CASTAÑEDA ON HE AND I

JOHN PERRY

In a series of thoughtful, original, and thought-provoking papers published in the late 1960s, Hector-Neri Castañeda brought forcefully to the attention of analytical philosophers the importance of indexicals and demonstratives in thought and in descriptions of thought (1966, 1967a, 1968, 1968a). Castañeda put forward a theory to deal with these phenomena; in this paper I explain and criticize basic parts of his original theory, restricting myself to issues associated with the first person.

I allow myself great liberties in the exposition. Castañeda wrote these papers in a manner neutral among several philosophies of language. This hinders comprehension at times. The more central aspects of his theory are also easily lost in a barrage of detail and argument. I present his theory as an attempt at a conservative revision of a Fregean theory of language, thought, and action, in response to discoveries about indexicals and demonstratives. This may produce some distortion and unfairness; I hope it does not.

Elsewhere, I have argued for an account of the issues discussed here that goes along different lines from Castañeda's (Essays 1–4). This account also began as a conservative revision of Frege, making a different choice of what to save and what to give up (Essay 1). But it has led to an increasingly radical departure from the Fregean approach, most recently in joint work with Jon Barwise, who had come to a similar skepticism about the Fregean perspective through his work on definability and on the logic of perception (Barwise and Perry 1980, 1981; Barwise 1981; Essay 6). I do not develop these views here. Rather, in writing this paper, I have tried to recapture the Fregean spirit from which I originally approached Castañeda's pioneering work, in order to indicate as clearly as possible the reasons Castañeda's approach seems to have difficulties, even from within a basically Fregean perspective. In the last section, however, I offer a diaphilosophical fantasy intended to suggest some grounds for dissatisfaction with the entire approach.

The Original Position

I shall describe as a starting point a theory of thought, language, and action that I characterize as "Fregean." I believe most of its elements are found in Frege, particularly in his later essays on the philosophy of logic (1892/1960, 1918/1967). But there are differences of emphasis, of detail; points of great importance to Frege are not mentioned, and so forth.

There are senses that, though not mental, can be directly grasped by the mind. Senses are wholes determined by their parts, called constituents.[1] Some senses are propositions. Minds can have the attitude of belief, as well as other attitudes, towards propositions. Belief in a proposition is a mental state; the propositions a person believes are an important fact about her or him, which, together with other facts, including facts about other propositional attitudes, are the basis of our explanations and expectations of purposeful actions.

Senses have unique references. Some senses have objects as references, others, properties;[2] propositions refer to truth-values. The reference of a complex sense depends on the way things are in the world. We shall use "ref" for the function that assigns a reference to a sense.

Expressions in a language express senses and are the means by which thoughts, including beliefs, are communicated. We shall use "sen" for the function that assigns to each expression its sense in English. Thus,

$$\text{ref}(\text{sen}(\textit{The President of the U.S.})) = \text{Reagan}.$$

We shall say that expressions *designate* objects, properties, etc. When an expression α has a sense, its designation must be the reference of that sense:

---

[1]This useful notion of a *constituent* of a proposition is probably one Frege did not have, and one of which he would not approve, since it suggests that a given proposition ("thought") would be the value of a unique set of arguments.

[2]I ignore the differences between Frege's *Begriffe* and more traditional properties, and, in particular, the idea that they are functions.

$\text{des}(\alpha) = \text{ref}(\text{sen}(\alpha))$.
$\text{des}(\textit{The President of the U.S.}) = \text{Reagan}$.

But we shall not assume that expressions only achieve designation by having sense.

Sentences that do not embed other sentences, like *The President of the U.S. is a Republican*, we shall call simple. Those that do, like *Agnes believes that the President of the U.S. is a Democrat* and *The President of the U.S. is a Republican or the President of the U.S. is a Democrat*, we call complex. The sense of a simple sentence is a whole made up of the senses of the expressions in the sentence. The reference of such a sentence is determined by the references of the senses of its parts. But with some types of complex sentence, these principles do not hold.

In particular, *A believes that S* does not have the proposition expressed by S, but a sense of the proposition expressed by S, as a constituent; this seems quite reasonable, given what was said above, that belief reports describe the relations minds have to senses.

A problem must be noted here. Each successive embedding of a sentence brings in a new level of senses. Thus the proposition expressed by S is just $\text{sen}(S)$. The proposition expressed by *A believes that S* has, as a constituent, a sense that has $\text{sen}(S)$ as reference. The proposition expressed by *B believes that A believes that S* has, as a constituent, a sense that has as reference a sense that has $\text{sen}(S)$ as reference. The problem is the use of the indefinite article in the last two sentences. Which of the indefinitely many senses that refer to $\text{sen}(S)$ is to be a constituent of the proposition *A believes that S*? Any serious Fregean theory must solve this problem, which involves, in this case, providing a route from a reference ($\text{sen}(S)$) to a sense of it.

Beliefs affect action. Most obviously, beliefs affect how one describes the world. Let us say that a person *accepts* a sentence when that person uses it, or is disposed to use it given appropriate conditions, to sincerely describe the way things are. Then, in general, a person's acceptance of S, which is a disposition to act in a certain way, is explained by the person's belief that S. But acceptance is only the most obvious way in which beliefs affect action; beliefs, together with other propositional attitudes, explain purposeful actions in general.

Similarly, we can perceive that the conditions for the truth of a proposition are met, and this results in belief. Thus propositions, and the senses that constitute them, have psychological roles. These would be difficult to articulate, given that action depends not just on belief, but belief together with various other propositional attitudes. But in our ordinary explanations of belief in terms of perception, we show great familiarity with, and facility in, dealing with the psychological roles of senses.

The fact that senses have unique references provides a method of individuation. If $\text{ref}(s) \neq \text{ref}(s')$, then $s \neq s'$. But the psychological roles of senses require and provide a more finely grained principle of individuation.

