

SRI International

DONNELLAN'S DISTINCTION AS AN ADEQUACY TEST FOR A REFERRING MODEL

Technical Note 436

April 20, 1988

By: Amichai Kronfeld

Artificial Intelligence Center
Computer and Information Sciences Division

**APPROVED FOR PUBLIC RELEASE:
DISTRIBUTION UNLIMITED**

The research reported in this paper was supported, in part, by the National Science Foundation Grant DCR-8407238. The views and conclusions expressed in this paper are those of the author and should not be interpreted as a representative of the views of the National Science Foundation or the United States Government.



333 Ravenswood Ave. • Menlo Park, CA 94025
(415) 326-6200 • TWX: 910-373-2046 • Telex: 334-486

Donnellan's Distinction as an Adequacy Test for a Referring Model*

Amichai Kronfeld

April 20, 1988

Contents

1	Introduction	2
2	Donnellan's distinction(s)	2
3	Having a particular object in mind	4
4	A three-tiered model of referring	11
4.1	Individuating sets	12
4.2	Referring intentions	14
4.3	Choice of referring expressions	15
4.4	Donnellan's distinction: final chord	16

*The research reported in this paper was supported, in part, by the National Science Foundation Grant DCR-8407238. The views and conclusions expressed in this paper are those of the author and should not be interpreted as representative of the views of the National Science Foundation or the United States Government. I am grateful to Doug Appelt for his comments.

1 Introduction

Let us recall the two crucial features of Donnellan's distinction: whereas, in the attributive use, the description must be satisfied for reference to succeed, in the referential use this is not so. In the latter case, moreover, the speaker has a particular object in mind, while in the attributive he does not.

As I have shown elsewhere [14], Donnellan's distinction is used extensively in arguments against the descriptive approach to reference. But opinions differ sharply as to its true significance. Chastain, [4], for example, thinks that the paper in which Donnellan's distinction is introduced is "the first significant advance beyond Russell" (196). Castañeda [3], on the other hand, takes such enthusiasm to be "too much ado about practically nothing" (186). Even those who accept Donnellan's distinction simply as characterizing interesting linguistic behavior disagree strongly among themselves as to how it should be interpreted or represented in a computational model. For Cohen [5], the speaker's intention that the hearer identify the referent constitutes a crucial difference between the referential and the attributive. Barwise and Perry [1], on the other hand, take the referential use to be an instance of a definite description whose interpretation is *value-loaded*. That is, its denotation is a particular object in a particular situation that is "accessible" to participants in the conversation. However, as is pointed out by Grosz et al. [7], Barwise and Perry's analysis ignores an essential aspect of the referential use, namely, the speaker's ability to refer to an object, whether or not that object satisfies the description at all.¹

In this paper, I provide an analysis of the cognitive structures that underlie referential and attributive uses of definite descriptions. This endeavor should be worthwhile in and of itself, but it is also intended here as a means toward another end. As we shall see, this analysis of Donnellan's distinction represents an initial step toward a construction of a referring model.

2 Donnellan's distinction(s)

Let us now consider a new version of an old example. John, a police investigator, finds Smith's body. It is an unnerving sight, but John is a well-trained officer and, though repelled by his discovery, is determined to do his job. Finding the murder weapon, a knife, he checks it for fingerprints. Fortunately, the apparent culprit has left clear fingerprints on the handle. At this

¹In fairness to Barwise and Perry, they do acknowledge that their notion of the value-loaded use of a definite description does not entirely account for the referential use [1, p. 152 n. 3].

point John utters in total revulsion: "The man whose fingerprints these are, whoever he is, whoever he may be, is insane!"

As it happens, Smith's murderer is quite sane and he was careful to wear gloves during the murder. Moreover, the fingerprints are actually those of another man, Max, who used the knife an hour before the murder, and by a strange twist of fate, this Max (Mad Max) is known to be insane, having spent most of his life in an asylum.

Now, does John use the description "The man whose fingerprints these are" referentially or attributively? Let's look at the facts. John intended to speak about Smith's murderer, not about Max, and what he said about Smith's murderer was that he was insane. Hence he said something true or false about Smith's murderer regardless of whether Smith's murderer was in fact the man whose fingerprints were found on the murder weapon. Thus, John must have been using the description *referentially*. On the other hand, John had no particular person in mind. He said what he said about Smith's murderer, whoever he might be. Thus, the description must have been used *attributively*.

The problem is that Donnellan's distinction between the referential and the attributive uses of definite descriptions cuts across the boundaries of two independent linguistic parameters:

- Sometimes an object must uniquely fit the description used if the utterance is to be about anything, while, at other times, it does not.
- Sometimes a speaker refers to a particular object he has in mind — one that may or may not satisfy the description employed. At other times, the speaker has no particular object in mind at all.

