A Context Change Semantics of Donnellan’s Distinction

Atsushi Shimojima

Department of Philosophy
Indiana University
ashimoji@bronze.indiana.edu

The common ground of a discourse is the shared information that each participant of the discourse assumes (and often believes) to be true in order to interpret other participants’ utterances. In a context change model of discourse, the semantic value of an utterance is identified with its context change potential, namely, the impact the utterance would make on a common ground if interpreted and accepted on that common ground. The semantic value of an utterance is hence modeled as a function from (a model of) a common ground to (a model of) a common ground. The primary purpose of this paper is to develop a context change model of what Donnellan (1966) called the “attributive” use and the “referential” use of definite descriptions.

For this purpose, I present a scenario of typical discourse that involves both the attributive and the referential use of a definite description. After showing that classical models, with simple applications of the “familiarity theory,” fail to simulate the context change in the scenario, I propose a model that augments a classical model with partial functions, called “replacer functions,” defined on discourse referents. I argue (i) that the attributive-referential distinction is accurately modeled as the difference of the replacer functions on which the speaker uses and the interlocutors interpret a definite description, and (ii) that there is no need in a correct model of Donnellan’s distinction to posit a semantic ambiguity of a definite description. If we interpret a discourse referent as one of the several “guises” of an individual object, and a replacer function as the specification of the guises we refer to by definite descriptions (referential perspective), the proposed model can be seen as a formal development of the intuitive idea that the attributive-referential distinction is the difference of guises (of a same individual) a definite description selects in each occasion.
1. Basic Scenario

Act 1

We are speaking with Holmes in front of Smith’s mutilated body. We (and Holmes himself) trust in Holmes’ ability in criminal investigation so that any of his findings automatically enters into our common ground. First, Holmes says (1). Later, Holmes comes to say (2). Holmes then interviews Jones, and reports his finding by (3).

(1) Smith’s murderer is insane. (attributive)
(2) Jones is Smith’s murderer.
(3) Smith’s murderer is a Buddhist. (referential)

To verify the “contents” of our common ground after these utterances, we let one of the interlocutors, say Watson, make various utterances, and try to evaluate his statements on the basis of the information we currently possess.1

<table>
<thead>
<tr>
<th>Watson utters:</th>
<th>We evaluate:</th>
<th>Watson utters:</th>
<th>We evaluate:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jones is insane.</td>
<td>YES</td>
<td>Jones is a Buddhist.</td>
<td>YES</td>
</tr>
</tbody>
</table>

At this point, we have information about both Jones’ mental state and his religion.

Act 2

Later, Holmes revises his previous statement, saying (4).

(4) Bond, not Jones, is Smith’s murderer.

Our evaluations will be:

<table>
<thead>
<tr>
<th>Watson utters:</th>
<th>We evaluate:</th>
<th>Watson utters:</th>
<th>We evaluate:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Jones is insane.</td>
<td>DON’T KNOW</td>
<td>Jones is a Buddhist.</td>
<td>YES</td>
</tr>
<tr>
<td>Bond is insane.</td>
<td>YES</td>
<td>Bond is a Buddhist.</td>
<td>DON’T KNOW</td>
</tr>
</tbody>
</table>

Roughly, we have removed insanity from Jones, and conferred it to Bond. We on the other hand left Buddhism with Jones.

Crucially, upon Holmes’ revision (4) of the identity of Smith’s murderer, the information about insanity moved from Jones to Bond, while the information about Buddhism stayed with Jones. Looking back to the original statements (1)

---

1 In each of the following acts, however, we will evaluate Watson’s utterances only to verify the impacts of Holmes’ utterances upon our common ground, and will “forget” Watson’s utterances after each act, as though he did not make any statements at all. Also, to avoid tacit influences of conversational implicature on our judgement, we assume that Holmes is not that kind of person who conveys something by means of implicature. That is, whenever his utterance is about to implicate something, he asserts it. In a sense, his assertions is closed under implicature.
and (3) in which each information was introduced, we have the impression that
the definite description “Smith’s murderer” in the statement (1) has switched its
refferent from Jones to Bond upon (4), while the same description in (2) has never
switched its referent from Jones, i.e. the impression that the former refers to

\textit{whoever murdered Smith}, while the latter refers to Jones \textit{rigidly}. This is a main
intuition that supports Donnellan’s distinction of attributive and referential use
of definite descriptions. Our first task is to develop a model that simulates the
context change involved in this scenario.

