REFLEXIVITY IN FIRST-PERSON REFERENCE

KELLY ALBERTS
University of Wyoming

As an alternative to intentionalist accounts Direct Reference approaches to designation have achieved several successes, notably in the area of indexical reference. Arguably, however, these approaches have failed to produce a refined treatment of first-person, singular reference as it operates in natural language. In this paper I want to add to the growing corpus of work on Direct Reference Theory (DRT) and propose a way of treating the first-person nonintentionally.

I

For strategic reasons we begin by briefly presenting the intuitions and arguments that motivate DR approaches to reference. Chisholm advises that we distinguish between theories which emphasize the "primacy of the intentional" and those which emphasize the "primacy of the linguistic."[1] As I interpret it, DRT, at least of the variety for which David Kaplan is primarily responsible, belongs to this second classification.

If we ask what is essential to any intentionalist theory of reference (IRT), we find a general assumption: that thought has a "directedness" to it. After Castaneda, we might call this basic datum "thinking reference."[2] There are, however, strong and weak versions of IRT. The strong version of the theory claims that thinking reference is somehow fundamental to all other forms of reference. As Castaneda expresses it, "*the semantics of a means of thinking*...is a climactic part of the

---

[1]Roderick Chisholm, *The First Person: A Study in Intentionality and Reference* (Minneapolis, Minn.: Univ. of Minnesota Press, 1981), p. 1.
[2]Actually Castaneda distinguishes between "speaker's thinking reference" and "hearer's thinking reference" in his presentation of a five-fold classification of reference. Nothing is lost, I believe, by ignoring the Gricean overtones in Castaneda's classificatory scheme. See, Hector-Neri Castaneda, "Direct Reference, Realism, and Guise Theory," (Unpublished, Indiana Univ., 1984). p. 3.

phenomenon of reference"[3] (emphasis his). Reference that takes place by means of language must, again as Castaneda sees it, "eventually be tested against speaker's thinking reference."[4] It is a theory such as this that Chisholm has in mind when he speaks of the primacy of the intentional. On the other hand, the weak version of IRT does not insist upon primacy. It claims only a place for thinking reference.

By comparison, the fundamental idea of DRT in its application to indexical expression is that the mechanism of reference operates independently of thought. Suppose I say, "That doesn't belong here." On the version of this theory proposed by Kaplan,[5] the indexical terms 'that' and 'here' in this sentence possess both a "character" and a "content." The character of an indexical is the implicitly understood linguistic rule(s) governing the use of the word.[6] The content is the referent of the term. Therefore, understanding the meaning-rule or character under which "that" and 'here' operate, I express the above sentence in a given context, and in that context certain references (say, the ordered pair [dirty shirt, bathroom floor]) are automatically assigned to the indexical expressions at issue. Hence,

KT.1. Reference is a function from particular, actual or possible contexts to individuals.

The proposition expressed by the sentence in question is Russellian in nature. Hence,

---

[3]Castaneda, p. 3.

[4]Castaneda, p. 3.

[5]David Kaplan, *Demonstratives: An Essay on the Semantics, Logic, Metaphysics, and Epistemology of Demonstratives and Other Indexicals* (Unpublished, UCLA, 1977).

[6]There is a difficulty here. The idea of a rule "in the sense of character" is ambiguous in Kaplan's theory. It has application both inside and outside of language. From the inside, as a user of a given indexical term, the character of that indexical allows one to fix the reference of it on a given occasion of use. From the outside, as a theoretician of language, the rule functions more as an analysis of whatever indexical expression is at issue. In its application as an analysis a rule may contain concepts which make it unrecognizable to a user. In this paper we are interested in the analysis-rule.

KT.2. Propositions consist, *inter alia*, of the contents of the indexicals used in the sentences that express them.

If DRT is to serve as a challenge to intentionalist approaches to reference, however, these two theses are not sufficient. We need an additional claim. Consider the case of the ignorant heiress. Kidnapped, locked in the trunk of a car, and not knowing where she is or what time it is, she nevertheless utters, "Well, here I am now." What singular proposition is expressed by this sentence? According to Kaplan, the actual place and time of the heiress' entombment function as the contents of the expressed proposition even though the heiress herself is ignorant of these referents. Hence,

KT.3. Ignorance of the referents does not defeat the referential character of indexical expressions.

For Kaplan propositions must be detachable from the actual or possible contexts that produced them. It is only in this way that they may be evaluated from the perspective of different circumstances (i.e., possible worlds). It is this requirement that justifies KT.3.