These facts provide two intuitive criteria for deciding whether a particular use of a definite description is referential or attributive. The first criterion is based on the role of the descriptive content of the referring expression in determining truth conditions. I shall call it the *denotation criterion*:

Denotation Criterion: If the description must denote one and only one object for the utterance to be about anything, its use is attributive. Otherwise, it is referential.

The second criterion is based on an intuitive (and rather vague) characterization of the speaker's mental state:

Mental-State Criterion: If the speaker has a particular object in mind when he refers, his use of the referring expression is taken to be referential. Otherwise it is attributive.

Whenever Donnellan's distinction is discussed, it is tacitly assumed that the two criteria are equivalent: any use of a definite description that is referential according to the denotation criterion should also be classified as referential according to the mental-state criterion and similarly as regards the attributive use. But this is simply not the case. As my example shows, some uses are attributive according to one criterion, referential according to the other. The two criteria, along with the underlying linguistic phenomena, are indeed in need of elucidation, but there is no reason to lump them together. It is therefore misleading to talk about *the* referential/attributional distinction, since there are actually two distinctions: there is Donnellan's distinction as interpreted by the denotation criterion, and there is Donnellan's distinction as interpreted by the mental-state criterion. Each is logically independent of the other.²

The denotation criterion itself is quite clear; what remains to be seen is how to interpret the phenomena it characterizes within a computational model of referring. The mental-state criterion, on the other hand, is much less tangible. All we have at this stage is an intuition and a metaphor to express it, namely, the metaphor of "having a particular object in mind." Definite descriptions can be used (so it is claimed) either while the speaker is in a mental state that is directed toward a particular object, or while the speaker is not in any such a state. But what could such a mental state be? An answer to this question requires that we untangle two conflicting interpretations that are employed indiscriminately in formal characterizations of *de re* beliefs.

3 Having a particular object in mind

As I have argued elsewhere [14], there is a tendency to associate the referential use with *de re* propositional attitudes. If a description is used referentially to convey a belief, that belief, it is claimed, is *de re*. This idea clearly reflects the mental-state criterion: to have an object "in mind" is to have a *de re* attitude toward it. But what is it for an agent to have a *de re* attitude? Many have tried to answer this question by studying the conditions under which *de re reports* of attitudes are true. Suppose I report Ralph's belief by saying

1 *Ralph believes that someone is a spy.*

²A similar point is made by both Loar [16] and Wettstein [18]. Their analysis of Donnellan's distinction, however, is entirely different from mine.

My report is open to two interpretations. I may have meant to say that Ralph believes that spies exist, or that there is a *particular* person whom Ralph suspects of being a spy. The former interpretation is a *de dicto* report of belief, while the latter is *de re*. We can characterize the *de re* interpretation by saying that Ralph has a particular person in mind whom he suspects of being a spy. This is why we regard the study of *de re* reports as being relevant for understanding the mental-state criterion.

Formally, we obtain *de re* reports of beliefs by a simple application of existential generalization. From the truth of

2 *Ralph believes that Saul Kripke is a philosopher,*

we should be able to conclude that

3 *There is a person about whom Ralph believes that he is a philosopher.,*

which is a *de re* report of Ralph's belief. But, as Quine [17] has pointed out, constructions such as (3) lead to serious difficulties. For, although the move from (2) to (3) is justified, a similar move from

4 *Ralph believes that Santa Claus lives at the North Pole*

to

5 *There is a person about whom Ralph believes that he lives at the North Pole*

is not so justified, since Ralph's belief in Santa Claus is not enough to bring Santa Claus into existence. Similarly, from the obviously true statement,

6 *Ralph believes that Smith's murderer is Smith's murderer,*

we cannot derive

7 *There is a person about whom Ralph believes that he murdered Smith,*

since Ralph may have no idea who the murderer might be. Note that, if we substitute "Jones" for one occurrence of "Smith's murderer" in (6), the resulting sentence,

8 *Ralph believes that Jones is Smith's murderer,*

may very well be false even if Jones is indeed the culprit. Thus, not only existential generalization fails. Substitution of coreferring terms also does not work.