2. Simple Applications of the Familiarity Theory

According to the familiarity theory (Heim 1982, 1983a; Karttunen 1968, 1976),
the contribution of a definite NPs to context change process can be specified as
the following instructions:

(I) Let a definite NP pick an old discourse referent.

(II) Let a definite description do so according to its descriptive con-
tent.

It turns out that simple applications of the familiarity theory fail to capture
the context changes in our scenario. I will use the standard DRT (Kamp 1981;
Kamp & Reyle 1993) for illustration of the problems, and indicate occasionally
how the problems carry over to Heim’s file change semantics (Heim 1982, 1983a).
This is, however, by no means to criticize those frameworks. As we will see, the
half of the required mechanism is already provided by the notion of “discourse
referents,” and all that has to be done is to augment the mechanism to make a
fuller use of that notion.

Suppose that we have already introduced Jones, Bond, and Smith’s murderer
into our conversation. Our initial DRS then contains three different discourse
referents, say \(x\), \(y\), and \(z\), for those people respectively. Upon the utterance
(1) “Smith’s murderer is insane,” the familiarity constraints force us to ascribe
insanity to \(z\). That produces the following DRS:

\[
\text{K1} \quad \begin{array}{c}
\text{Jones}(x) \quad \text{Bond}(y) \quad \text{murder}(z, \text{Smith}) \quad \text{insane}(z)
\end{array}
\]

What shall we do upon the utterance (2) “Jones is Smith’s murderer”? Since
the utterance seems to assert the co-reference of \(x\) and \(z\), one of the simplest
ways of incorporating this utterance into our DRS is to replace \(z\) with \(x\). Let
us call this “the simple replacement strategy.” On this strategy, \(x\) takes over all
the positions of \(z\), producing the following DRS\(^2\):

\[
\text{K2} \quad \begin{array}{c}
\text{Jones}(x) \quad \text{Bond}(y) \quad \text{murder}(x, \text{Smith}) \quad \text{insane}(x)
\end{array}
\]

\(^2\) We could of course replace \(x\) with \(z\) or introduce a new referent and let it replace both \(x\)
and \(z\). But it will only produces an alphabetic variant of K2, which makes no difference to
prediction.
Now we receive the utterance (3) “Smith’s murderer is insane.” Since $x$ is now Smith’s murderer in our DRS, the whole utterance ascribes the property of being a Buddhist to $x$. The resulting DRS will be:

$$\begin{array}{c}
\{x, y\} \\
Jones(x) \quad Bond(y) \quad murder(x, Smith) \quad insane(x) \quad Buddhist(x)
\end{array}$$

Notice that $K_3$ justifies both the utterances “Jones is insane” and “Jones is a Buddhist.” It thus correctly predicts our affirmative judgments of Watson’s utterances in act 1.\(^3\)

To predict our judgments in act 2, however, our DRS must contain the following conditions after hearing the utterance (4) “Bond, not Jones, is Smith’s murderer”:

$$\{x, y\}$$

In a word, we have to replace $x$ with $y$ only in selected places. But there is nothing in the DRS $K_3$ or in the utterance (4) that guides us to this particular replacement. $K_3$ treats every occurrence of $x$ equally, without giving any hint about which occurrence to be affected by the utterance (4). The simple replacement strategy thus fails. (In Heim’s file change semantics, the simple replacement strategy corresponds to merging two Heim cards into one. It runs into a similar problem.)

This failure teaches us the following lesson:

**Lesson 1** Even after two (or more) discourse referents turn out to coincide in reference, the DRS must keep using separate discourse referents so that the conditions accumulated on them can be separated properly when the co-reference is canceled or further co-reference is found.\(^4\)

Let us take the “equation strategy” then. Instead of replacing one referent with another upon the utterances (2) and (4), we add new conditions that equate those referents. This strategy incorporates Lesson 1, since it allows us to keep two discourse referents separate while recording their co-reference. On this strategy, we find the following DRS after the utterance (2):

$$\begin{array}{c}
\{x, y\} \\
Jones(x) \quad Bond(y) \quad murder(y, Smith) \quad insane(y) \quad Buddhist(x)
\end{array}$$

\(^3\) I believe it sufficiently clear under what conditions I say that a DRS justifies an utterance. A DRS is said to justify an utterance if and only if updating the DRS with the utterance results in a logical consequence of the original DRS, with “logical consequence” defined as in the classical DRT (Kamp & Reyle 1991; 1993). Also, a DRS is said to justify an utterance if and only if updating the DRS with the utterance results in a contradiction, i.e. a DRS that is true in no model.