In particular, even if $\text{des}(\phi) = \text{des}(\psi)$ where ϕ and ψ are sentences, it is entirely possible that both

> A believes that ϕ.
> Not-(A believes that ψ).

In this case, ϕ and ψ must have different senses. Acceptance is a test for belief, so we can speak of the acceptance test for sense identity.

Given the distinction between sense and reference, we get two concepts of the truth conditions of a sentence. Consider these sentences

> The editor of *Soul* smokes cigars.
> The author of "It" smokes cigars.

Given a grammatical analysis of each sentence into a noun phrase and a verb phrase, we can ask (i) what conditions the truth of the sentence imposes on the references of the expressions, or (ii) what conditions the truth of the sentence imposes on the senses of the expressions. We shall call (i) the referential level of analysis of truth conditions. Given that the editor of *Soul* is the author of "It," we get the same condition for both sentences at the referential level. But when we rise to (ii), the sense level, we get different conditions of truth for the two sentences.

The linchpin of this theory is sense. Reading, writing, hearing, speaking, and thinking with expressions play a big role in our lives. All of the properties of expressions that make this so derive from the fundamental property of having a sense. We might represent this central role of sense in the theory with a diagram:

psychological role

expression   sense

(+ facts)

reference
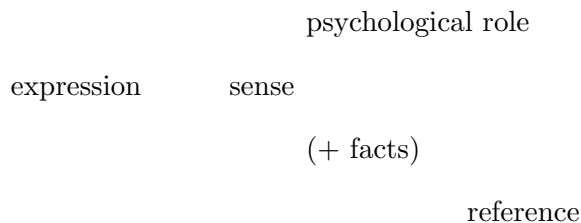
The Problem Posed by I

This theory needs revision or supplementation to deal with indexicals (*I, now, you, here, today, yesterday*, etc.) and demonstratives (*this, that*). Here is a simple, intuitively plausible candidate for a rule that gives the meaning of *I* in English:

> **K-I::** In any statement in which it occurs, *I* designates the speaker of the statement.[3]

---

[3]I use the expression "K" for David Kaplan, not the first to state such a rule, but deserving credit for *pondering* it so fruitfully (1979). Note that in Kaplan's formal theory, we do not quite have the K-I rule, for "I" gets us to a "rigid" intension, not an individual. Indeed, given the way I have set up Castañeda's theory (benefiting, of course, from Kaplan 1969 as well as 1979), it is strikingly similar to Kaplan's. With Kaplan, character and

From this we see immediately that something has to give in the original theory. There we had

meaning of $(\alpha) = \text{sen}(\alpha)$.
$\text{des}(\alpha) = \text{ref}(\text{sen}(\alpha))$.

According to K-I, the designation of I varies from speaker to speaker. Therefore, the sense must vary. Therefore, the meaning must vary. But K-I appears to give a perfectly general meaning for I; the same rule applies when you use it as when I use it.

Castañeda's approach is best represented as recognizing a break between meaning and sense. I has the same meaning, no matter who uses it, but picks up a different sense, with a different reference, and so designates someone different. This result can be achieved by a rule like this:

> **C.I.1::** In a statement made by speaker a, I expresses a sense s such that $\text{ref}(s) = a$.

Rule K-I follows from this rule, so its apparent plausibility is explained.

But this won't quite do. It does not explain the fact that I has a definite psychological role. The sentence *I am wanted on the telephone* has a definite psychological state associated with it, one that, in conjunction with the other beliefs, desires, and values of most of us, would lead us to go answer the telephone. If I can express any of the countless senses that have the speaker of a particular statement as reference, this would be mysterious.

Furthermore, none of the senses already in the theory, to serve, for example, as the senses of definite descriptions or proper names, will work. This is shown by the acceptance test. Given any name or description α that does not contain the first-person pronoun, Castañeda can invent a story in which a reasonable person, who is the reference of α, accepts a sentence containing I but not the sentence just like it containing α, and another in which the opposite happens.

To solve this problem, Castañeda postulates a realm of *special senses*, as I shall call them. Each person's special sense has, for him, the psychological role one associates with I. We shall use "ego" for the function that takes us from any person to that person's special sense. Then Castañeda's basic rule for I is:

> **C-I.2::** In a statement of speaker a, I expresses $\text{ego}(a)$.

───────────

context yield content, a rigid individual concept; with Castañeda, meaning and (something like) context yield a rigid special sense. The great difference is the importance accorded to the "intermediate" property of character or meaning. The step Kaplan took, which Castañeda did not, is developing a theory of character. This allows Kaplan to avoid taking the step Castañeda did: Kaplan's rigid contents need not be *special*; they can be the same contents used, for example, with proper names. The difference is at the intermediate level. Which step was an opportunity missed, and which a cul-de-sac avoided can be argued; my view, in the papers cited, is that it is Kaplan's step that we must take before we can get clear about the psychological role of indexicals. We should also mention A. W. Burks' important work here; in particular, the use of duplication arguments to demonstrate the indispensability of indexicals and demonstratives (1949).

K-I is also a consequence of this rule, so its intuitive plausibility is still explained.

This account of the meaning of I leaves a gap in the theory of belief reports, and Castañeda sees that it also must be revised. Suppose that $i =$ ego(Ivan). How do we report that Ivan has the belief

> that i am wanted on the telephone

(i.e., the proposition consisting of i combined with the sense of *am wanted on the telephone*)? We do not say, *Ivan believes that I am wanted on the telephone*. Instead, we would say, *Ivan believes that he is wanted on the telephone*. But this, as Castañeda brings out by a number of delightful examples, need not be taken to absolutely require that Ivan believes the proposition indicated. To bring out the reading in which it does absolutely require this, Castañeda introduces the term he^*. This word disambiguates *he*; it is only a linguistic accident that we use the same word for imputing self-knowledge as we use for other purposes. Castañeda calls he^* a *quasi-indicator*.