The problem is well known. Reports of beliefs are *intensional*: they create contexts in which the laws of substitution and existential generalization are not guaranteed to be valid forms of inference. What should be ascertained, therefore, is when existential generalization and substitution work in *de re* reports of beliefs, and when they do not. The various suggested solutions to this problem originate in two distinct ideas that should be carefully separated.³

The first idea is that existential generalization and substitution are allowable only if the agent's conception of the referent is "vivid" enough [11]. The strongest version of this requirement specifies that the agent should know *who* or *what object* the referent is [8]. This is supported by the fact that the more the agent knows about the object of his belief, the more comfortable we feel in substituting other referring expressions for the one used by the agent when we report his belief. As we have seen, it would be preposterous, on the basis of Ralph's belief in (6) and the fact that Jones is the murderer, to assert (8). Yet, had Ralph known who the murderer was, we would have then felt perfectly comfortable in asserting either (7) or (8). Such considerations I call the *epistemic* intuition underlying *de re* reports of belief.

In contrast to the epistemic intuition, we have what I call the *modal* intuition. This idea stems from a consideration of what is logically possible: from an intuitive standpoint, it makes sense to ask not only whether a true proposition could have been false under different circumstances, but also whether a given person could have been different from the way he or she is. For example, Saul Kripke is the author of *Naming and Necessity*, but surely he might not have been. That is, although Saul Kripke in fact has the property of being the author of *Naming and Necessity*, we can easily conceive of alternative circumstances in which he would lack this property. Note that for any property that Kripke may possess, it makes sense to *ask* whether he might not possess it. The *answers* to such questions may vary: some of Kripke's properties are perhaps essential to his identity. Nevertheless, no matter what the answers are, the questions themselves are not meaningless.

³Intensional contexts such as belief reports give rise to a host of other problems. Why do existential generalization and substitution sometimes fail? Do words in intensional contexts change their meaning? What values can be assigned to free variables in such contexts? Important as they are, these questions are not relevant to our concern, namely, determining when an agent can be said to "have a particular object in mind."

When we ask such questions about Kripke, what we are doing is this: through a mental process of abstraction, we regard Kripke not as being the author of *Naming and Necessity*, or as someone named "Saul Kripke," or, for that matter, as anything else. We simply focus our attention upon *him* and, for a while, ignore any attributes he in fact has. The ability to do this is what I call the modal intuition, of which Kripke himself has provided an eloquent presentation [12]. The modal intuition does not mean, of course, that there really exists such an entity as Kripke-without-properties to whom all of Kripke's attributes are affixed. Although this view has a long and respected history, I do not share it, nor does it really matter for our purpose whether in the last analysis, it is right or wrong. My only point is that, when we contemplate an object, we are capable of considering *this very object* without, necessarily, reflecting upon any particular property that the object may have. It is the modal intuition that enable us to make sense of expressions such as

9 *The inventor of the light bulb might not have been the inventor of the light bulb.*

In this particular case, the first occurrence of "The inventor of the light bulb" only *fixes the reference* (to use Kripke's term). Once the reference has been fixed, we consider Edison himself — not as being the inventor or as being anything else — and we ask ourselves whether a possible world exists in which Edison is *not* the inventor. This is the modal intuition at work, and the way it expresses itself in natural language is through the use of what Kripke (*ibid.*) has called *rigid designators* — i.e. referring expressions that designate the same object in each possible world in which the object exists. Note that the converse of the modal intuition is consideration of the referent *qua* having a certain property. In some circumstances, we may be interested not so much in Kripke himself, but in Kripke only in so far as he is the author of *Naming and Necessity*. We may refer to Kripke during a discussion of the causal theory of proper names, but it is obvious that, if *Naming and Necessity* had been written by Gödel, we would have been talking about Gödel, instead of Kripke.

Now, the modal intuition plays a role similar to that of the epistemic intuition in interpreting *de re* reports of belief. To see how, note first that the modal notions of necessity and possibility create intensional contexts in the same way as do belief reports. For example, the statement

10 *Necessarily Smith's murderer is Smith's murderer*

is surely true, since nothing can fail to be identical to itself. But from (10) we cannot conclude that

11 *There is a person who is necessarily the one that murdered Smith,*

since it is hardly plausible to assume that being a murderer is a necessary property of the actual killer. Note the similarity in structure between the failure to derive (11) from (10) and the failure to derive (7) from (6). At the same time, as was the case in belief reports, there are occasions when existential generalization does work. For example, from the true statement

12 *Necessarily the number nine is odd,*

we can certainly conclude that

13 *There is a number that is necessarily odd.*

Existential generalization works here for reasons that clearly have nothing to do with the epistemic intuition. They have to do rather with our *modal* intuition, i.e., with the fact that the numeral “nine” in (12) is a rigid designator. When using the numeral “nine”, we are not considering the integer 9 as being, say, the number of planets, the square root of 81, or anything else, but are regarding it rather as a thing “in itself,” so to speak. Since statement (12) is true, the proposition *nine is odd* will be true in all possible worlds. Since “nine” is a rigid designator, one and the same number (i.e., the number 9) will be odd in all possible worlds; hence (13) is true as well.

