivial Principle:** It is impossible both to hold and not to hold the same belief.

---

sentation. The dispute between the descriptive and the causal theorists is whether facts outside the mind are relevant to the characterization of *belief*. Incidentally, the descriptive theorist does not deny, of course, that causal explanations play an important role in understanding *why* someone holds a *de re* belief. For example, it is obvious that, in general, evidence is causally grounded. The descriptive theorist would insist, however, that what the belief is about is determined solely by its descriptive content.

It is important that the triviality of this claim be appreciated. This is not a characterization of our rationality. I am not actually claiming that it is impossible to hold both a belief and its negation. Although this is not particularly recommended, one can certainly hold the belief that P and, at the same time, hold the belief that not P. A belief and its negation are two distinct beliefs and, if one chooses to hold both, one is free to do so. But an agent cannot both hold and not hold the *same* belief any more than a geometric shape can be both a square and a circle.

Now, what is the content of a *de re* belief? Let us take a concrete example. Suppose Ralph believes of Wiley that he is a spy. Or — to attribute a *de re* belief to Ralph in a more dramatic way — let Ralph himself point to Wiley and say: "I believe that this man is a spy!" Thus, there is no doubt that Ralph has a *de re* belief about Wiley, and its content, according to the causal theory (at least in its original form), is the singular proposition:

1 SPY(*Wiley*).

But the singular proposition cannot be the *complete* specification of the content of Ralph's belief because of the *trivial principle*. Suppose that, on a certain occasion (say, on the beach), Ralph points to Wiley and says "I believe this man is a spy." Suppose that at another time (say, in a supermarket), Ralph points to Wiley and says "I do not believe this man is a spy." Let us assume that Ralph is sincere on both occasions, and that his only problem is his failure to recognize Wiley in the supermarket as the man he saw earlier on the beach. If the *complete* content of Ralph's belief is a singular proposition, Ralph's first utterance shows that he holds a belief whose content is expressed by (91), while his second utterance shows that he does *not* hold the very same belief. But this is impossible according to the *trivial principle*.[10]

---

[10]Donnellan once suggested to me that (1) is indeed the complete content of Ralph's belief. When Ralph thinks to himself in the supermarket "I do not believe this man is a spy," Ralph is simply wrong. As a matter of fact, Donnellan suggested, Ralph *does* believe the man in the supermarket is a spy, but Ralph himself is not aware of this fact about himself (I do not know whether Donnellan still holds this view). Such a position, of course, amounts to abandoning the presumption of *positive introspection*, i.e., the presumption that, when an agent believes P, he also believes that he believes P. Note that Ralph's case can also be described in terms of *knowledge:* Ralph may know that the man he sees on the beach is a spy, while not knowing whether the man in the supermarket is one. Hence, if Donnellan's suggestion is to be maintained, Ralph may know something without knowing that he knows. Giving up positive introspection with respect to both knowledge and belief is not only extremely unintuitive, but, as Moore [14, 16] notes, has dire consequences for formal accounts of planning.

It is important not to confuse the problem this example poses for the causal theorist with that of the cognitive significance of *utterances*. Two utterances differ in their cognitive significance if a rational agent can accept one as true while rejecting the other. In the foregoing example, it is clear that an utterance on the beach of "This man is a spy" (coupled with a pointing gesture toward Wiley) would be accepted as true by Ralph, while the very same utterance and gesture in the supermarket would be rejected. Since, according to the causal theorist, the two utterances express the same proposition, it is not apparent what should account for the difference in cognitive significance. This is indeed a thorny problem for a causal theorist [24], but the problem of the cognitive significance of *utterances* is not our concern. In fact, my point has nothing to do with natural language at all. If we could somehow grasp the content of Ralph's beliefs directly, we could have dispensed with Ralph's utterances altogether. My argument is simply this: (1) *believing* is a relation between an agent and a content; (2) Ralph cannot hold and yet not hold a belief any more than a beer bottle can both be in the cooler and not in it; (3) if the content of the belief attributing "spyhood" to Wiley is a singular proposition, Ralph would then have to both hold and not hold the same belief; hence, singular propositions cannot by themselves be contents of beliefs. Nothing here hinges on the cognitive significance of utterances.

If the singular proposition is not the complete content of Ralph's belief, some element of content is missing. Let Ralph's mode of presentation (of Wiley) be *by definition* that missing element. We do not know at this point what a mode of presentation is. What we do know is that a mode of presentation, in the sense described above, is necessary regardless of what one's theory of the content of *de re* beliefs is. The *trivial principle* simply requires it.