The most straightforward way to interpret what Castañeda says about he^* is embodied in this rule:

C-He*.1:: With antecedent α, he^* expresses the sense ego(des(α)).

Thus,

> Ivan believes that he* is wanted on the telephone

attributes to Ivan just the belief we wanted,

> that i am wanted on the telephone.

Here des(*Ivan*) is Ivan, and ego(Ivan) is his special sense, i.

He*: Some Problems

Consider,

> (1) *Ivan believes that he* is wanted on the telephone.*

As I have stated the theory, (1) expresses the proposition that Ivan believes a certain proposition, namely,

> (2) that i is wanted on the telephone,

where $i =$ ego(Ivan). This theory seems to have grave difficulties.

First, note that on this theory, anyone who can believe anything about Ivan can believe (2). Suppose, for example, that this is true:

> (3) *Sheila believes that Ivan believes that he* is wanted on the telephone.*

And suppose further that Sheila believes the proposition

> (4) that Ivan is wanted on the telephone.

(3) tells us that Sheila believes a proposition with a sense of the sense i as a constituent.

It seems that if the belief of Sheila's reported by (3) leads her to expect Ivan to go answer the phone, as it surely might, she *must* believe (5).

(5) that Ivan is i.

But from (5) and (4), (2) follows. So it would be very surprising if Sheila did not believe (2).

Some things that Castañeda says have led readers to suppose that this is not possible on his theory, that "first-person propositions," like (2) are private, that only those referred to by the special senses therein can grasp them, much less believe them. But this is a misreading; Castañeda quite explicitly says the opposite, in his reply to Kretzmann 1966.

Kretzmann claimed that since there are certain propositions only one person can know, as when Jones knows that he (himself) is in the hospital, there are things God cannot know, and so there is no omniscient God. This seems like a pretty good argument, if there are private first-person propositions. But Castañeda says, in his comments on Kretzmann's paper, that Kretzmann's view does not follow from his theory. The example in question is

(7) *Jones knows that he (himself) is in the hospital.*

Castañeda says,

> The expression "he (himself)" in sentence (7) is a quasi-indicator. It cannot be eliminated from (7) by any name or description of Jones that includes no first-person quasi-indicator. It is this fact that leads Kretzmann to say that the statement expressed by the occurrence of "he (himself) is in the hospital" in (7) cannot be known by any other person. But this does not follow. If Kretzmann, or the reader, knows that Jones knows that he (himself) is in the hospital, then, by principle (P) above, Kretzmann or the reader know the very same proposition that by (7) Jones knows to be true. Hence, theism is not, by the present route, incompatible with omniscience (1967a, 209).

Principle (P) is,

> If a sentence of the form "X knows that a person Y knows that . . ." formulates a true statement, then the person X knows the statement formulated by the clause filling the blank ". . ." (1967a, 209).

Castañeda is using *statement* here as I am using *proposition*.

Thus it seems clear that Castañeda intends it to be possible for Sheila to believe exactly what Ivan believes, when Ivan believes that he* is wanted on the telephone.

(He does maintain that there is no simple way in unadorned English to state the proposition that Sheila believes (2); the relevance of this is discussed below.)

There seem to be at least three fairly serious difficulties with this theory.

First, suppose that Sheila and Ivan both believe (2). If the theory is to explain the data, this belief must have different psychological roles for Sheila

and Ivan. Ivan should accept *I am wanted on the telephone* and be disposed to answer the phone but Sheila should not be disposed to speak or act in these ways. What is to explain the difference?

It does not seem an adequate explanation to say that i has Ivan, but not Sheila, as reference. It seems that it would have to be added that Ivan was aware that he* was the reference of i. But in what could this awareness consist? It cannot simply be his belief in the proposition,

> that i is i

for there is no reason to doubt that Sheila has that belief also.

The second problem has to do with reports like (1). This belief report seems like the sort of thing that might explain Ivan's moving towards the telephone to answer it. But it is not at all clear why it should do so. What (1) says, on this theory, is that Ivan believes a certain proposition,

> (2) that i is wanted on the telephone.

Now if the first difficulty can be solved, then it must be that not everyone who believes any first-person proposition accepts *I am wanted on the telephone* and tries to answer the telephone. One only does this when it is one's own special sense that is a constituent.

Therefore, reporting that a person believes a first-person proposition does not provide any explanation of that person's accepting the relevant I-sentence, or performing the appropriate actions. The information that it is his or her *own* special sense that is a constituent of the believed proposition must also be conveyed. On the theory being considered, (1) does not provide this information. Remember that the meaning of he^*, given by rule C-He*.1, is not a constituent of the proposition expressed by (1). The proposition expressed by (1) says nothing about the expression he^*, or its meaning, or about i being the value of the function ego for the argument Ivan. It simply says that Ivan believes (2).[4]

The third difficulty is provided by Castañeda himself (1966). It is also brought up by an example due to Rogers Albritton, which I have altered somewhat (Adams and Castañeda 1983).

Suppose that Ivan has just been appointed the editor of *Soul*, but does not know it, although he has heard that *Soul* has once again appointed a male editor. The following might be true:

> (6) *Ivan believes that the editor of Soul believes that he* is alive.*

---

[4]This is really an ad hominem argument. One is inclined to say that one simply sees the he^* in the sentence, and realizes that since it can be used to identify Ivan's belief, *his* first-person sense must be a constituent of the proposition he is said to believe. We pick up the crucial information from *how* a certain proposition is expressed—by use of he^*—even though the information is not part of what the proposition says. I think this a fine reply, but anyone who uses it is on the verge of accepting the point of view suggested later on anyway.