Thus, two distinct ideas are employed to justify applying existential generalization within intensional contexts. The epistemic intuition requires that the agent have knowledge of the object if a *de re* report of his belief is correct. The modal intuition requires that a rigid designator be used if existential generalization is to be allowed. Characterized in this manner, the two notions appear to exhibit a natural “division of labor.” The modal intuition is applied to the explication of necessity and possibility, while the epistemic intuition is applied to beliefs. But as the logic of modal concepts and the logic of belief are both couched in terms of possible-world semantics, *both* intuitions serve as a basis for applying existential generalization to reports of beliefs. Consider the formula

14 $(\exists x)\text{BELL}(\text{Ralph}, \text{Insane}(x)),$

which roughly states that there is someone Ralph believes to be insane. Under the standard possible-world interpretation, (14) is true just in case there is a person (at least one) such that, in each possible world compatible with Ralph’s beliefs, that very same person is insane. Now suppose that the following statement is a correct report of Ralph’s belief:

15 *Ralph believes that Smith's murderer is insane.*

If Ralph knows who the murderer is, then, in each possible world compatible with Ralph's beliefs, Ralph can identify the man who actually murdered Smith, and in each such world that man is insane — which means that the conditions under which (14) is true are satisfied. Obviously, this is a manifestation of the epistemic intuition. But now consider the statement

16 *Ralph believes that Jones is insane.*

If Kripke is right, the name "Jones" is a rigid designator. That is, it picks out the same individual in every possible world in which that individual exists, independently of any property he may have. This, of course, is the modal intuition at work. Suppose that Ralph does *not* know who Jones is. Nevertheless, since "Jones" is a rigid designator, it denotes the same individual (i.e., Jones) in each possible world in which Jones exists. Since Ralph believes that Jones exists, it follows that Jones does indeed exist in every possible world compatible with Ralph's belief, and that, in each such world, Jones (i.e., whoever the use of the name "Jones" in (16) designates) is insane. Thus, we can again conclude that (14) is true, even though Ralph does not know who Jones is. The move from (15) to the formula in (14) is justified on epistemological grounds, while the same move from (16) is justified on modal grounds.

Let us first consider the modal intuition and its relation to Donnellan's distinction. The class of referring expressions most frequently associated with the property of rigid designation is the class of proper names, but Kaplan [9,10] has argued that demonstratives are also endowed with this property. Moreover, we can use a definite description as a verbal mode of pointing the sole purpose of which is to help "fix the referent." This stratagem turns definite descriptions into rigid designators; consequently, when a definite description is employed in this manner, we can report what is said by using the *de re* form. Kaplan introduces the artificial operator DTHAT to mark a definite description that is used merely as a verbal way of pointing. For example, the sentence

17 *The inventor of the light bulb was a genius*

is about whoever invented the light bulb. But the sentence

18 DTHAT(*"The inventor of the light bulb"*) was a genius

is about Edison. The difference between the two is this: in a possible world in which the light bulb was invented by Einstein and in which Edison was the village idiot, (17) is true, whereas (18) is false.

As Kaplan acknowledges [10, 234, 238], there is a close relation between the referential use and his treatment of definite descriptions as demonstratives. When a speaker uses a description referentially, he uses it as a rigid designator. But which of the two criteria of Donnellan's distinction should be so characterized? In the fingerprint example (p. 3 above), the description "The man whose fingerprints these are" is classified as referential according to the denotation criterion. Since it is *not* being used as a rigid designator (in every possible world the referent is the murderer, whoever he or she may be in that world), the analysis of the referential use in terms of rigid designation is *not* an explication of the denotation criterion. Hence it is somehow connected to the mental-state criterion (which is not surprising, given the close relation between rigid designation, the modal intuition, and the semantics of *de re* reports of belief). But the *epistemic* intuition is surely relevant to the mental-state criterion as well. According to the epistemic intuition, knowledge of the referent is a necessary condition for having the referent "in mind." Note that one cannot simply decide to possess knowledge, so that the question of whether a given definite description is used referentially (in the original, intuitive sense of "referential") is not entirely a function of the speaker's intention. At the same time, Kaplan's DTHAT, shows that a speaker is always free to decide whether or not to use a referring expression as a rigid designator.⁴

Thus, the mental-state criterion has two separate aspects. According to the epistemic aspect, knowledge of the referent is essential for a description to be used referentially. According to the modal aspect, the intention to use a definite description as a rigid designation is the defining feature of referential usage. If we add these two aspects to the *denotation criterion*, we get *three* aspects of Donnellan's distinction — the epistemic, the modal, and the denotational — that are conceptually independent of one another. They can be expressed in terms of three dichotomies:

- **Mental State Criterion (epistemic aspect)** — Knowing who or what the referent is, or at least having knowledge of the referent (referential use), in contrast to lacking such knowledge (attributive use).