\(^4\) In fact, it was one of the original motivations for the notion of discourse referents that discourse referents can be treated as distinct objects while coinciding in reference. See Heim (1983a: 166).
By the familiarity constraint, however, we have to add the property of being Buddhist to $z$ upon the utterance (3):

$$\begin{array}{l}
x \quad y \quad z \\
Jones(x) \quad Bond(y) \quad murder(z, Smith) \quad insane(z) \quad x = z \\
\end{array}
$$

K4

This gives us a big trouble upon the next utterance (4). The best thing we can do upon (4) is to cancel the equation $x = z$, and add a new equation $y = z$. So we get:

$$\begin{array}{l}
x \quad y \quad z \\
Jones(x) \quad Bond(y) \quad murder(z, Smith) \quad insane(z) \quad Buddhist(z) \\
\end{array}
$$

K5

This DRS justifies the utterance “Bond is a Buddhist,” without justifying “Jones is a Buddhist,” failing to predict our evaluations of Watson’s utterances in act 2. To get the right result, then, we should have added the property of being a Buddhist to $x$ rather than to $z$ in K4 upon the utterance (3). But again, there is nothing in the DRS K4 or the utterance (3) that motivates that particular addition. The familiarity constraints directed us even to the contrary. Thus the equation strategy also fails at this point. (In Heim’s model, the equation method corresponds to concatenating two cards without overlap, or “connecting” two cards by an equation sign using cross-reference. It also leads to the same difficulty we have seen here.) We learn a second lesson:

**Lesson 2** A DRS needs an additional mechanism that directs which discourse referent an utterance should accumulate a condition upon when two (or more) discourse referents are equated and an NP picks one of them.

### 3. Proposing the Model

Lessons 1-2 indicate the (minimal) degree of complication a correct context change model has to assume. In the following, I will propose a way to incorporate these lessons into the standard DRT.

I propose to augment a standard DRS with a partial function, called the “replacer function,” defined on the set of discourse referents of the DRS. Thus, a DRS in my model is a triple of a set of discourse referents $U$, a set of DRS-conditions Con, and a replacer function $f: U \to U$. The idea is to let the replacer function serve as the additional mechanism that lesson 2 calls for. For this purpose, I make the following stipulations:
Stipulation 1 A definite NP follows the familiarity constraints to choose its initial discourse referent, applies the current replacer function to it, and chooses the obtained value as its final discourse referent.

Stipulation 2 Let $K = \langle U, \text{Con}, f \rangle$ be a DRS. The replacer function $f$ is warranted in $K$ if and only if $f$ is a total function from $U$ to $U$, and for all discourse referents $x$ and $y$ in $U$, if $f$ assigns $x$ to $y$, then $x$ is identical to $y$ or $\text{Con}$ contains the condition $x = y$ or $y = x$. The replacer function of a DRS must be warranted in the DRS.

Let us go through the context changes in our scenario again with our augmented notion of DRS and the above stipulations. Before Holmes’ first utterance (1), our DRS contains the following set of discourse referents and set of conditions:

$$\{x, y, z\}$$
$$\{\text{Jones}(x), \text{Bond}(y), \text{murder}(z, \text{Smith})\}$$

By Stipulation 2, then, there is only one replacer function that our DRS can employ at this point, i.e. the total identity function upon $\{x, y, z\}$. Thus, our initial DRS must be:

$$\begin{align*}
  & x \quad y \quad z \\
  & \text{Jones}(x) \quad \text{Bond}(y) \quad \text{murder}(z, \text{Smith}) \\
  & f_0 : x \to x \\
  & \qquad y \to y \\
  & \qquad z \to z \\
  \end{align*}$$

$$K_7$$

Now we receive the utterance (1) “Smith’s murderer is insane.” Since $f_0$ is a vacuous identity function, Stipulation 1 does not affect the way the NP “Smith’s murderer” picks its discourse referent—The NP picks $z$ as its initial referent, applies the replacer function $f_0$ to $z$, and ends up with the same referent $z$. Accordingly, the whole utterance (1) ascribes insanity to $z$, producing:

$$\begin{align*}
  & x \quad y \quad z \\
  & \text{Jones}(x) \quad \text{Bond}(y) \quad \text{murder}(z, \text{Smith}) \quad \text{insane}(z) \\
  & f_0 : x \to x \\
  & \qquad y \to y \\
  & \qquad z \to z \\
  \end{align*}$$

$$K_8$$

Then comes the statement (2) “Jones is Smith’s murderer.” Again, since $f_0$ is an identity function, it does not affect the ways the NPs “Jones” and “Smith’s murderer” in (2) pick their discourse referents. They simply pick $x$ and $z$ respectively, and the whole utterance (2) adds the equation $x = z$ to $K_8$, just as the simple equation strategy instructs. The crucial point is that, because
of this addition of the equation \( x = z \), our DRS now warrants four alternative replacer functions:

\[
\begin{align*}
  f_0 &: x \rightarrow x, \ y \rightarrow y, \ z \rightarrow z \\
  f_1 &: x \rightarrow x, \ y \rightarrow y, \ z \rightarrow x \\
  f_2 &: x \rightarrow z, \ y \rightarrow y, \ z \rightarrow z \\
  f_3 &: x \rightarrow z, \ y \rightarrow y, \ z \rightarrow x \\
\end{align*}
\]

It is not our concern, however, which replacer function our DRS employs as an immediate consequence of the utterance (3). Our main concern is which replacer function our DRS will employ later on, when we interpret Holmes’ third utterance (3) “Smith’s murderer is a Buddhist.”

It is a part of our scenario that Holmes makes this utterance to report the finding of his interview with Jones. Let us assume that this information makes the interlocutors employ the replacer function \( f_1 \) among other options\(^5\). Then, the DRS on which we interpret the utterance (3) is:

\[
\begin{array}{c}
\text{Jones}(x) \ \text{Bond}(y) \ \text{murder}(z, \text{Smith}) \ \text{insane}(z) \ x = z \\
\text{f}_1 : x \rightarrow x \\
\quad y \rightarrow y \\
\quad z \rightarrow x \\
\end{array}
\]

This time, the replacer function does affect the way the NP “Smith’s murderer” picks its referent: the NP picks \( z \) as its initial referent, applies the replacer function \( f_1 \) to \( z \), and ends up picking \( x \) as its final referent. The replacer function \( f_1 \) re-directs the NP to pick a discourse referent other than the one the NP picks initially. Accordingly, the whole utterance ends up ascribing Buddhism to \( x \) rather than \( z \), and produce:

\[
\begin{array}{c}
\text{Jones}(x) \ \text{Bond}(y) \ \text{murder}(z, \text{Smith}) \ \text{insane}(z) \ x = z \ \text{Buddhist}(x) \\
\text{f}_1 : x \rightarrow x \\
\quad y \rightarrow y \\
\quad z \rightarrow x \\
\end{array}
\]

It should be clear that the resulting DRS justifies both of Watson’s utterances in act 1.

What context change does the next utterance (4), “Bond, not Jones, is Smith’s murderer” bring about? Naturally, the insertion “not Jones” in (4) cancels the effect that the utterance (2) brought about, removing the equation

\(^5\) See section 5 for a more detailed description of how we select the replacer function \( f_1 \) at this point. Notice that we could assume, without affecting the later prediction, that we employ the replacer function \( f_3 \).
By definition, this removal of equation makes the replacer function \( f_1 \) unwarranted, and we are forced to replace \( f_1 \) with the identity function \( f_0 \). Thus, we have the following DRS when we interpret the remaining part of the utterance (4):

\[
\begin{align*}
\text{Jones}(x) \quad \text{Bond}(y) \quad \text{murder}(z, \text{Smith}) \quad \text{insane}(z) \quad \text{Buddhist}(x) \\
f_0 : x \rightarrow x \\
y \rightarrow y \\
z \rightarrow z
\end{align*}
\]

Since \( f_0 \) is an identity function, the NPs “Bond” and “Smith’s murderer” in (4) simply choose their initial referents, \( y \) and \( z \), as their final referents, and the part “Bond...is Smith’s murderer” of (4) introduces the equation \( y = z \). The set of conditions of the resulting DRS is:

\[\{\text{Jones}(x), \text{Bond}(y), \text{murder}(z, \text{Smith}), \text{insane}(z), \text{Buddhist}(x), y = z\}\]

The DRS justifies the utterances “Jones is a Buddhist” and “Bond is insane” without justifying “Jones is insane” or “Bond is a Buddhist.” Our model thus correctly predicts our judgments on Watson’s utterances in act 2. Again, what replacer function we employ as an immediate consequence of (4) is not our concern.