The reply in effect countenances a third level of truth conditions, in addition to the referential level and the sense level; it is by understanding these that we *see* that it must be Ivan's special sense that is involved.

Where e = ego(the editor of *Soul*), it follows that Ivan believes the proposition

(8) *that e is alive.*

The reasoning is similar to that credited to Sheila above. Ivan believes the proposition

(9) that the editor of *Soul* is e

and the proposition

(10) that the editor of *Soul* is alive.

So he concludes (8). But Ivan surely also believes the proposition that he* is alive, that is,

(11) that i am alive.

But since Ivan is the editor of *Soul*, ego(Ivan) is the same sense as ego(the editor of *Soul*), that is, $i = e$. Therefore, (8) = (11).

It seems that if Ivan would but reflect upon (8) and (11), he would realize that they are one and the same proposition. But then he would know the proposition

(12) that $i = e$.

But then he would believe that he* has been selected editor of *Soul*, for this follows from (12) and (9). But this is contrary to the original hypothesis.

Castañeda considers the possibility that "the heaviest man of Europe could come to know that he* weighs more than anybody else without resorting at all to the scales and comparison of weights" (1966). Castañeda says that "This absurdity arises simply from allowing the tokens of 'he*' to function as independent symbols, i.e., as referring expressions in their own right, without the need of a grammatical and logical antecedent" (152). Castañeda says, "Propositions about a given I can be the full objects of belief (knowledge, assumption, assertion, etc.) only if the belief (knowledge, etc.) in question belongs to that same I" (1968a, 263). And in reply to Kretzmann, Castañeda says, "an omniscient being does not know every proposition in *oratio recta*: indexical propositions he must know in *oratio obliqua*, in the form of quasi-indexical propositions ..." (1967a, 210).

I believe these remarks may be responsible for some people taking Castañeda to be asserting that persons cannot grasp first-person propositions other than their own. But, as we saw, this is not his view.

In these remarks, Castañeda is drawing our attention to the fact that in English there is no simple way of reporting, say, that Sheila believes (2). Neither of the following will do:

> *Sheila believes that I am wanted on the telephone.*
> *Sheila believes that she* is wanted on the telephone.*

The first seems like it ought to work, according to the theory, as long as Ivan is the speaker. But, as we shall see below, Castañeda amends the rule for I so that in such sentences it does not express a first-person sense.

(3) reports that Sheila believes a proposition of which a sense of (2) is a constituent; but there seems to be no straightforward way to report her belief in (2).

In the remarks quoted, this apparent fact about English seems to somehow be used to resolve the third difficulty. But if this is the strategy, I do not believe it succeeds.

There is nothing in Castañeda's theory to explain why English has no way of saying that Sheila believes (2). To stipulate that since English has no way to report such beliefs, there are none to report, would be at best an ad hoc way of solving the problem. But even this is not available to Castañeda; he maintains, in the reply to Kretzmann, that we can have such beliefs and that we can have such beliefs does seem to follow from the theory by two good arguments, the one I gave and the one he gave. So if English has no way to express such beliefs, that just seems to be an odd fact about English. If it is a deep fact about English, it seems an embarrassment to the theory that it provides no explanation of it.

My argument that the editor of *Soul* could figure out virtually a priori that he* was the editor of *Soul* did not depend on allowing *he** to function as an independent symbol, as examination of it will show. But, in fact, the theory seems to give us no reason not to introduce a perfectly intelligible close relative of *he**. We simply let *he*(Ivan)* do, in a sentence where *Ivan* is not available as an antecedent, just what *he** does when it is available. In this slightly augmented English, we can say

> *Sheila believes that he*(Ivan) is wanted on the telephone*

and

> *God knows that he*(Jones) is wanted on the telephone*

to express Sheila's belief and God's knowledge. Finally, the phrases "*oratio obliqua* knowledge" and "*oratio recta* knowledge" seem to me misleading in the quote given above about what an omniscient being can know. Castañeda has argued, against Kretzmann, that an omniscient being can know Jones' first-person propositions in the way of knowing a proposition that is usually reported with *oratio recta*. That unaugmented English has no way of reporting such knowledge does not seem to change this.

So I conclude that appeal to these facts about English will not provide a solution to the third difficulty I have raised.

He*: A Revised Theory

I believe the account of *he** can be amended in a way that takes care of the three difficulties I have raised. The theory as so amended is one that, I think, people have often taken Castañeda to be putting forward, and is in the spirit of a suggestion made by Carl Ginet (Castañeda 1966, 143). This account is compatible with first-person propositions being accessible only to those to whom the special senses in them refer.

Consider,

(3) *Sheila believes that Ivan believes that he* is wanted on the telephone.*

On the revised theory the proposition

(2) *that i is wanted on the telephone*

is not a constituent of the proposition expressed by (1). Instead, (1) says that Sheila believes that Ivan believes the proposition resulting from combining his special sense with the sense of *is wanted on the telephone*.

Let s be a variable ranging over senses. Then we can say such things as,

For some s, Ivan believes that s is wanted on the telephone.

This does not mean that Ivan thinks someone wants to talk to a sense, of course. It means, rather, that there is a proposition, resulting from the combination of some sense s with the sense of *is wanted on the telephone*, that Ivan believes.

To achieve the desired effect, we need to treat he^* as an expression that abbreviates a construction that binds such sense variables. The rule may be put as follows:

He*.2: With α as antecedent,

> *A believes that he* is so-and-so*

has the sense of

> *There is an s such that s = ego(α) and A believes that s is s-and-so.*

Using this rule, (3) says,

> *Sheila believes that there is an s such that s = ego(Ivan) and Ivan believes that s is wanted on the telephone.*

On this theory, i, Ivan's special sense, is not a constituent of the proposition Sheila believes, but is a constituent of the proposition she believes that Ivan believes. On the other hand, the sense of the expression

(13) *ego(Ivan)*

is a constituent of the proposition Sheila believes. It is easy to get confused here, since (13) has a sense and designates another sense. Only the former is a constituent of the proposition (3) expresses.