⁴In discussions about the formal representation of knowledge, it is sometimes argued that an agent knows who a certain person is if the agent possesses a referring expression that rigidly designates that person. This view confuses the epistemic with the modal intuition: as Kaplan's DTHAT shows, any referring expression can be turned into a rigid designator, but this has nothing to do with knowing who the referent is.

- **Mental-State Criterion (modal aspect)** — Intending the referring expression to be interpreted as a rigid designator (referential use) in contrast to considering the referent *qua* having a particular property (attributive use).
- **Denotation Criterion (denotational aspect)** — Using a definite description “the so-and-so,” to refer to whoever is the so-and-so; if nothing is, the speech act cannot succeed (attributive use). This is in contrast to using a definite description to refer to an object x , whether or not x is indeed denoted by that description (referential use).

These three aspects are summarized in Figure 1.

4 A three-tiered model of referring

The three aspects of Donnellan’s distinction — the epistemic, the modal, and the denotational — are conceptually independent of one another. Yet they must all be bound together in an obvious way, or else Donnellan’s distinction would not have such a persuasive ring to it. What needs to be explained, therefore, is the intuitive immediacy of the distinction, given its complexity. How is it that there is in fact a single referential/attributive distinction and not three?

The answer, no doubt, must lie in the way the three aspects interact with one another. To show how this is done, I propose to correlate aspects of Donnellan’s distinction with the logical components that a plan-based model of referring must possess. Any such model must contain the following elements: (1) a *database* that includes representations of objects, (2) a *planner* that constructs strategies for carrying out referring intentions, and (3) an *utterance generator* that produces referring expressions. I show how each aspect

Donnellan’s Distinction					
Mental-State Criterion				Denotation Criterion	
Epistemic Aspect		Modal Aspect		Denotational Aspect	
Referential	Attributive	Referential	Attributive	Referential	Attributive
Knowledge of the referent	No knowledge of the referent	Rigid designation	Nonrigid designation	Reference without denotation	No reference without denotation

Figure 1: Aspects of Donnellan’s distinction.

of Donnellan's distinction could be represented in the corresponding computational component. Then, by demonstrating how each component interacts with the others in the course of a speech act, we can see how mind and language cooperate to produce what we call the referential and the attributive uses of definite descriptions. Then, leaving Donnellan's distinction behind us, we can concentrate on each of its aspects and the emergent referring model.

4.1 Individuating sets

According to the *individuation principle* [14], if an agent has a belief about a particular object under a presentation mode, the latter determines a unique object in all possible worlds that are compatible with the agent's beliefs. It is not yet clear what, precisely, presentation modes are supposed to be. Yet, given the fact that, whatever they are, they must satisfy the individuation principle, the most plausible accounts take them to be some sort of internal representations. Such internal representations serve as a mechanism through which an agent can individuate objects within his conceptual scheme. This is done, under such an interpretation, by the relation of denotation: each presentation mode is believed by the agent to denote one and only one object.

Such a view sounds very much in line with the descriptive program of reference, but it need not be. First, as I have argued elsewhere [14], the causal theorist can accept such an account of presentation modes while insisting that they alone do not determine the actual referent. Second, it is still left open to question whether *every* belief must contain a mode of presentation. All that the individuation principle entails is that, *whenever* a mode of presentation is needed, the agent must believe that it individuates an object. A causal theorist can still claim that some *de re* beliefs are special in that they do not require presentation modes. Third, to say that modes of presentation should be interpreted as internal representations still does not tell us what kind of representations they are. They need not individuate objects, for example, by means of general terms alone. Some modes of presentation may very well individuate objects only relative to other objects. Thus, as far as I can see, nothing prevents a causal theorist from accepting an account of modes of presentation as internal representations. Doing so, however, does not mean that presentation modes can always be articulated as linguistic descriptions. Nor should the claim that a presentation mode is believed by the agent to denote a unique object be taken too literally. It does not mean, for example, that we are always *aware* of the presentation modes under which we have beliefs about particular objects. Sometimes we

are, but certainly not always.