Several causal theorists have recognized that singular propositions do not suffice to individuate beliefs, and have suggested various remedies [1,7,15]. But these remedies do not seem to me adequate, because they do not take seriously enough the need for modes of presentation. Whatever we take such modes to be, they must satisfy the following condition, which I call the *basic constraint*.

> Basic Constraint: For every mode of presentation M_1 and M_2,
> if $M_1 = M_2$, then, if Ralph believes Wiley to be a spy under M_1,
> he also believes Wiley to be a spy under M_2.

I take the *basic constraint* to be as self-evident as the *trivial principle*. It is nothing more than an instantiation of Leibnitz's law: if two things are identical, whatever is true of one is true of the other. However, if one feels

uncomfortable applying Leibnitz's Law in a context of beliefs, all we need do for reassurance is to transform the *basic constraint* into its equivalent:

> **Basic Constraint (second version):** For every mode of presentation M_1 and M_2, if Ralph believes Wiley to be a spy under M_1 and does *not* believe Wiley to be a spy under M_2, then $M_1 \neq M_2$.

In its second version, the *basic constraint* is simply a restatement of the initial motivation for introducing modes of presentation into the content of *de re* beliefs. If the constraint is not satisfied, it is easy to construct circumstances in which the *trivial principle* is violated.

Now, from the *basic constraint* on any theory of presentation modes it is possible to derive another principle, which I shall call the *individuation principle*. It is this:

> **Individuation Principle:** If M is a mode of presentation under which Ralph believes Wiley to be a spy, then, in each possible world that is compatible with Ralph's beliefs, one and only one object is presented to Ralph under M.

What the *individuation principle* means is that modes of presentation — whatever they are — must carry out an individuation function within one's network of beliefs. In other words, if M, the mode of presentation under which Ralph believes Wiley to be a spy, is a concept, it is an individual concept instantiated by a unique object in each possible world compatible with Ralph's beliefs. If M is a description, it is a definite description denoting a unique object in each world compatible with Ralph's beliefs. If M is a causal chain, it "determines" a unique object in each world compatible with Ralph's belief, and so on with regard to anything a theory of presentation modes might come up with.

Incidentally, the argument for the *individuation principle* does not depend upon the concept of "possible worlds." States of affairs, alternative circumstances, situations, etc., would do equally well. As a matter of fact, as far as I am concerned, we can dispense with such terminology altogether and rephrase the *individuation principle* as follows: if M is a mode of presentation under which Ralph believes Wiley to be a spy, then Ralph believes that one and only one object is presented to him under M. This version, however, can be interpreted as implying that Ralph is aware of presentation modes, a claim that the causal theorist may deny (for example, by arguing that modes of presentation are really causal chains). In other words, phrasing the *individuation principle* in terms of possible worlds leaves open

23

the question of whether the content of Ralph's belief is "in his head" or not. Another reason for using the apparatus of possible worlds is that the argument for the *individuation principle* is easier to state in this manner.

The argument for the *individuation principle* is, in essence, identical to the original motivation for modes of presentation. Let us suppose that the *individuation principle* is false and, to make the argument more concrete, let us assume that, in our theory of content, modes of presentation are nonindividuating *concepts*. Suppose, for example, that the mode of presentation under which Ralph believes Wiley to be a spy is *man-on-the-beach*. Since Ralph takes Wiley to be a man on the beach, this concept is instantiated by at least one individual in every world compatible with Ralph's beliefs. As this is a nonindividuating concept, however, it is certainly possible for Ralph to believe that, at another time, he sees a different man on the beach; Ralph has no opinion as to whether or not he is a spy. Thus, on one occasion Ralph thinks to himself "I believe this man on the beach is a spy," whereas at another time he thinks "I do not believe this man on the beach is a spy." But suppose that on both occasions the man is really is Wiley (although Ralph does not realize that). If the complete content of Ralph's belief is

2 SPY(*Wiley*) [under mode of presentation: *'man-on-the-beach'*],

then we again find Ralph both holding and not holding the same belief. But this would be impossible, since it, too violates the *trivial principle*.

I have borrowed this argument from Schiffer [20], who uses it to show that nonindividuating concepts cannot be modes of presentation. But the same argument can be easily generalized to show that the *individuation principle* must be right, no matter what one takes modes of presentation to be. The schema of the argument is as follows. Let us assume that the *individuation principle* is false. There are then possible worlds compatible with Ralph's beliefs in which two distinct individuals are presented to Ralph under M. Nothing prevents Ralph from believing that one of them is a spy while *not* believing that the other is. But, since Ralph may fail to recognize that the man is Wiley in both cases, we have a situation in which Ralph believes Wiley to be a spy under M while *not* believing Wiley to be a spy under M. This contradicts the *basic constraint*, as well as, of course, the *trivial principle*. Hence, our hypothesis that the *individuation principle* is false must be incorrect.