KT.3 has the anti-intentionalist effect of driving a wedge between the semantic proposition and what's going on in the heiress' head. It is for precisely this reason that Castaneda does not find DRT acceptable. The mechanism of reference it offers allegedly leaves out something essential (if one holds a strong version of IRT) or something important (of one holds a weak version). On the standard view propositions are the bearers of truth-value. According to Castaneda, while Kaplan's Russellian proposition may be what is truth-valued from a purely external point of view, it omits the "truth" that is experienced by the heiress. He comments,

> The truth she [the heiress] proclaims seems indeed to lie right there in the middle of her experience; it is the truth that there is quiet at the time of her experience as she experiences it and at the place where she experiences herself to be. She seems to be asserting a purely experiential and experienced truth, which

must, therefore, by thoroughly differentiated from the Russellian-Kaplanian proposition...[7]

Castaneda believes that we must admit of an "internal accusative of thought" to function as the truth that the heiress experiences. It is this internal accusative, omitted in DRT, that constitutes a structure (or the basic structure) of thinking-reference for intentionalist theory.

If we are to apply consistently the insights of DRT to I-reference, the problem Castaneda raises demands an answer. The correct reply, I think, can be expressed as a dilemma. Either the internal accusative to which he refers is truth-valued or it is not. If this structure is truth-valued, then it turns out to be irrelevant to semantics. If it is not truth-valued, then it is irrelevant to semantics for another reason. By either horn, DRT is warranted in neglecting the internal accusative, if there is such a thing.

Suppose we assume that the accusative is a bearer of truth-value. In what does its structure consist? If the heiress were not suffering from a condition of complete ignorance, we might picture this internal accusative as a structured entity containing certain mental representations of what, *inter alia*, 'here' and 'now' single out. The referent of 'here' would be a mental representation of the actual place of the utterance, and the referent of 'now' would be a representation of the actual time of the utterance.

This picture is no help to Castaneda, however. These representations were they to exist, would simply be modeled on--that is to say, their analysis would be dependent on--the primary referents contained in the K-R proposition. The referent of 'here' would presumably be a phenomenological representation of the car trunk and its spatial context, and the referent of 'now' would be whatever phenomenological representation the heiress could have of the actual time. The point, though, is that the internal accusative of thought would merely be a shadowy, mentalese imitation of the external proposition. There is no reason for a specifically semantic

[7]Castaneda, p. 21.

analysis to elevate this structure to relevance. Thought would be "directed" in virtue of its function as internal speech.[8]

But of course the heiress is in no position to display *these* representations to her consciousness. What, then, could function as the content of the mentalese counterparts to 'here' and 'now'? What Castaneda says is, "there is quiet at the time of her experience as she experiences it and at the place where she experiences herself to be." This is by no means clear, but two interpretations suggest themselves.

According to the first interpretation we simply abandon the claim that the internal accusative or "truth" is singular in structure. Perhaps the general proposition "There is an x and there is a y such that x is an experience of place for S and y is an experience of time for S and x and y have property F," more accurately represents the correct construction of the internal accusative. If this is the case, however, the burden of proof would seem to be entirely on Castaneda. For even if we assume this construction represents the general features of the phenomenological experience the heiress is undergoing, we are still left with the task of explaining what that construction has to do with the terms 'here' and 'now' as employed in the sentence, "Well, here I am now." Those terms do not appear to be operating semantically in the way the interpretation requires, i.e., as propositional functions. Instead they operate as singular designators. In this case and in the more general case the proper semantic treatment of sentences containing indexicals would seem to demand existential commitment to singular propositions.

The second interpretation involves construing the internal accusative as a singular truth but taking 'here' and 'now' to be operating as definite descriptions for something on the order of "the quiet and dark place..." and "the eerie time..." respectively. On no construction, though, does this solution work.

If a Russellian interpretation is placed upon these definite descriptions, singular propositions are surreptitiously abandoned in favor of general ones, and we are back to the

---

[8]*cf.*, W. Sellars, "Language as Thought and as Communication," *Philosophy and Phenomenological Research* 29 (1969): 506-527; and Christopher Gauker, "Thought as Inner Speech" (Unpublished, University of Cincinnati, 1987).

preceding solution. If the definite descriptions are interpreted along Donnellan's lines, then they are in fact not definite enough. They can accomplish neither referential nor attributive reference.[9] 'Here' and 'now' cannot latch on to something referentially because there is no antecedent something in the mind of the heiress which these terms are to aid us (or her) in singling out. Furthermore, 'here' and 'now' fail to refer attributively--i.e., to refer to whatever happens to satisfy their disguised description--because there is no reason to assume that these descriptions or their mental representations are sufficiently definite to admit of unique satisfaction. Finally, if these descriptions are interpreted as elliptical for "the quiet and dark place where I am" and "the eerie time that it is," unique satisfaction is ensured only at the expense of affirming Kaplan's general point. For if the expressions "the place where I am" and "the time that it is" succeed in referring to the actual place and time of the heiress' captivity, they do so even though she is entirely ignorant of them; even though, as it might be put, she is not *de re en rapport* with them.