We thus have modes of presentation, each believed by the agent to denote a particular thing. Let an *individuating set* be an exhaustive list of presentation modes, all taken by the agent to denote the same object. For example, the set

{*author of the Odyssey, author of the Iliad*}

is an individuating set for me, since (1) I happen to believe that one and only one person composed both the *Odyssey* and *The Iliad*, and (2) I must admit that I do not know of any other fact that *individuates* that Greek author.⁵

Individuating sets vary a great deal in informational content. The singleton {*shortest spy*} is an individuating set representing the shortest spy for me. On the other hand, the rich cluster of images and definite descriptions representing my mother for me is also an individuating set, although I could not even begin to enumerate all the presentation modes it contain. When all is well, and the presentation modes in a particular individuating set do denote a unique object, I shall say that the individuating set *determines* that object. But many things can go wrong. A speaker may possess two distinct individuating sets that, unbeknownst to him, determine the same object (e.g., Oedipus's distinct conceptions of his mother and his wife). On the other hand, an agent may possess an individuating set containing two presentation modes that actually denote different objects (e.g., when a person mistakenly thinks that the author of *Sense and Sensibility* also wrote *How to Do Things with Words*). Of course, some individuating sets may include presentation modes that do not denote anything (e.g., an individuating set containing *present king of France*), while some individuating sets may determine nothing at all (e.g., a child's conception of Santa Claus).

Whether or not an agent can have knowledge of an object, or knows who or what the referent is (the epistemic aspect of Donnellan's distinction) depends on the relevant individuating set. On first approximation, the richer in content the individuating set, the more likely it is that the agent knows who or what the referent is. This, however, is far from being universally true. Some modes of presentation may be *privileged*, depending on circumstances [2]. For the purpose of discussing computation theory, I can say that I know who Stephen Cook is: he is the person who proved that the satisfiability

⁵This is not entirely accurate. I do, in fact, have other modes of presentation for the Greek Homer: for example, *person believed by my mother to have composed both the Iliad and the Odyssey*. My concern here, however, is with an abstract data structure for modes of presentation, whatever they are. I am not interested at this point in the correct enumeration of presentation modes.

problem is NP-complete. For most other purposes (e.g., introducing him at a party), I cannot claim to know who he is, since I have never made his acquaintance. Nevertheless, as we shall see, the different contextual requirements can be expressed as constraints on the relevant individuating set.

Individuating sets are part of the database, and this is where the epistemic aspect of Donnellan's distinction should be represented. It is useful to distinguish between the quasi-permanent individuating sets that represent relatively stable knowledge of objects and what may be called *local* individuating sets that are constructed and discarded in the course of a conversation.

4.2 Referring intentions

Individuating sets, per se, have nothing to do with speech. Nevertheless, they become involved when a speaker plans to make a hearer recognize what the speaker is talking about. We should distinguish, though, between two ways in which individuating sets are employed in the process of such planning, corresponding to two types of referring intentions. First, a speaker may select a mode of presentation from the relevant individuating set, intending that *this particular presentation mode* be recognized by the hearer. Second, the speaker may intend to refer to an object determined by an individuating set, but without at the same time intending that any *particular* presentation mode from the set be part of the proposition he wishes to express. Consider, for example, the following two statements:

19 *The author of Othello wrote the best play about jealousy.*

20 *Shakespeare was born in Stratford-upon-Avon.*

In making both statements, a speaker would normally be referring to Shakespeare. But note the difference between the two referring intentions. In statement (19), the speaker selects a particular presentation mode of Shakespeare, namely, that he is the author of *Othello*, and intends to make the hearer think of Shakespeare in terms of this presentation mode. If the hearer fails to do so, he will miss the whole point. In statement (20), on the other hand, the speaker does not select any particular presentation mode of Shakespeare from the relevant individuating set. Indeed, he may not care at all *how* the hearer makes the connection between the name "Shakespeare" and the referent, as long as he appropriately identifies who the speaker is talking about.

In [13], the difference between these two types of referring intentions is spelled out more precisely. It should be clear, however, that the two types

of intentions result in two distinct types of propositions that a speaker may express. In (19), the presentation mode is a constituent of the proposition expressed. In (20), no particular mode of presentation is meant; consequently, we may say that the proposition expressed is a *singular* one, involving, as it were, (1) the property of being born in Stratford and (2) Shakespeare himself. This seems to me the most natural interpretation of singular propositions in terms of speech act theory.⁶

The two types of referring intentions correspond to the modal aspect of Donnellan's distinction. In (19), Ralph's hearer is supposed to think of Shakespeare *qua* being the author of *Othello*. In (20), the hearer is expected to interpret the referring expression as a rigid designator. Since the planner is where referring intentions (as well as the procedures for carrying them out) are represented, this is where the modal aspect belongs. Note that the two types of referring intentions can be described as intentions to impose various constraints (including the null constraint) on the way the hearer should think of the referent. In [15], this is generalized to other referring intentions — for example, to the intention that the hearer identify appropriately what is being talked about.