Notice that the proposed model is an exact reflection of the lessons we learnt in the previous section. First, it makes a full use of the feature of discourse referents that two discourse referents can be distinct while coinciding in reference (i.e. being equated by an identity condition), and (ii) it has an explicit mechanism (replacer functions) that directs which of two (or more) equated discourse referents a definite description should pick as its final referent. With this mechanism at hand, our model distinguishes an attributive reading from a referential reading by a difference of the replacer functions on which the interlocutors interpret a definite description.

In our model, an attributive reading arises when the replacer function directs a definite description to pick the same discourse referent that the definite description initially picks. In the case of the utterance (1) “Smith’s murderer is insane,” since the only warranted function that the interlocutors can employ is a vacuous identity function, the definite description “Smith’s murderer” inevitably picks its initial discourse referent \( z \) as its final referent, and the whole utterance ends up ascribing insanity to \( z \). Accordingly, whenever an identity statement “NP is Smith’s murderer” introduces a new equation \( y = z \), insanity re-applies to the discourse referent \( y \) equated to \( z \). This simulates our intuition that insanity applies to whoever turns out to be the murderer of Smith.

On the other hand, a referential reading arises when the replacer directs a definite description to pick a discourse referent that is equated to, but different
from, the initial discourse referent the definite description picks. In the case of the utterance (3) “Smith’s murderer is a Buddhist,” the replacer function directs the definite description to pick the discourse referent \( x \), which is equated to, but different from, the initial discourse referent \( z \). As a result, the whole utterance ends up ascribing the property of being a Buddhist to \( x \), and hence, even when an identity statement introduces a new equation \( y = z \), Buddhism does not re-apply to \( y \). This simulates our impression that Buddhism stays with Jones no matter who turns out to be the real murderer.

4. Local Accommodation of Replacer Functions

The referential use of “Smith’s murderer” in our scenario is a “standard” case in which the speaker and the listeners agree (in common ground) on who is Smith’s murderer. On the other hand, Donnellan (1966) also points out cases in which a referential use is rather successful without any such consent among the interlocutors. Here is one of such cases in the words of Kripke:

Someone sees a woman with a man. Taking the man to be her husband, and observing his attitude towards her, he says, “Her husband is kind to her”.... Suppose the man in question is not her husband. Suppose he is her lover, to whom she has been driven precisely by her husband’s cruelty. (Kripke 1977: 250)

In this example, while the speaker assumes the man dating with the woman to be her husband, the interpreters are supposed to know that the man is not her husband. Donnellan’s point is that, even under this circumstance, we can interpret the speaker to speak about the woman’s lover, rather than her real husband. Thus, assuming the lover is in fact kind to her, we judge the speaker to have said something true.

Kripke (1977) agrees with Donnellan, but with reservation. He emphasizes that our judgment on the utterance in question is rather mixed. If we are not under the misimpression that the man the speaker referred to was her husband, Kripke claims, “we would not express the same assertion by ‘Her husband is kind to her’” (254), and “it seems hard for us to say that when he uttered, ‘Her husband is kind to her,’ it expressed a truth” (255). In Neale’s words, when evaluating such an utterance, we “feel an uneasy tension” in that, although the speaker said something true about the person he wanted to talk about, the description he used failed to fit the person he wanted to talk about, and “to that extent the speech act was defective” (Neale 1991: 91).

This consideration led Kripke and Neale to appeal to the Grician distinction of sentence meaning and speaker meaning, and to claim that the utterance in question is true with respect to speaker meaning, but not with respect to sentence meaning. On the other hand, I claim that our conflicting judgments about the utterance can be accurately modeled by the difference of the replacer functions on which we interpret the definite description in question. In that
regard, our model is a simulation of what Kripke and Neale understand as the speaker and the sentence meaning in the present case.