We are thus led to the second disjunct of our original dilemma. I assume most proponents of DRT would find this the most plausible alternative. Whatever else is said of the internal accusative, it would seem not to be truth-valued. If we could project the stream of consciousness that characterizes the heiress' mind as she contemplates her problem and utters "Well, here I am now," there is no reason to suppose this stream possesses propositional structure. The heiress is, as Castaneda says, "experiencing" quiet (and darkness) at the time and place of her cramped captivity. But this clearly does not warrant the thesis that "she seems to be asserting a purely experiential and experienced truth." What she is asserting is the K-R proposition; what she is experiencing is no truth at all.

Therefore, the internal accusative insofar as it is an object of experience rather than a bearer of truth, would seem to be irrelevant to the project of semantic analysis. Alternatively, one might say that if the heiress' expression marks a "truth" which functions as an experiential accompaniment to the

---

[9]Keith Donnellan, "Reference and Definite Descriptions," *The Philosophical Review* 75 (1966): 281-304.

semantic proposition, this sense of "truth" has no referential cash value. On this horn, in other words, the idea that thought has "direction" independent of language is simply not accepted.

II

Although this answer does not and could not serve as a comprehensive response to a matter this deep, it is sufficient to block the checkmate that intentionalist intuitions may seem to put in the way of a DR account of 'I'. However, the puzzle we are left with is this. I have number of devices at my disposal for referring to myself. Consider just two of these--'I' and my proper name "KA." If what we said in the preceding section is correct, then I-propositions share in part an identical content with singular propositions that get produced by means of my proper name. A certain person, me, is the subject constituent of any token of this type. It might seem to follow, therefore, that the first-person pronoun 'I' and the proper name 'KA' demand the same semantic treatment. So how can DRT account for I-reference without assimilating it to a proper name?[10]

On the standard DR view of proper names 'KA' refers to the person who bears the correct causal or historical connection to the person who was baptized 'KA' in a certain time, at a certain place, in the actual world. In their recent book *Knowing Who*, Boer and Lycan mistakenly assume that DRT will treat 'I' as a proper name too. They write, "Note also that for some singular terms, such as indexical pronouns like 'you' and 'I', the 'appropriately shaped causal chains' are so short and direct as to be degenerate cases."[11] But with 'I' there are no causal chains, appropriately shaped or otherwise. To think of a use of 'I' as a degenerate case of tagging something with a name is to endorse an entirely misleading picture of both the Causal

---

[10]Note that IRT, in either its strong or weak form, is not faced with this puzzle. For the referent of 'I' *qua* internal accusative will not be identified with the referent of 'KA'*qua* external proposition. The idea that the internal 'I' refers to a metaphysical self of some description is a normal corollary to both versions of IRT.

[11]Stephen E. Boer and William G. Lycan, *Knowing Who* (Cambridge, Mass.: MIT Press, 1986), p. 128.

Theory of Reference (one variation on DRT) and first-person reference by means of 'I' and 'me.'

This is not to say that 'I' (or even 'me') cannot be used *as* a proper name. It can even be used *as* a definite description. Compare two cases. In the first case you wear a moustache. One day while shaving you notice splotchy bald spots, *alopecia areata*, in your moustache. You utter, "I am losing my moustache." Contrast this with a second case. You are an extremely famous scientist, well known for your public stands on controversial social problems. You are watching a play, and the play is about you. Midway through the play you are amused to notice that the false moustache on the actor portraying you is beginning to slip. You turn to your companion and utter, "Look. I am losing my moustache."[12]

The same sentence is uttered in both of these cases and that sentence contains a use of 'I' in the subject position. However, the first example discloses a "first-person" use of 'I' whereas the latter example does not. Why is this? The answer is that the indirect discourse proxy for the first-person use of 'I' would necessarily make use of the indirect reflexive locution "he himself." By contrast, the non-first-person use of 'I' would not and could not utilize this indirect reflexive.

Consider. In the first case it is appropriate to say of you, "He knows that he himself is losing his moustache." The reflexive pronoun 'himself' is essential to this transformation because it grammatically represents the fact that in the first case your utterance of 'I' is directed upon yourself *qua* performer of the utterance. In the second case the indirect use of the reflexive is entirely inappropriate. In this case we would not be tempted to say "He knows that he himself is losing his moustache." We know from the description of the case that you are in no danger of mistaking the actor for yourself. Hence we would use a nonreflexive, singular term for the indirect discourse transform, (e.g.) "He knows that the actor portraying him is losing his moustache."

This suggests the following criterion. An employment of 'I' in a sentence of direct discourse is relevant to the semantic analysis of first-person singular reference if and only if the

---

[12]Those familiar with Castaneda's "'He': A Study in the Logic of Self-Consciousness," *Ratio* 9 (1966): 130-157; will recognize this case as the one originally suggested by Norman Kretzmann in footnote #12.

transformation of that sentence into indirect discourse necessarily involves (after the relevant that-clause) the use of a grammatically indirect reflexive expression.[13] This criterion filters out those conceivable uses of 'I' that have very little to do with the study of first-person semantics. If a parent were to name her child 'I', we would feel sorry for the child but would not confuse this idiosyncratic employment with the use of the pronoun 'I' in natural language. I think we should assume a similar attitude toward the use of 'I' as a definite description in the second case above. Clearly it is a possible use, but this does not imply that it will prove relevant to the analysis of first-person, singular reference.