The *individuation principle* holds no surprises for the descriptive theorist. The only thing missing there, he would say, is the realization that modes of presentation do not merely determine a unique object in worlds that are compatible with Ralph's beliefs. They determine the *referent*, i.e.,

what the belief is about in the *actual* world, whether this world is compatible with Ralph's beliefs or not. As this is, indeed, what modes of presentation are supposed to do in the descriptive program, we can begin to see why the latter is so attractive. Since the *individuation principle* holds in any case, individuating representations seem made to order for a theory of reference. Thus, in the descriptive program we have a stronger version of the *individuation principle*, which I call the *fregean principle*:

> Fregean Principle: Modes of presentation determine what a belief is about.

Now, according to the descriptive theory, whatever anchors a belief to the actual world is conceptual, or at least mental in nature. But the causal theorist insists that the object a *de re* belief is about is determined by a referential, presumably causal *chain*, beginning with the object outside Ralph's mind and — as if by a domino effect — progressing through Ralph's various mental states (perceiving Wiley, remembering him, etc.), and terminating in Ralph's belief state concerning Wiley and the property of being a spy. Given such a position, there are but two ways in which the causal theorist might incorporate the *individuation principle* into his account (which, in the last analysis, he must do). He can either accept the *fregean principle* or reject it.

If the causal theorist accepts the *fregean principle*, he must reinterpret causal chains as modes of presentation, which of course means that causal chains are part of the content of *de re* beliefs. I cannot imagine what this would mean in terms of, say, reasoning about the content of another agent's belief. But even if we can make sense out of this position, two of its consequences are (1) that a change in a causal chain should alter the corresponding belief and (2) that, when two causal chains are identical, so are the corresponding beliefs. This is extremely unintuitive. First, if Ralph had seen Wiley through a periscope, the causal chain from Wiley to Ralph would have been different, but it is hard to see why this should affect Ralph's belief. Second, suppose that the two causal chains that are responsible for two of Ralph's beliefs concerning Wiley are identical replicas of each other — right down to the subatomic level. Unless we adopt a very crude version of mechanical behaviorism, it still does not follow that the beliefs must be the same. To verify this, just imagine that, in spite of the identity of his two visual experiences, Ralph fails to recognize the same person in both. Finally, consider cases in which the causal chain is totally absent (for example, in cases of illusion). Let the entire scene on the beach be a massive hallucination. Ralph indeed no longer has a *de re* belief about Wiley (or anybody else, for that matter), and many causal theorists would argue that

25

the content of Ralph's belief in this case is an "incomplete" proposition (one with a "gap"). But Ralph's belief seems to be the same whether he is under an illusion or not.

If, on the other hand, causal chains are not modes of presentation, the only alternative left for the causal theorist is to reject the *fregean principle*. Modes of presentation are needed no matter what one's theory is and, in each possible world compatible with Ralph's beliefs, the mode of presentation determines a unique individual. But, the causal theorist would claim, the mode of presentation need not pick out a unique individual in the *actual* world. Even if it does pick out a unique individual, it does not have to pick out the right one. Let us suppose that modes of presentation are individual concepts. Given that premise, there is an individual concept under which Ralph believes Wiley to be a spy, and this concept is instantiated by a unique individual in each possible world compatible with Ralph's beliefs. But this concept may be one that *in fact* does not fit Wiley at all. Nevertheless, Ralph's belief may still be about Wiley by virtue of the causal connection between the person Wiley and Ralph's mental state.

This position is derived from referential usage of definite descriptions. When I say "Smith's murderer is insane," referring to Jones at the dock, my utterance is about Jones whether he is guilty or not. Similarly, the causal theories may argue, my *belief* that Smith's murderer is insane may still be about Jones no matter who really murdered Smith. In such a case, the mode of presentation (roughly, the property of being the murderer of Smith) indeed determines a unique individual in all possible worlds that are compatible with my beliefs, in accordance with the *individuation principle*. However, whom in the *actual* world my belief is about is determined not by the mode of presentation, but by something else entirely.

But the inference from utterances to beliefs is very misleading in this case. Even though the description "Smith's murderer" does not apply, I can still use it precisely because I can identify Jones independently of his being the murderer of Smith. In other words, I have more than enough modes of presentation that pick out Jones in the *actual* world. To make the current suggestion plausible, it must be shown that, despite my inability to identify Jones independently of the concept *Smith's murderer*, and despite the fact that Jones is *not* the murderer, my belief can still be said to be about Jones. What we are asked to imagine, indeed, is the following: I believe that Smith's murderer is insane; I cannot associate any other mode of presentation with the person I take to be Smith's murderer; Jones is *not* the murderer; but my belief is still about Jones. Donnellan [3] has attempted to show that this is indeed possible, but as I have argued elsewhere [13], his argument does not work.