Let us turn to (1). We want (1) to attribute to Ivan not simply the belief that a first-person proposition of his is true, but belief in that proposition. The proposal gives this result. (1) says,

> *there is an s such that s = ego(Ivan) and Ivan believes that s is wanted on the telephone.*

Although (2) is not a constituent of the proposition expressed by (1), Ivan has to believe (2) for (1) to be true.

Given this account of he^*, the arguments for the possibility of believing (3), or God's believing Jones' first-person proposition in Kretzmann's example, do not work. This undercuts Castañeda's criticisms of Kretzmann, but it allows Castañeda to claim that the special senses are private. He can

then claim, against the first objection raised, that any normal person who believes (2) will accept "I am wanted on the telephone" and go to answer the telephone. He could claim that it is intrinsic to (2) that belief in it has this role. But since Ivan is the only person who can believe it, there is not a general rush to the telephone.

The second objection, that (1) does not explain Ivan's telephone-answering behavior, is also now answered. The sense of the expression (13) is a constituent of the proposition expressed by (1), although not of (2), the proposition Ivan must believe for (1) to be true. So (1) now tells us what the connection is between Ivan and the special sense that is a constituent of the proposition he is said to believe. Given general knowledge of the psychological role of first-person propositions, (1) does tell us that Ivan may be expected to accept *I am wanted on the telephone* and to go to answer the phone, assuming he is relatively normal.

The a priori argument by which anyone who is editor of *Soul* could figure out that fact will not now pose a problem. (6) now attributes to Ivan a belief that has the sense of the expression

  ego(the editor of Soul)

as a constituent, but not e (the editor's special sense) or i (Ivan's special sense). The argument by which Ivan was to overcome his ignorance of his appointment does not get started.

Toward the Unity of He

Once we have adopted the revised theory of *he**—and I am not saying Castañeda ever has or will—certain questions naturally arise. On the revised theory, *he** does not seem so distinct from other uses of *he* as Castañeda argued. But when we pursue this, we find ourselves wondering whether there really is a separate sense of *he* corresponding to *he** after all.

Castañeda isolates what he calls the "(F)-use" of *he*:

> (F) "He" is often employed as a place-holder for some *unspecified* description which refers to a previously mentioned object .... Examples are "Paul said (believes, knows) of (someone who in point of fact is) Mary that she is happy."
> ... and, of a different linguistic form "Paul saw Mary and believes that she is happy" (1966, 133).

Castañeda gives an analysis of such uses. Consider,

(14) *Paul believes of Mary that she is happy.*

Castañeda's analysis is

> *There is a sense s such that ref(s) = Mary and Paul believes that s is happy.*

Consider (1) once more,

(1) *Ivan believes that he* is wanted on the telephone.*

If we were to take the he* in (1) to be simply an (F)-use of *he*, the analysis would be

> *There is a sense s such that ref(s) = Ivan and Ivan believes that s is wanted on the telephone.*

This is not so far off; if we could restrict the domain of senses to the special senses, we would have just what we want. But then, why not simply treat he* as complex; *he* is an (F)-use of *he*, while * restricts the domain of senses to the special senses?

The relevant rules look like this:

(F).1: With α as antecedent,
> *B believes that he is so-and-so*

has the sense of
> *there is a sense s, such that ref(s) = α and B believes that s is so-and-so.*

(*): Appended to an F-use of *he*, * restricts the domain of senses to special senses.

(Note that Castañeda's explanation of (F)-uses mentions unspecified *descriptions* of previously mentioned objects. But there is no obvious reason why we should limit ourselves to senses of descriptions.)

Castañeda 1966 gives arguments against treating *he** as an (F)-use of *he*. But I do not believe these arguments are effective against the present proposal. The first argument (135–36) is directed against an attempt to

treat uses of he* simply as uses of *he*, not against a proposal like the present one, which treats *he** as semantically complex, built from the (F)-use of *he*. Arguments that show that uses of he* do not go proxy for nonspecial senses (senses of names or descriptions not containing *he**) are irrelevant, so long as rule (F) is not limited to such senses.

Castañeda says,

> It is only a linguistic freak that "he" in the sense of "he*" looks exactly like the third-person pronoun "he," which occurs, for instance in "Arthur came, but he knew nobody he saw; he left early" (1966, 132).

For those who are skeptical of such linguistic freaks, the present suggestion may seem plausible. And perhaps we can go further.

There seem to be many useful restrictions on the domain of senses that could be combined with (F)-uses of *he* for various purposes. Indeed, Castañeda suggests one, by way of an alternative treatment of (F)-uses. On rule (F), or at least analogous rules for *it*, we can say of B that he believes of the morning star that it is the evening star, so long as B believes that the evening star is the evening star. To avoid this result, Castañeda suggests incorporating into (F) a requirement that restricts the domain of senses to those not expressed by expressions in the relevant embedded sentence, in this case eliminating the sense of *the evening star*. And Castañeda notes that he "can easily conceive of" other implicit restrictions of this sort.

Here are some examples in which it seems plausible to appeal to such implicit restrictions.