In the case in question, the evaluators are supposed to have the information that the person dating with the woman is not her husband, that he is in fact kind to her, and that her real husband is not kind to her. So, if $x$, $y$, and $z$ are the discourse referents for the lover, the woman, and the husband respectively, we are in the following information state:

\[
\begin{array}{c}
x \ y \ z \\
\neg \text{husband}_o((x, y)) \ \text{kind}_o((x, y)) \ \text{husband}_o((z, y)) \ \neg \text{kind}_o((z, y)) \\
f_4 : x \to x \\
y \to y \\
z \to z \\
K_{12}
\end{array}
\]

If we judge the utterance “Her husband is kind to her” on this DRS, then the judgment will be negative. To wit, try to update $K_{12}$ with this utterance, assuming that both occurrences of “her” in the utterance picks $y$ as its discourse referent. Then, the whole NP “Her husband” will initially pick the discourse referent $z$, applies the current replacer function $f_4$ to $z$, and chooses the value $z$ as its final referent. Accordingly, the whole utterance adds the condition $\text{kind}_o((z, y))$, which, together with the condition $\neg \text{kind}_o((z, y))$ in $K_{12}$, would give rise to a contradictory DRS. $K_{12}$ thus injustifies the utterance. This simulates the negative side of our conflicting judgment on the utterance.

On the other hand, it is part of Donnellan-Kripke’s assumption that the evaluators also know that the speaker is “taking the man to be husband” and “saying of him” that he is kind to her. Intuitively, this means that the evaluators are in the position of accommodating the perspective that the speaker is taking, and interpret his utterance on that accommodated perspective. In our model, the perspective accommodation is a temporal switch of the replacer function. Thus, instead of evaluating the utterance on $K_{12}$, we could evaluate it on $K_{13}$, in which the speaker’s perspective is accommodated:

---

6 In the following, “$\neg \text{husband}_o((x, y))$” and “$\neg \text{kind}_o((z, y))$” abbreviate “$\neg \text{husband}_o((x, y))$” and “$\neg \text{kind}_o((z, y))$”, which are commonly used to make it explicit that a negative condition subordinates a DRS in it.

7 In the original definition of Lewis (1979), “accommodation” is an adjustment of common ground (an adjustment of DRS in our case) that takes place only when it is required for the utterance to be true or otherwise acceptable. Heim (1983b) uses the term to indicate an addition of presuppositional items required for the felicity of an utterance. Thus, perspective accommodation in our case is accommodation only in Lewis’s broader sense, not in Heim’s sense. See the next section for more detailed description of how perspective accommodation takes place.
Since $f_5$ assign $x$ to $z$ on this DRS, the NP “Her husband” in the utterance “Her husband is kind to her” ends up picking the discourse referent $x$, rather than $y$, as its final referent. Accordingly, the whole utterance prompts to add the condition $\text{kind_to}(x, y)$, which is already contained in $K_{13}$. Thus, the result of updating $K_{13}$ with that utterance is (vacuously) a logical consequence of $K_{13}$, and $K_{13}$ justifies the utterance. This simulates the positive side of our judgment.

Notice, however, that the accommodated replacer $f_5$ is “unwarranted” in $K_{13}$ in that $K_{13}$ lacks the equation $x = z$ or $z = x$. Thus, the accommodation of $f_5$ is bound to be “local” in the sense that the accommodation is only for the interpretation of a particular utterance, and has to be canceled as soon as the interpretation is completed.\footnote{See Heim (1983b) for fuller definitions of the terms “local accommodation” and “global accommodation.”}

The fact that the positive evaluation of the utterance requires a local accommodation of an unwarranted replacer function gives us the impression that the utterance does not “express a truth” in its own right. This fact also explains why “we would not express the same assertion” by the same sentence: we cannot generally expect the listeners to have an access to the particular perspective we are taking, and to locally accommodate our unwarranted perspective to evaluate our utterance favorably.

5. What Affects the Choice of Replacer Functions?

Our main concern in this paper is to develop a model that addresses the question of (i) what it is to interpret (or use) a definite description attributively and to interpret it referentially, and how the two readings gives rise to the observed inferential differences. Our formal model, on the other hand, is silent about (ii) under what conditions one reading arises more naturally than the other. Since we model the referential-attributive distinction as the difference of replacer functions on which we interpret a definite description, (ii) amounts to the question what factors affect the choice of the replacer function in each occasion of interpretation. For further elucidation of the phenomena we have considered, and to hint at a fuller model that can address the question (ii), I propose four constraints on the choice of replacer function.

**Constraint 1** The interlocutors chooses only a replacer function that is warranted in the current DRT.
This is just a restatement of Stipulation 2. In many cases, where the current DRS contains no equations, this constraint is strong enough to identify a unique replacer function. A case in point is the utterance (1) “Smith’s murderer is insane” in our original scenario. Since the set of conditions of the DRS $K_7$ contains no equations, the only replacer function warranted in $K_7$ is an identity function $f_0$ on $\{x, y, z\}$. This explains why only the attributive reading is likely for “Smith’s murderer” in (1).