4.3 Choice of referring expressions

Once the speaker knows what particular mode of presentation (if any) he means, he must pick out a referring expression to be used in the speech act. It is natural to assume that if, say, the mode of presentation is *author of Othello*, the definite description "the author of *Othello*" would be a natural choice; if no particular presentation mode is relevant, a proper name would be most appropriate. But this, obviously, is not at all necessary. It would be very tiresome for the speaker to keep saying "the author of *Othello*" over and over again whenever reference to Shakespeare is called for. Conversely, even if it makes no difference to the speaker how his audience comes to think of the referent, he may surmise that no one around is familiar with any of the names he might offer, and that a definite description would better serve his communication goals. In any case, whether or not a definite description must denote one and only one object for the speech act to succeed (the *denotation criterion*) depends on the relation between that description and the particular presentation mode (if any) that the speaker regards as important. If the description expresses that same presentation mode, unique denotation

⁶A hard-core descriptive theorist might be tempted to say that the mode of presentation associated with "Shakespeare" in Statement (20) is the entire individuating set. But a speaker cannot *mean* such a mode of presentation, since the individuating set is not accessible to the hearer.

is required. If the relevant presentation mode is only implied (as was the case in the fingerprint example), or if no particular mode of presentation is meant, denotation is superfluous.

As the choice of the actual referring expressions to be used is made within the utterance generator, it is there that the denotational aspect of Donnellan's distinction should be represented.

4.4 Donnellan's distinction: final chord

We end up with a correspondence between Donnellan's distinction and an outline of a referring model. The nature of the relevant individuating set, the type of referring intention, and the actual choice of a referring expression correspond to the epistemic, modal, and denotational aspects, respectively. Such a correlation is not a magical coincidence, of course. It is another reflection of the three tiered structure of beliefs, propositions, and expressions [14] It should not be too surprising, then, that the epistemic aspect of Donnellan's distinction is concerned with representations of objects in the mind, that the modal aspect is manifested by the structure of the proposition that a speaker intends to express, that the denotational aspect is concerned with utterances of referring expressions, and that Donnellan's distinction itself is an ambiguous formulation of two natural ways in which mental representations, propositions and utterances are interrelated in the course of a speech act. We can now utilize this correspondence among Donnellan's distinction, cognitive elements in the production of a speech act, and the computational components in explaining the structures underlying typical referential and attributive uses of definite descriptions.

In the paradigmatic examples of attributive usage, the speaker selects a *particular* mode of presentation (the modal aspect) and, not surprisingly, chooses a definite noun phrase that expresses that presentation mode (hence the denotational aspect). As Donnellan himself notes [6], *lack* of knowledge concerning the referent (the epistemic aspect) is not a prerequisite of at-

REF/ATTRIB ASPECT	THEORETICAL INTERPRETATION	COGNITIVE ELEMENT	COMPUTATIONAL COMPONENT
Epistemic	Type of individuating set	Attitude	Database
Modal	Type of referring intentions	Proposition	Planner
Denotation	Choice of noun phrase	Expression	Utterance generator

Figure 2: Donnellan's distinction and its computational interpretation.

tributive usage. One can know who murdered Smith, but still insist that *anyone* who would have murdered Smith in such a brutal way must be insane. Nevertheless, the attributive use typically occurs when the identity of the referent is not known, because, under such circumstances, the intention to refer to an object *qua* being the such-and-such is easier to recognize. When the identity of the referent is well known to all participants in the conversation, the speaker must frequently make his intention to refer to the object *qua* having a certain property more explicit. Moreover, when the speaker has no knowledge of the referent whatsoever, his use of a definite description is not likely to be intended as a rigid designator. Now we can see why the attributive side of Donnellan's distinction encompasses a natural category of uses of definite descriptions. Nonrigid designation tends to accompany a particular choice of a definite description, usually when knowledge of the referent is lacking.

In the paradigmatic examples of referential usage, on the other hand, knowledge of the referent must be possessed; at the same time, the speaker does not take any *particular* mode of presentation to be of importance. The speaker's goal is simply to have the hearer identify appropriately what is being talked about, a task for which any well-suited referring expression will be acceptable. Moreover, when knowledge (including mutual knowledge) of the referent exists, identification can be achieved, to a large extent, apart from the descriptive content of the referring expression. Hence, strict denotation is frequently unnecessary.

This completes my analysis of Donnellan's distinction. Figure 2 shows in condensed form the correspondence between Donnellan's distinction and the proposed outline of a referring model. The first column lists the three aspects of Donnellan's distinction; the second summarizes how these aspects are to be interpreted; the third enumerates the corresponding cognitive elements that contribute to the production of a speech act; the fourth shows the computational components of the model.