To return, then, to our puzzle: we want to put me in the K-R proposition whether I say first-personally, "I intend to write this paper," or whether I say "KA intends to write this paper," (or even "This man intends to write this paper"). But importantly we want the referential mechanism by means of which this is accomplished to differ in the two cases. The difference is to be sought in the divergent character of these two kinds of expression. We found the "he himself" locution to be essential to the formulation of a criterion for first-person uses of 'I'. Grammatically, as noted, the term 'himself' is a reflexive pronoun. It is the concept of reflexivity, I believe, only now applied to semantics, that can help us to see how a first-person use of 'I' differs in character from other devices of singular reference.

III

The idea of reflexivity analytically entails the notion of "turning back upon." Consider another reflexive expression, "this very phrase." Semantically we might, with Nozick,

---

[13]Castaneda is primarily responsible for bringing the attention of the American philosophical community to the importance of the indirect reflexive in the analysis of 'I.' But in attempting to show that 'I'has an irreducible sense he did not think to use the "he*" device as a way of distinguishing between first-person and non-first-person uses of 'I.' For that , Geach's original note is more helpful. My use of the term "proxy" comes from him. See, P.T. Geach, "On Beliefs About Oneself," *Analysis* 18 (1957): 23; H-N Castaneda, "'He': A Study in the Logic of Self-consciousness," *Ratio* 8 (1966): 130-157.

describe this expression as a case of reflexive "self-reference."[14] What the phrase refers to of course is itself; on each occasion that "this very phrase" is tokened or replicated it refers to the phrase so tokened or replicated. It is important to note that to say a term self-refers in this sense is just to mean, neutrally, that it refers to itself. Reflexivity of this sort--i.e., that which "turns back upon itself" in the tightest possible way--may be termed reflexivity of the "narrow sort." In the present example this reflexivity is provided by the introduction of the term "very" into the phrase "this phrase." Although one might use "this phrase" to single out itself, one can use it to refer to the same phrase on different occasions of use. The addition of "very" makes the phrase sensitive to its own tokening.

It would be a mistake to suppose that 'I' reflexively self-refers in the same sense. Plainly 'I' does not refer to itself. Thus if self-reference is predicated of 'I', the sense of this concept cannot be explained in the way that we explained narrow reflexivity.

What sense, then, is to be given to the idea of self-reference in this application? While the reference of 'I' is not itself, its referent is causally connected to itself. What 'I' refers to on each occasion of its replication or tokening is the *producer* of that token. Since the reflexivity involved here is once removed from an explanation solely in terms of the reflexive pronoun "itself", we may term it reflexivity of the "broad sort." (Note that we may remain noncommittal about whether 'I' refers to a "self," in the sense of a nonempirical agent who may somehow take itself as an object of its own reflection. A proponent of DRT will have no special inclination in virtue of his theory to endorse such an entity.)

Initially one might believe that the description of broad reflexivity could be piggybacked upon the description of the reflexivity of "very," as in "this very phrase." On a view once held by Riechenbach, 'I' just means "the person who utters this

---

[14]See, Robert Nozick, *Philosophical Explanations* (Cambridge, Mass.: Harvard University Press, 1985), p. 74. Nozick, too, makes use of the concept of reflexivity, but the analysis he offers differs in fundamental ways from the one being proposed in this paper.

token";[15] and it is clear that he intended to say "the person who utters this very token." A refined version of this, incorporating the point made above, might be: 'I' just means "the producer of this very token." One objection to this meaning-postulate is that it would make inexplicable the difference between the analytic and necessary statement:

> The producer of this very token is the producer of this very token,

and the synthetic and contingent statement:

> I am the producer of this very token.

Another difficulty is this. The reflexivity of "very" in "this very phrase" is a function of the reflexive used in its explanation, i.e., "refers to *itself*." If the explanation by means of the reflexive is rendered inapplicable--as it is by the fact that 'I' does not refer to itself--then some other explanation of reflexivity of the broad sort must be produced.

The advent of DRT shows why meaning-postulates of the type suggested by Riechenbach cannot function in a semantic analysis in the way that was once supposed.[16] The expression "the producer of this very token" is not a synonym of 'I'. By KT. 1, this expression acts as a character which, together with a context, fixes the reference of 'I'.

The idea of using a character to fix a reference rather than to supply a synonym is borrowed by Kaplan from Kripke.[17] What sets Kaplan's application of this idea apart from Kripke's, however, is that Kripke is concerned with the use of an accidental property of a thing (e.g. being the length of the

---

[15]H. Reichenbach, *Elements of Symbolic Logic* (New York, N.Y.: Macmillan, Inc., 1947), p. 284. See also, Kaplan, *Demonstratives*, pp. 43-44. And Hector-Neri Castaneda, "Indicators and Quasi-Indicators," *American Philosophical Quarterly* 4 (1967): 87.