The *fregean principle* certainly does not follow from the *individuation principle*, but surely a mere assertion to that effect is insufficient. The least the causal theorist should do is sketch out a plausible theory in which modes of presentation satisfy both the *basic constraint* and the *individuation principle*, and in which reference is nevertheless determined by something entirely outside the content of belief. I do not think any such theory is forthcoming, especially since the notion of a causal chain as determining reference is so hopelessly vague at this point.

As I have said earlier, this is not an attempt to "refute" the causal approach. My only point is that the *individuation principle* makes the descriptive research program more promising. Modes of presentation are needed no matter what one's theory of the content of beliefs is, and modes of presentations, whatever they may be, must individuate objects for agents. Given these facts, the ease with which the notion of individuating representation can accommodate them, and the difficulties of incorporating a theory of modes of presentation within the causal approach, it seems to me that the descriptive approach is the logical choice.

The proof of the pudding, however, is in the eating. The most convincing argument for the descriptive program is made by showing that its adoption as a philosophical foundation represents a productive step in developing a computational model.

References

[1] Jon Barwise and John Perry. *Situations and Attitudes*. The Massachusetts Institute of Technology Press, Cambridge, Massachusetts, 1983.

[2] Taylor Burge. Belief *De-re*. *The Journal of Philosophy*, LXXIV:338–362, 1977.

[3] Keith S. Donnellan. Proper names and identifying descriptions. *Synthese*, 21:334–358, 1970.

[4] Keith S. Donnellan. Reference and definite descriptions. In J.F. Rosenberg and C. Travis, editors, *Reading in the Philosophy of Language*, pages 195–211, Prentice Hall, Englewood Cliffs, New Jersey, 1971.

[5] Gottlob Frege. Sense and reference. In Donald Davidson and Gilbert Harman, editors, *The Logic of Grammar*, Dickenson, Encino, CA, 1975.

[6] Jakko Hintikka. Individuals, possible worlds, and epistemic logic. *Nous*, 1:33–62, 1967.

[7] David Kaplan. Demonstratives. 1977. University of California, Los Angeles. Photocopy.

[8] David Kaplan. Dthat. In Peter Cole, editor, *Syntax and Semantics, Volume 9*, Academic Press, New York, New York, 1978.

[9] Saul Kripke. *Naming and Necessity*. Harvard University Press, 1980.

[10] Saul Kripke. Speaker reference and semantic reference. In Jr. Peter A. French, Theodore E. Uehling and Howard K. Wettstein, editors, *Contemporary Perspectives in the Philosophy of Language*, pages 6–27, University of Minnesota Press, Minneapolis, Minnesota, 1977.

[11] Amichai Kronfeld. Donnellan's distinction and a computational model of reference. In *Proceedings of the 24th Annual Meeting*, pages 186–191, Association for Computational Linguistics, 1986.

[12] Amichai Kronfeld. *Donnellan's distinction as an adequacy test for a referring model*. Technical Note, Artificial Intelligence Center, SRI International, 1988.

[13] Amichai Kronfeld. *The Referential Attributive Distinction and the Conceptual-Descriptive Theory of Reference*. PhD thesis, University of California, Berkeley, 1981.

[14] Robert C. Moore. *Reasoning about Knowledge and Action*. Technical Note 191, SRI International Artificial Intelligence Center, 1980.

[15] John Perry. Frege on demonstratives. *Philosophical Review*, 86:474–497, 1977.

[16] John Perry. The problem of the essential indexical. *NOUS*, 13, 1979.

[17] Hilary Putnam. The meaning of 'meaning'. In Keith Gunderson, editor, *Language, Mind, and Knowledge*, University of Minnesota Press, Minneapolis, 1975.

[18] Stanley J. Rosenschein. Formal theories of knowledge in ai and robotics. *New Generation Computing*, 3(4):345–357, 1985.

[19] Bertrand Russell. Knowledge by acquaintance and knowledge by description. In *Misticism and Logic*, Penguin, 1953.

[20] Stephen Schiffer. The basis of reference. *Erkenntnis*, 13:171–206, 1978.

[21] John R. Searle. *Intentionality: An Essay in the Philosophy of Mind*. Cambridge University Press, Cambridge, England, 1983.

[22] John R. Searle. Proper names. *Mind*, LXVII(266):166–173, 1958.

28

[23] John R. Searle. *Speech Acts: An Essay in the Philosophy of Language.* Cambridge University Press, Cambridge, England, 1969.

[24] Howard Wettstein. Has semantics rested on a mistake. *The Journal of Philosophy*, LXXXIII(4):185–209, 1986.