Suppose I am explaining why Albert went over and asked Mary if he could borrow her copy of *Word and Object*:

> *Albert wanted to look up something in Word and Object, and he knew, of Mary, that she had a copy, so he walked up to her.*

This is really an incomplete explanation, as it stands. To make it work we need to assume that the sense with Mary as reference that is a constituent of the proposition Albert knew is, say, the sense of *the woman I see in the corner* and not, say, the sense of *the person I met last week who is now in Madrid watching a bullfight, for all I know*. Why do we assume this? One explanation is that there is a special *she* quasi-indicator, whose meaning forces the restriction to senses whose constituency in the proposition believed by Albert explains his actions. Another is that we simply have an (F)-use of *she*, but that its occurrence in the context of explanation and common sense force the assumption. It is not the *she*, but the *so*, that forces it. One argument in favor of this is that we might criticize the explanation by saying, "That is not why he walked up to her, because although he did know that, he did not recognize her." We would be allowing that the part of the sentence before the *so* was true, but denying that the facts that made it true are explanatory. If *she* were a quasi-indicator, we should deny that Albert knew what he is said to know.

You ask me why I have bet so much money on the outcome of the Orange Bowl. I say, *Jimmy the Greek has seen the team from my state play, and said that it would win the Orange Bowl*. This would be rather misleading if what Jimmy the Greek had said was,

> *The next winner of the Orange Bowl will win the Orange Bowl*

or

> *The best team will win the Orange Bowl*.

It would be misleading even if the team from my home state is the best team and the next winner of the Orange Bowl. We assume that he expressed a proposition with a sense that would give me some reason for betting on the team from my home state. We do not assume this because of any special meaning of the word *it* contained in my sentence, but simply because unless he expressed such a proposition my explanation is stupid.

Or consider: *Ivan saw Sheila from a distance, and believed she was not Sheila*. Here we assume that Ivan thought something like *That person is not Sheila*. Suppose Sheila is also the new editor of *Soul*, but Ivan did not know it. He sees her from a distance, recognizes her, but thinks, *The new editor of Soul is not Sheila*. It would be very misleading to report this last incident with the sentence indicated. We restrict the candidates for the proposition Ivan believes to those with "demonstrative senses" as constituents, but the restriction is an assumption guided by common sense, not sensitivity to a special meaning of *she*.

I think, in fact, that nearly all (F)-uses of *he*, *she*, and *it* would naturally be regarded as placing some implicit limitations on the senses that can be parts of the incompletely specified propositions. The limitations are not placed by a number of different senses of these pronouns, but by common sense and explanatory relevance. The restrictions seem to come in *not* when we understand the proposition expressed by the sentence containing the pronoun, but when we figure out how the facts would have to be to make the proposition true and explain what is in question.

It is a very natural step to begin to think of he^* not as marking a separate sense of *he* at all, but as an important use of *he*, conforming to a well-known pattern of implicit restriction to special senses.

That is, we might conjecture that where the antecedent of an (F)-use of *he* is an expression that designates the believer of a proposition expressed by a sentence in which it occurs, we always assume, unless the suggestion is cancelled, that the domain of senses is restricted to the special senses. This is not a matter of entailment by a special sense of *he*, but a matter of common sense and Gricean implicature.

Suppose I say,

> *Privatus believes that he is rich*.

I think one would ordinarily expect that I intended to imply that Privatus accepted "I am rich." But this suggestion can be cancelled:

> *Privatus believes that he is rich, but in an odd way, since he is one*
> *of Castañeda's characters.*

The additional information cancels the expectation; instead one expects Privatus to accept something like *The editor of Soul is rich* or *The man in the mirror is rich*.

This is not to say that one cannot *introduce* an expression that works as we have defined * to work. But it is an interesting fact that English can get by without it, and, as we shall see, a fact that leaves the way open for more drastic revisions.

Himself

We do have in English the expressions *herself* and *himself*. And Castañeda often inserts *himself* after *he* in his initial explanations of quasi-indication. One might conjecture, then, that *himself* in ordinary English does just what we imagined * to be doing in the last section, systematically restricting the domain of senses. Note that it is treating he^* as complex that makes this suggestion possible.

I do not think this is the way *himself* works, however.

First, note that this expression can be used in nonepistemic contexts:

(15) *Elwood bites himself*.

(15) may be contrasted with

(16) *Elwood bites Elwood*.

(15) seems to involve an intransitive verb phrase, *bites himself*; (16) a transitive verb phrase, *bites*. A natural proposal is that *himself* is governed by these syntactic and semantic rules:

> (Himself):
> (i) Where δ is a transitive verb, δ *himself* is an intransitive verb.
> (ii) α δ *himself* is true if and only if α δ α is true, and α designates a male.

The device *himself* thus gives us two ways of stating what is intuitively a single fact, that Elwood bites Elwood. Why should we want two ways of saying this?

Note that the intransitive verb phrases, *bites himself* and *bites Elwood*, give us two quite different principles of classification. The class of dogs that bite themselves may have interesting attributes in common, and so may the class of dogs that bite Elwood, but they are probably not the same attributes. The real point of *himself* shows up in sentences like

(17) *Like every dog that bites himself, Elwood is covered with sores*.
(18) *Like every dog that bites Elwood, Elwood has broken teeth*.

Similarly, the class of people picked out by the verb phrase,

> *believes himself to be rich*

differs from that picked out by

> *believes Privatus to be rich*.

This makes it quite easy to see why

(19) *Privatus believes himself to be rich*

generates quite different suggestions than

(20) *Privatus believes Privatus to be rich*.

In day-to-day discourse, sentences are often proffered as explanations in a way that requires a lot of filling in by a listener. Even if (19) and (20) are both true just in case Privatus believes Privatus to be rich, they naturally answer different questions. Suppose that the generalizations *Everyone who believes himself to be rich is a snob* and *Everyone who believes Privatus to be rich is amazed* have both been asserted and left unchallenged in a conversation. Then (19) would be a good answer to the question *Why is Privatus so snobby?* and (20) would be a good answer to *Why is Privatus so surprised?*

There is, in fact, a very common and important way of believing oneself to be rich, which, on Castañeda's theory, involves believing that s is rich, where s is one's special sense. It is not surprising then, that the use of the verb phrase *believes himself to be rich* so strongly suggests that Privatus believes it in this way.