[16]Kaplan, *Demonstratives*, pp. 10-27. See, also, Kaplan, "On the Logic of Demonstratives," in French, Uehling, and Wettstein (eds.) *Contemporary Perspectives in the Philosophy of Language* (Minneapolis, Minn.: Univ. of Minnesota Press, 1979), pp. 401-412.

[17]See, Saul Kripke, *Naming and Necessity* (Cambridge, Mass.: Harvard University Press, 1981), pp. 54-56.

standard meter stick in Paris) to fix the reference of a singular term naming that thing (e.g., "one meter"). While rules of the type envisioned by Kaplan have nothing to do with accidental properties, if the idea can be extended in the way proposed, Kaplan argues that indexical expressions fall into either one of two categories.

These categories are defined by the manner in which the expressions subsumed get their respective reference fixed. Demonstratives, terms such as 'this,' 'that,' 'he' or 'she' require two elements to accomplish reference. They need, first, some appropriate character--e.g., "the reference of 'she' must be of the female gender." They need, second, an appropriate demonstration of the exact thing which meets the condition(s) specified by the character--e.g., "She (pointing to the person with the book in her hand) is a fine lecturer." Pure indexicals, on the other hand, terms such as 'today,' 'tomorrow' or 'yesterday' require only an appropriate character--e.g., "'today' refers to the day in progress." In their case no associated demonstration is required.

Where does 'I' fit? According to Kaplan, it is not a demonstrative; for, aside from emphasis, any demonstration of oneself accompanying an utterance of 'I' is redundant. It therefore is to be classified with the pure indexicals. On his view the rule in the sense of character under which 'I' operates is formulated by what Perry and Castaneda have called the K-rule: In each of its utterances, 'I' refers to the speaker who utters it.[18]

The advantage offered by Kaplan's account is immediately clear. It provides us with a mechanism for distinguishing I-reference from other types of singular reference. As just noted, 'I' is not a demonstrative. Further, the manner in which the reference of a proper name is fixed--even one's own--is causal rather than indexical in character. Finally, definite descriptions, at least in their basic use, accomplish reference by means of satisfaction; in the expected situation one and only one person satisfies the description "the conqueror of the North Pole." Hence, on Kaplan's view we can register the semantic

---

[18]See, John Perry, "Castaneda on 'He' and 'I'" and Hector-Neri Castaneda, "Reply to John Perry," in *Agent, Language, and the Structure of the World*, James E. Tomberlin (ed.) (Indianapolis, In.: Hackett Publishing Co., Inc., 1983), pp. 15-42 and 313-328.

differences among reference by means of 'I', demonstratives, proper names, and definite descriptions.

What we cannot do, however, is register the difference between the semantics of I-reference and the semantics of reference by means of the pure indexicals 'today,' 'tomorrow' and 'yesterday.' What I which to argue is that there is a "demonstrative-like" aspect to I-reference. This aspect is entirely absent in the paradigm pure indexicals. For note that in contrast with the pure indexicals, a demonstration of oneself may take the place of a use of 'I'. This suggests the hypothesis that the broad reflexivity characterizing I-reference demands explanation of a rather unique sort.[19]

---

[19]We have not considered the indexicals 'here' and 'now,' and perhaps they demand a slight qualification to the above statement. Many have noted semantic similarities between these two terms and 'I.' Kaplan argues that 'here' and 'now' can be used either demonstratively or as pure indexicals. Consider for simplicity only 'here.' I believe we need to distinguish three different cases of this term's employment.

First, there is the case where one might say, "We were here," pointing to, say, Tahiti on a map. This would seem to be a demonstrative use of the term 'here.' Second, there is the odd case where it might be appropriate to say "We are here," taking here to be wherever we in fact are. This case, made much of by Kaplan, seems to me *not* to constitute a pure indexical use of 'here.' For, arguably, 'here' is operating simply as the definite description (interpreted attributively) "the place where we are." Regardless whether the case constitutes a pure indexical use or not, 'here' would seem to be accomplishing reference nonreflexively. Third and finally, one can imagine oneself lost, in a cloud, on a mountain. Someone yells "Where are you?" and one replies "I am *here*."

It may be better not to interpret this third example as a case of reference at all. "HERE!," one might contend, directs our attention to a certain location, but it does not single out that location. But if one does interpret it as a case of reference, it may perhaps seem to constitute a use of 'here' exhibiting broad reflexivity. It reflexively refers in this case, not to the producer of the very token but to the spatial location (defined arbitrarily) from which the voice producing the utterance "Here!" originates. Similarly, 'now' refers to the time (also defined arbitrarily) at which the voice producing the utterance 'now' occurs. So, only with this third case do we get something semantically similar to 'I.'