The next two constraints are instances of the rule of accommodation in Lewis (1979).

**Constraint 2** If at time $t$ an utterance is made that requires the current replacer function to assign a discourse referent $x$ to a discourse referent $y$ if the utterance is to be topically suited; and if the replacer function does not assign $x$ to $y$ just before $t$; then at $t$, the replacer function assigns $x$ to $y$.

A case in point is Holmes’ third utterance (3) “Smith’s murderer is a Buddhist” in our original scenario. Because of the equation $x = z$, we have four alternative replacer function to choose:

$$
\begin{align*}
\text{f}_0 &: x \rightarrow x, y \rightarrow y, z \rightarrow z \\
\text{f}_1 &: x \rightarrow x, y \rightarrow y, z \rightarrow x \\
\text{f}_2 &: x \rightarrow z, y \rightarrow y, z \rightarrow z \\
\text{f}_3 &: x \rightarrow z, y \rightarrow y, z \rightarrow x 
\end{align*}
$$

Suppose the interlocutors chooses $f_0$ or $f_2$ to interpret (3). Then, “Smith’s murderer” in (3) picks $z$ as its referent and the whole utterance introduces the condition $\text{Buddhist}(z)$, where $z$ is the initial discourse referent for “Smith’s murderer.” Thus, (3) becomes an utterance about whoever murdered Smith. In view of the fact that (3) is a report of Holmes’ finding in an interview with Jones, this choice would make (3) topically unsuited. On the other hand, interpreting (3) on the replacer function $f_1$ or $f_3$ would make (3) topically suited. (3) would add the condition $\text{Buddhist}(x)$, and become an assertion about Jones. Thus, by Constraint 2, the interlocutors chooses $f_1$ or $f_3$. This explains why the referential reading is likely in this case.9

For another example, change our original scenario slightly, and suppose that Holmes utters (3) not after an interview with Jones, but at the site of the murder, while inspecting Smith’s body. In this case, the choice of $f_0$ or $f_2$ would make the utterance topically suited in that (3) would become an assertion about whoever murdered Smith. On the other hand, the choice of $f_1$ or $f_3$ would make (3) an assertion about Jones, who is hardly a topic in the discourse. The interlocutors thus chooses $f_0$ or $f_2$, interpreting “Smith’s murderer” attributively.

---

9 The notion of topical suitedness is admittedly vague. I am simply assuming it possible to define the notion structurally, in terms of the DRS in which the utterance is embedded.
This reflects the fact that the attributive reading is possible even when the interlocutors have already identified the murderer\textsuperscript{10}.

Still another example is the utterance “Her husband is kind to her” discussed in the previous section. If the interlocutors chose a function \((f4)\) that assigns \(z\) to \(z\), then the utterance would become an assertion about the woman’s real husband, who is hardly the topic in the discourse. In contrast, the choice of a function \((f5)\) that assigns \(x\) to \(z\) makes the utterance an assertion about the man dating with the woman, who is a likely topic in the discourse. This motivates the interlocutors to select \(f5\) as the replacer function.

In this particular case, however, there occurs a clash between Constraint 1 and Constraint 2: while Constraint 2 motivates the adoption of a replacer function that assigns \(z\) to \(x\), Constraint 1 discourages it since such a function is unwarranted in the given DRS. This is exactly the way the interlocutors come to have two conflicting judgments upon the utterance: when we follow Constraint 1 and interpret the utterance on a replacer function that assigns \(z\) to \(z\), our judgment on the utterance becomes negative, while when we follow Constraint 2 and accommodate a function that assigns \(x\) to \(z\), our judgment becomes positive.

**Constraint 3** If at time \(t\) an utterance is made that requires the current replacer function to assign a discourse referent \(x\) to a discourse referent \(y\) if the utterance is to be evaluated true; and if the replacer function does not assign \(x\) to \(y\) just before \(t\); then at \(t\), the replacer function assigns \(x\) to \(y\).