It does seem to me that the suggestion is still cancelable, however. Here is an example from Jon Barwise. The dean has been complaining that professors who publish less than ten articles per year on the average are overpaid. He has particular ones in mind, Professors A, B, Q, and Z in the blind sample he has been studying. Then one day he counts the articles he has written and finds only ninety-three articles over the past ten years, agreeing exactly with the figure for Professor Z, which could not be a coincidence.

We say,

> *The dean was surprised to find that he believed himself to be overpaid*.

Toward a Little More Unity

Besides the quasi-indexical and (F)-uses of *he*, Castañeda distinguishes five other uses. Three of these are, roughly speaking, demonstrative uses. The other two are often grouped together by philosophers and linguists, as the use of *he* as a variable of quantification:

> *Somebody came when I was out and he returned my book.*
> *If Arthur comes late, he will call.*

Both of these can easily be translated into the predicate calculus in a way so that the function of the *he* is taken by a bound variable. One way to handle the second sentence is

> *There is an x such that x = Arthur and if x comes late, x will call.*

But it is much simpler to simply regard proper names, as Montague does, as capable of binding variables,

> *If Arthur x comes late, x will call.*

Those who find this strange should recall that in English *somebody* and *everybody* do not come equipped with variables any more than *Arthur* does, and that most quantifiers in English cannot be treated as logical expressions (Barwise and Cooper 1980).

Suppose we think that there is such a bound variable use of *he*. Can we see the (F)-use of *he* as simply what happens to the bound-variable *he* when it finds itself in a sentence embedded in a propositional attitude sentence?

When variables or pronouns regarded as variables are found in such embedded sentences, there is a familiar problem for Fregean theories. Within such embedded sentences, expressions are to refer to their usual senses, and contribute senses of those senses to the propositions expressed by the embedding sentence. But, on the usual treatment, variables have no sense, only designation. Castañeda's analysis of (F)-uses is in fact very similar in spirit to attempts to solve this problem. (Cf. Kaplan 1979.) The basic idea is that we go from the designation of the variable or pronoun to the set of senses that have that object as reference, usually restricted in some way or other. One of these senses has to be a constituent of the proposition, belief in which is attributed.

Two sections back I advocated seeing the *he* in

> *Ivan believes that he is wanted on the telephone*

as an (F)-use of *he*, which picks up, as virtually all (F)-uses do, commonsense restrictions on the domain of senses. In this case, the natural restrictions are to special senses, unless something to the contrary is indicated. Now we can regard this as simply a use of the bound-variable use of *he*.

I

Rule C-I.2 actually simplifies Castañeda's treatment of *I*. Consider Ivan's remarks:

(21) *Sheila believes, of me, that I am wanted on the telephone.*
(22) *Sheila believes that I am wanted on the telephone.*

In these cases, *i* clearly need not be part of the proposition Sheila is reported to believe.

Castañeda calls the use of *I* exhibited by (21) an (F)-use. Like an (F)-use of *he*, he regards it as a "place-holder for some unspecified description or name of the person" to whom the antecedent refers. This result is obtained by an additional rule:

> (P′) A statement of the form "X E's of me that ϕ (I)" is the same as the statement of the corresponding form "There is a way of referring to a certain person as Z, I am that person and X E's that ϕ (Z)" (1966, 147).

On rule C-I.2, *I* does provide a sense, unlike a variable or a use of *he* as a bound variable. The problem that arises with (21) cannot, then, be the same problem that arose with *he*. It is not that *I* provides no sense to serve as the reference within the embedded sentence, but that it provides the wrong one. Rule (P) is simply an additional rule, in no way motivated by C-I.2; Castañeda recognizes this fact when he says that this use of *I* "... is at bottom not an authentic first person use of *I*" (1966, 147).

Note that the rule K-I would leave us in exactly the same position with an *I* in an embedded sentence as we are with a *he*; designation but no sense. The fact that in both cases we seem to be left with an unspecified sense for the proposition expressed by the embedded sentence would be no accident.

(22) has to be treated differently, since there is no authentic use of *I* or *me* to serve as antecedent. Castañeda introduces a third rule for it:

> (P″) A statement of the form of "X E's that ϕ (I)" is the same as the corresponding statement of the form "There is a way of referring to a certain person as Z, X can identify Z (in the relevant respect, or knows who Z is), I am Z and X E's that ϕ (Z)" (1966, 149).

Now consider

(23) I believe, of me, that I am wanted on the telephone.
(24) I believe that I am wanted on the telephone.

I do not sense much difference between these. But it seems that either of them might be used, by Ivan, to explain his acceptance of *I am wanted on the telephone* and his rising from the table to answer the phone. Principles (P′) and (P″) do not explain this, for according to them, the proposition in which Ivan expressed his belief has not been specified.

Now one might reply, at this point, that although the exact proposition has been left unspecified, one can see what it probably is. We naturally assume, when someone says (23) or (24), that it is their own special sense that is a constituent of the proposition they are reporting their belief in.

Since this is exactly the sort of reasoning I have been advocating, in arguing that we only need *he* and not *he**, how could I object to this reply?

I do not object, but I wish to draw a moral from this reply. On the Fregean view, the role of attributions of belief is to tell us which propositions the believer has the attitude of belief toward. This knowledge then can play a role in forming expectations or explanations of the believer's actions, in conjunction, of course, with other knowledge.

On the original view, the attributions told us directly which proposition was believed. It was simply whatever proposition was the sense of the embedded sentence. On the accounts of *he* and *I* that we have arrived at, however, things do not work this way at all. The expressions, embedded in a belief report, designate entities referred to by their antecedents. We ask ourselves, based on a number of factors, including the relation of the person to whom belief is attributed and the designation of the antecedent, and the purpose for which the belief report was given, which sense or senses are reasonable completions of the proposition. Then we take the report to attribute belief in one of those propositions.