I wish to propose that the explanatory notion necessary to explain the special reflexivity of 'I' is that of surrogate demonstration. Think of a case where you raise your hand in response to a "Who?" question, (e.g.) "Who wants the last piece of cake?". The act of raising your hand takes the place of uttering an expression that singles you out. The demonstrative act is a surrogate for an act of singular reference. With 'I', it seems to me, the surrogate relation is reversed. The producer of 'I' is performing the symbolic surrogate of overtly pointing to him or herself. It is as if 'I' in language were an arrow, and the arrow always points at the agent who produced it.

If this idea were to prove analytically instrumental, we would have a refined DRT treatment of 'I'. This treatment would explain why an utterance of 'I', unlike the utterance of a demonstrative such as 'this', requires no additional demonstration. The demonstration has already been accomplished by the use of reflexive language. It would register, moreover, the demonstrative-like aspect of 'I' in contrast to the pure indexicals. Finally, it would explain why the narrow reflexivity implied by the self-reference of "this very phrase" will not stretch to an explanation of the broad reflexivity of 'I'. For this latter explanation we require the concept of a demonstration and its surrogate in language.

IV

But can the idea prove analytically instrumental? The notion that 'I' operates in language as a symbolic arrow pointing to its producer will be seen as naive in several respects. There are, in fact, three rather strong objections to the very idea of using surrogate demonstration to explain the reflexivity of 'I'. My own view is that each objection ultimately helps us change our hypothesis into a genuine analysis.

The semantic account now on the table, modeled upon the K-rule, is: In each of its first-person uses, 'I' refers to the producer of that very token. Let us call this the Revised K-rule. But suppose the wicked witch, saying nothing, thinks to herself, "I am the fairest of them all." Assume that this use of 'I' serves in thought to distinguish the witch from them all. Surely, it would be wrong to suggest, wouldn't it, that 'I' accomplishes this task in virtue of a surrogate of self-pointing?

This is the alleged counterexample that will be presented against us by the proponent of IRT. The person who holds the strong version of that theory will conclude that our DR account omits just what is primary--namely, the nondemonstrative directedness of the internal accusative. The person who holds the weak version will find our account insufficiently generalizable. For it cannot apply to one form of I-reference, namely I-thought reference.

Our earlier response to IRT indicates how we should answer this objection. If, as it seems to me most plausible to hold, the thought, *qua* internal accusative, is experiential but not propositional, then it lacks truth-value. If it lacks truth-value, then it is not relevant to semantics. If, on the other hand, it is assumed that truth-value is to be predicated of the witch's internal accusative, then the mental image she presents to herself of the content of 'I' is simply a representation of what she would have demonstrated in a surrogate way had she spoken 'I'. (Note that there is no problem about ignorance in this case.) The internal accusative in this counterfactual way will be modeled on the external proposition.

Consider an analogous case: the written instance of 'I'. Since no act of utterance occurs in this case, it may seem that our analysis is inapplicable. Now written uses of 'I' fall into one of two categories: the fictional or the autobiographical. When 'I' is employed by a character in a novel, it refers reflexively in virtue of a surrogate self-pointing to the fictional individual who produced that very token. Hence, the Revised K-rule, interpreted according to the referential mechanism we have accorded it, and qualified by a provision notifying the reader of the fictional character of the discourse, is an appropriate formulation of the semantic rule controlling the fictional 'I'.

The autobiographical use is more problematic. One could adopt a simple amendment to the Revised K-rule: In each of its first-person written instances, 'I' refers to the producer of that very written inscription. However, now the idea of explaining the notion of broad reflexivity in terms of surrogate self-demonstration begins to look less perspicuous. What we are not inclined to do in this situation is make up an entirely novel and distinct structure of reference. The more reasonable approach is to adopt a counterfactual application of our antecedent analysis:

Rev. K-rule (i): If a writer S uses 'I' as an element in a written sentence of autobiography A, then S has performed what would have been an act of surrogate demonstration had 'I' been used by S to communicate A verbally.

Of course the pragmatic circumstances of the case may make it impossible to know to whom 'I' refers in any given autobiographical use. The point to be underlined, though, is that even in a first-person use of 'I' where no utterance-act occurs, the idea of surrogate demonstration, applied counterfactually, is the correct explanatory notion.

The use of 'I' in written autobiography provides the model for integrating the I-thought into our DRT analysis:

Rev. K-rule (ii): If a speaker S uses 'I' as an element in an I-thought T, he has performed what would have been an act of surrogate demonstration had 'I' been used by S to communicate T verbally.

This amendment can be viewed as a transformation rule. It gives the mechanism for transforming the use of 'I' in an internal accusative into a use of 'I' suitable for producing a proposition of which semantics can deal. As with the case of the autobiographical use of 'I', we reject the intentionalist manuever in favor of a counterfactual application of the Revised K-rule. This reemphasizes the adherence of DRT to a theory that holds consistently to the primacy of the linguistic.