For a case in point, slightly change the example in the previous section, and suppose the speaker says, “Her husband is unkind to her,” instead of saying “Her husband is kind to her.” If all the other conditions to be the same, how do we interpret the utterance? Although we are somewhat inclined to take the speaker to talk about the lover, the inclination is much stronger to take him to speak about the real husband. Where does this difference come from? On the one hand, since the topical context still indicates the lover to be the topic, Constraint 2 motivates us to accommodate a function, \(f5\), that assigns \(x\) to \(z\), and to interpret the definite description to pick \(x\), the discourse referent for the woman’s lover. On the other hand, as before, we have a reverse Constraint 1 that discourages the accommodation of \(f5\) because of its unwarrantedness. In this case, however, Constraint 3 plays a more important role. Observe that the choice of a function, \(f4\), that assigns \(z\) to \(z\) allows us to evaluate the utterance to be true, while the choice of \(f5\) forces us to evaluate the utterance to be false. To wit, if the utterance in question is interpreted on the function \(f4\), it would introduce the condition \(\neg\text{kind}_t(z, y)\), which is already in the current DRS. In contrast, the same utterance, if interpreted on the function \(f5\), would introduce

\textsuperscript{10}See Donnellan (1966). Ishikawa (To Appear) correctly points out that my earlier model in Shimojima (1993) fails to simulate the cases in question. I owe this observation also to William Ladusaw and Hans Kamp.
the condition \( \neg \text{kind_to}(x, z) \), which contradicts with the condition \( \text{kind}(x) \) already in the DRS. Thus, Constraint 3 strongly motivates the employment of the function \( f4 \), and together with Constraint 1, discourages the accommodation of the function \( f5 \). This is a case in which all the three constraints affect the choice of replacer function, and the power balance among them produces our strong preference of the “husband” reading over the “lover” reading.

**Constraint 4** *In the absence of other constraints, the interlocutors preserve an assignment of the previous replacer function.*

In many cases, constraints 1-3 are not strong enough to identify a single replacer function down to one. In the case of the utterance (3) in our original scenario, we are not constrained to choose either of \( f1 \) or \( f3 \). In the modified case of the same utterance, we still have a choice between \( f0 \) and \( f2 \). Constraint 4 dictates that, in such a case, the interlocutors choose the replacer function that maximally preserves the assignments of the previous replacer function.

**Conclusion**

It is important to notice that, in all the examples we have discussed, definite descriptions keep doing a single task: They pick the discourse referents that satisfy their descriptive contents in the given DRSs, apply the current replacer functions, and pick the obtained values as their final discourse referents. In particular, we did not have to specify separate instructions to a definite description to differentiate its attributive-referential potentials. Thus, in the proposed context change model, a definite description is not semantically ambiguous, not at least between the attributive and the referential reading. Our model distinguishes the attributive reading from the referential reading by the difference of the replacer functions on which the interlocutors interpret a definite description.

What is a replacer function then? What is it the model for, if it is ever a model for an entity in the real world?\(^{11}\) Here is one interpretation of the proposed model. As we have seen, our model fully exploits a feature of discourse referents that two (or more) discourse referents can be distinct while coinciding in reference. This makes a discourse referent, as used in our model, an apt model for a “guise” of an individual, a guise that is constructed in the course of a discourse. Suppose two discourse referents \( x \) and \( y \) coincide in reference, namely, are equated in an equation in the current DRS. When we accumulate a condition upon \( x \) in the course of interlocution, we are enriching one guise of an individual. When we remove a condition on \( y \), we are impoverishing the other guise of the same individual. From this standpoint, a discourse is a continuous process of updating the various guises of individuals. Now, a replacer function is intended

\(^{11}\) A particular object we use in a model need not itself be a model for a real-world entity. A “compositional model,” in which each model-theoretic object is a model for some real-world entity, is only one of the many ways of modeling phenomena.
to be an mechanism that directs which discourse referent a definite NP should pick when two (or more) discourse referents are equated in a DRS. Hence, a shift of replacer functions amounts to a shift of the guises to which the interlocutors refer by definite NPs. This feature makes a replacer function a model for a referential perspective of the interlocutors, a point of view from which they use definite NPs. Under this interpretation, our main claim is that the attributive-referential distinction is the difference of the guises a definite description selects in each occasion. For example, when we use “Smith’s murderer” attributively, the description selects the guise of an individual as Smith’s murderer, while when we use the same description, it selects the guise of the same individual as the person that Holmes has just interviewed (in our scenario), the person put in trial (in Donnellan’s original example), etc. Our model is then a proof that this intuitive “guise” theory in fact captures the crucial differences in inference pattern arising from Donnellan’s distinction.

References