In [14], I have used Donnellan's distinction as a methodological tool in presenting objections to the descriptive program. The role its analysis plays in the present chapter is twofold. First, because of the attention accorded to Donnellan's distinction in both computational linguistics and philosophy, an understanding of its underlying cognitive structure is valuable in itself. Second, the process of dissecting the various aspects of referential and attributive usage provides an outline of a computational model of referring whose main data structure is that of an individuating set and in which the three aspects of Donnellan's distinction can be represented.

From the standpoint of this study, however, Donnellan's distinction can-

not be regarded as an end in itself. We do have bigger fish to fry. The real point is not to represent a complex distinction, but rather to formulate general principles of a computational model of referring. We must therefore shift our perspective; instead of assigning central stage to referential and attributive usage, we can treat Donnellan's distinction as a *test case* for a putative theory of referring as a speech act. Any such theory must take Donnellan's distinction into account. That is, it must at least specify (1) what it takes for an agent to have *de re* beliefs, (2) what makes an agent express a general proposition rather than a singular one (or vice versa), and (3) what the reasons are for preferring one referring expression to another. If a theory of referring as a speech act cannot handle these questions, it stands little chance of being adequate.

My contention is that the three tiered structure proposed in this chapter is a suitable outline for such a computational theory of referring. That a theory based on this structure is likely to handle Donnellan's distinction correctly may not be very surprising. After all, the foregoing questions are precisely the ones that motivated the three tiered structure in the first place. Nevertheless, answering these questions is far from being a trivial matter: once the complexity of Donnellan's distinction is understood, it is clear that accommodating it is a challenging objective for any theory of referring. Thus, it may be methodologically advantageous to proceed with the three aspects of Donnellan's distinction kept constantly in mind. Of course, the theory must account for numerous other issues as well. There is much more to using a referring expression than simply choosing between the referential and the attributive.

References

- [1] Jon Barwise and John Perry. *Situations and Attitudes*. The Massachusetts Institute of Technology Press, Cambridge, Massachusetts, 1983.
- [2] Stephen E. Boër and William G. Lycan. *Knowing Who*. Bradford Books, The MIT Press, Cambridge, Massachusetts, 1986.
- [3] Hector-Neri Castañeda. On the philosophical foundations of the theory of communication: reference. *Midwest Studies in Philosophy, II: Studies in the Philosophy of Language*, 1977.
- [4] Charles Chastain. Reference and context. In K. Gunderson, editor, *Language, Mind and Knowledge*, University of Minnesota Press, Minneapolis, Minnesota, 1975.

- [5] Philip R. Cohen. Referring as requesting. In *Proceedings of the 10th International Conference on Computational Linguistics*, pages 207–211, 1984.
- [6] 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.
- [7] Barbara J. Grosz, A. Joshi, and S. Weinstein. Providing a unified account of definite noun phrases in discourse. In *Proceedings of the Twenty-first Annual Meeting*, pages 44–50, Association for Computational Linguistics, 1983.
- [8] Jakkko Hintikka. *Knowledge and Belief*. Cornell University Press, Ithaca, New York, 1962.
- [9] David Kaplan. Demonstratives. 1977. University of California, Los Angeles. Photocopy.
- [10] David Kaplan. Dthat. In Peter Cole, editor, *Syntax and Semantics, Volume 9*, Academic Press, New York, New York, 1978.
- [11] David Kaplan. Quantifying in. In Davidson and Hintikka, editors, *Words and Objections: Essays on the Work of W. V. Quine*, D. Reidel Publishing Co., Dordrecht, Netherlands, 1968.
- [12] Saul Kripke. *Naming and Necessity*. Harvard University Press, 1980.
- [13] Amichai Kronfeld. *The Referential Attributive Distinction and the Conceptual-Descriptive Theory of Reference*. PhD thesis, University of California, Berkeley, 1981.
- [14] Amichai Kronfeld. *The descriptive approach to reference: why it is difficult to live with, and why we have to*. Technical Note, Artificial Intelligence Center, SRI International, 1988.
- [15] Amichai Kronfeld. *The literal goal and discourse purpose of referring*. Technical Note, Artificial Intelligence Center, SRI International, 1988.
- [16] Brian Loar. The semantics of singular terms. *Philosophical Studies*, 30:353–377, 1976.
- [17] W.V. Quine. Quantifiers and propositional attitudes. In D. Davidson and G. Harman, editors, *Semantics of Natural Language*, D. Reidal Publishing Co., Dordrecht, Holland, 1972.
- [18] Howard Wettstein. Demonstrative reference and definite descriptions. *Philosophical Studies*, 40(4):241–257, 1981.