The explanatory device we have employed in our analysis will seem naive for a second reason. A demonstration would seem to be subject to mishaps that are inapplicable to a use of 'I'. Any given demonstration may fail to designate any object at all. It may fail to designate a unique object. And it may inadvertently designate an object other than the one intended by its performer. Are we to suppose that the semantic rule governing 'I' allows for similar possibilities?

This objection forces the recognition that the Revised K-rule is not sufficiently revised. The type of demonstration of which a first-person use of 'I' is an instance is, I believe, subject

to severe semantic restrictions. In this connection I wish to introduce the notion of a referential accident.[20]

To introduce the concept we require a distinction between pragmatics and semantics, and, within this framework, some Kripke-like distinction between the speaker's referent, on a given occasion of utterance, and the semantic referent of the linguistic expression used to make that utterance on that occasion.[21] Roughly this distinction (drawn from Grice) holds that there may be a difference between the object to which a speaker intends to refer by means of an expression, on an occasion of utterance, and the object to which his words actually refer in language. A referential accident thus occurs when a singular term is unable to successfully accomplish either the reference the semantic convention demands or the reference that the speaker of the term pragmatically intends.

Reference failure consists of course in the failure of an expression to designate any referent whatsoever. Some philosophers, influenced by Descartes, take 'I' to be immune to reference failure. The proposition expressed by the sentence "I do not exist" might be thought necessarily false. In my opinion one of the advances of contemporary possible world semantics is the insight that even first-person, existential statements are contingent. According to that semantics, the sentence "I do not exist" is possibly true. It is true "in," "of" or "at" all those worlds in which I do not exist. Hence our semantics of 'I' need not build in an immunity to reference failure.

The issue of referential ambiguity is more complex. Many context-relative indexicals, even sortally disambiguated ones such as "this phrase," fail (absent an accompanying demonstration) to designate a unique individual. If 'I' referred reflexively on the model of "this very phrase," then its reference could not fail to be unique. But 'I' always refers to its producer, and one might begin to wonder whether there might be, in some instances, more than one producer of 'I'. Imagine a case of speaking-through-the-mouth-of-another. The speaker, *qua* mouthpiece says, "I want you to come to me," but the person responsible for the sentence is the mad psychiatrist. Could 'I'

---

[20]This concept was first suggested to me by Rogers Albritton.

[21]Saul Kripke, "Speaker's Reference and Semantic Reference," in French, et. al. (eds.), *Midwest Studies in Philosophy*, 2 (Minneapolis, Minn: U. of Minnesota Press, 1977), pp. 258-259.

be convicted of referring to both the mouthpiece-producer and the psychiatrist-producer?

I would think not. If you were on the receiving end of this communication, you may not know whom to identify as the person wanting you. But this would seem to be a feature of the pragmatic context and not be salient to the semantics of 'I'. Depending on how the details of the story are filled in we may want to identify the psychiatrist as the referent of 'I' or we may want to identify the mouthpiece as the referent, but from this it does not follow that 'I' could refer to *both* producers. 'I' would seem not to be subject to that type of possible ambiguity.

Consider, finally, the issue of referential misfiring. I intend to refer to Smith by means of the expression "the man in the closet reading Shakespeare." But since it is Jones rather than Smith who satisfies this description, the reference has misfired. Referential misfiring can thus occur in cases where the speaker's referent (the object the speaker has in mind) may diverge from the semantic referent (the object to which the words in language actually refer). Is it possible for the speaker of 'I' to refer inadvertently to some producer other than the producer to whom he intends to refer?

Consider this case. Ernst Mach looks into the mirror and says, "I am a shabbily dressed academician." Unbeknowst to Mach the mirror is angled in such a way that he is in fact viewing the image of someone other than himself. That person is the shabbily dressed academician. Someone might claim, on the basis of such a description, that Mach intends to be making a statement about the person in the mirror. If so, the argument continues, the semantic reference of 'I' is Mach and the speaker's reference is the person reflected in the mirror. On the basis of such a possibility it might be claimed that 'I' is indeed subject to misfiring.

This conclusion betrays a confusion. We can agree that Mach's statement expresses a false belief. We would say of him, "He mistakenly believes that he himself is the shabbily dressed academician." We can agree, moreover, that if the statement were to be true, then it would have to be about the person reflected in the mirror. But this counterfactual formulation of the conditions under which Mach's I-statement is true does not introduce a speaker's referent possibly divergent from the semantic referent. Mach's use of 'I' signaled his intent to use that term in accord with the appropriate semantic rule,

and I am prepared to argue that this rule is sufficiently definite to rule out the possibility that he could intend that word to refer to someone other than himself.

The argument I have in mind is best seen as a response to an objection that may seem irresistible at this point. Surely, someone will say, the speaker of 'I', in this case Ernst Mach, must "intend" or "mean" to be referring to himself by use of that pronoun. Isn't that "intention" or "meaning," the objector will continue, the salient semantic feature--a feature that your analysis entirely omits? And isn't it possible for that "intended" reference of 'I' to diverge from what you call the "semantic" reference?

Here again we have a variation on the strong version of IRT resurfacing. The defender of a DR approach to the first-person will point out that the concept of an intention is ambiguous in this objection. We must distinguish between a general or first-order intention and a more specific or second-order intention. When a speaker utters (e.g.) "This...", he must have the first-order intention to employ that term in conformity to a certain rule ("'this' refers to things in one's near spatial and temporal range"), but this intention is not sufficient to fix a reference. The speaker must have the additional, second-order intention to refer to some given item among those things in that range.

The predication of reflexivity to I-reference amounts to the claim that this second-order sense of "intention" is inapplicable. Intention only comes in at the first-order level. That is to say, *if the speaker of 'I' has the first-order intention to use that term first-personally, then that speaker, qua producer of that very token, is singled out automatically.* In the case at hand, Mach indeed "means" or "intends" to be using'I' first-personally--not as a proper name or definite description. But given this, there is no further second-order intention controlling his first-person use of 'I'. The character governing 'I' assigns the reference to that term independently of what is going on in Mach's head.

It is for precisely this reason that with the word 'I' (in its first-person employment) there is only semantic reference. This is why no sense can be given to the thought that 'I' could fix a speaker's reference divergent from its semantic reference. There simply is no speaker's intended reference. Hence, the conditions

requisite for the possibility of referential misfiring would seem to be absent in the case of 'I'.[22]

Our analysis of the broad reflexivity of 'I' is based upon the idea of a symbolic demonstration. But, as just noted, 'I' operates under certain semantic restrictions inappropriate to the nonlinguistic act that informs its analysis, i.e., demonstration. In particular we must build into that analysis the apparent immunity of 'I' to the accidents of referential ambiguity and referential misfiring. A more satisfactory formulation is: In each of its first-person uses, 'I' refers to one and only one agent, the producer of that very token. This analysis involves, of course, amendments (i) and (ii).

Finally, our characterization of first-person reflexivity in terms of surrogate demonstration may seem objectionable on the grounds that it fails to recognize the use of 'I' in *oratio obliqua*. Here I would argue that we require not so much a revision of or amendment to the Revised K-rule, but rather a qualification. In the sentence, "Phillip believes that I ate the cherries," nothing significant turns on the fact that 'I' falls within the scope of the prefix "Phillip believes that." The 'I' still reflexively refers to the unique producer of that token.

Here Castaneda provides a distinction that is conceptually helpful.[23] It is the difference between the external and internal construction of a singular term. Consider Chisholm's example, "Columbus believed that Castro's island was China." If the singular expression "Castro's island" is construed as internal to the psychological prefix, then it is being suggested, anachronistically, that Columbus himself could have referred in some language to Cuba *as* Castro's island. To avoid this interpretation the phrase "Castro's island" should be construed externally: "Columbus believed of Castro's island that it was China." "Castro's Island," now falling outside the scope of the psychological prefix, shows how the speaker of the sentence referred to Cuba, but implies nothing about how Columbus

---

[22]This may be the point Sidney Shoemaker was trying to express some time ago when he noted that the speaker of "I" had "no latitude" in the determination of the reference of that pronoun. See, Sidney Shoemaker, "Self-Reference and Self-Awareness," *Journal of Philosophy* 65 (1968): 559.

[23]Hector-Neri Castaneda, "The Semiotic Profile of Indexical (Experiential) Reference," *Synthese* 49 (1981): 290.

himself referred to Cuba. Castaneda claims that all indexical occurrences must be construed externally; certainly the occurrence of 'I' demands that construction. Thus in the case of Phillip and the cherries we get, "Phillip believes of me that I ate the cherries." The *oratio recta* occurrence of 'me' in this sentence is a first-person use of that pronoun, and, as such, 'me' reflexively refers to the producer of it. We are then free to construe the *oratio obliqua* occurrence of 'I' as an anaphoric use, referring via the reference of 'me.'

Let us take stock. First-person semantics is defined by a customary use of the pronouns 'I' and "me" in natural language. We utilized the "he himself" locution to provide a criterion by means of which uses of 'I' irrelevant to first-person semantics could be filtered out. This criterion pointed to reflexivity as the salient feature of first-person, singular reference. In an attempt to explain the nature of the reflexivity involved here--i.e., what was called broad reflexivity--we introduced the notion of surrogate demonstration. A use of 'I' is the equivalent, in language, of an overt demonstration of oneself. However, this surrogate demonstration (by contrast with an overt demonstration) is constrained semantically in virtue of its immunity to the referential accidents of ambiguity and misfiring. Also, we showed how certain counterfactual analyses employing the idea of surrogate demonstration could account for the semantic reflexivity exhibited by the use of 'I' in writing and thinking. These results constitute, I believe, an answer to those who follow Chisholm in claiming that a theory of the first-person which entails "the primacy of the linguistic" has little hope of succeeding.