In this process, we go from designation to sense, but never from sense back to reference.

To see how this is so, first reconsider rule C-I.2. This rule tells us that each time *I* is used, a sense that refers to the speaker will be expressed. But to get to this sense, we must first identify the speaker. The speaker does not get into the picture through his special sense, but the other way round. To use rule C-I.2, we would first have to use K-I to find the speaker.

Now consider Ivan's use of (23). Rule K-I gives us Ivan as the designation of the first *I*. Rule (P′) tells us that Ivan believes some proposition with a sense that refers to him as initial constituent. Common sense, or Gricean implicature, or some mixture, leads us to suppose that it is Ivan's special sense that is needed. *This is the first use we have had for this sense.*

I think that, if we look closely at the way the theory would actually work in figuring out what people believe on the basis of belief-attributions, we would find that this result is quite general. What is important about special senses is that they are expressed by the use of *I* and that they have a certain psychological role. Their reference is immaterial.

Once we see this, certain alternatives to Castañeda's view present themselves. One is that we need nothing in our theory except the meaning of *I* and the designation of *I* on specific occasions of use. The former would be the same for everyone, and correspond to the same psychological role for everyone. It would therefore not be what determines reference, since it would no more refer to one person than another. But it would not *need* to determine reference, since *I* designates without sense—but not without meaning.

Can we get by without special senses? I think so. When I think to myself, *I need to do the taxes*, the meaning of *I*, and the fact that I am doing the

thinking, suffice to make the thought about me. No special sense is needed. *I*, of course, must have a meaning, but the K-I rule will do.

One might ask how *I* can have so crucial a role in the thinking of each of us, if not through expressing a special sense. But I do not think that one need look beyond the K-I rule for an answer. It insures that *I* will have a definite psychological role, the same for everyone who understands it.

These ideas have been developed in the series of articles mentioned at the beginning of the paper. I will not pursue them here, but rather, in the next section, try to put the basic point in a rather different way.

But first I would like to leave those who would still like to pursue Castañeda's line with a couple of challenging examples to chew on that have not otherwise come up:

(25) *Sheila and I both believe that I am wanted on the telephone.*

(26) *Sheila and I each believe that she and I are wanted on the telephone.*

(27) *I believe that I am wanted on the telephone and Sheila believes it too.*

A Fantasy

Let us suppose that God has created the heavens and earth, and populated the latter with humans. These humans have the powers of perception and movement, and there are wired-in connections between the two. When they see a carrot they grab it and eat it, for example. But they do not get enough carrots during the day to thrive. God assigns a very astute angel, Michael, to develop the capacity of belief for them. The idea is that having this capacity will help them: perceptions that do not immediately lead to action will lead to beliefs; later these beliefs may lead to effective actions that would not be occasioned by the perceptions they had then. They will be able to eat carrots they see during the day when they get hungry at night. Michael is given instructions to work out such a capacity and instill it in the humans.

Michael has read Frege, and proceeds as follows. He creates a realm of senses and creates in the humans the power to grasp those senses. By decree, the various senses stand for various conditions, with certain complex ones standing for the unique objects that meet the complex conditions, all this being done in a nicely compositional way. Some senses he calls propositions. He gives humans the power to have the attitude of belief towards propositions, and programs them to believe the propositions that correspond to what they see, and to take actions that according to the beliefs will satisfy their desires. (He plans to give them the capacity for more complex desires once he has solved the problem of belief.) He puts all of this into effect. Nothing happens.

Then Michael reads Castañeda. He diagnoses his problem as follows. The humans perceive and act from a certain position in space and time. The belief that a certain type of perceptual experience should lead a given individual to have, and the actions that certain beliefs should lead that individual to perform, depend on this position. When Ivan, the king of

France, perceives a carrot in front of him, he needs to be led to a belief that there is a carrot in front of the king of France. But Sheila, the editor of *Soul*, should be led to the belief that there is a carrot in front of the editor of *Soul*, when she sees a carrot in front of her. And the king of France, when he believes that there is a carrot in front of the king of France, and becomes hungry during the night, should reach out and grab the carrot in front of him. The editor of *Soul*, if she believes there is a carrot in front of the king of France, should not reach out, even if she is quite hungry, but she should if she believes there is a carrot in front of the editor of *Soul*. Michael realizes he has not given humans enough to use their beliefs in this way. The king of France's perceptions do not tell him who the carrot is in front of; he does not know whether to believe that there is a carrot in front of the king of France or one in front of the editor of *Soul*.

Michael decides to give each human a very special sense, which will refer to them, by his decree, no matter where they go or whether or not they become the editor of *Soul* or the author of *Waverly* or the king of France. All they have to do is be who they are. He puts such a sense into the system for each created person and resolves to create more as necessary. Ivan and Sheila and all the humans grasp these senses, as well as the old ones. Still, nothing happens.

Thinking this over, Michael sees the problem. There is still nothing in perception or action to tie each person's special sense into an effective network of psychological states. Ivan's special sense i refers to him; Sheila's special sense s refers to her. But this is not the sort of specialness that is important. This specialness guarantees that once they form the belief he wants them to, it will be true. But he has not done anything to get it formed, or to make it effective.

Michael decides he overdid things. He makes each person's sense graspable only by them; this was something that Castañeda sometimes seemed to suggest. It is not clear why this would help, but it does make each person's special sense even more special, so he gives it a try. Still, nothing happens. People are grasping sentences and entertaining propositions. Perhaps they are even writing novels in their heads. But they are not forming beliefs and eating carrots at night.

Michael thinks things over, and finally sees what must be done. He calls upon his subordinate, Penelope. Penelope is to take a list of each person and their special sense, and retool their minds so that their perceptions lead to right beliefs and the beliefs lead to the right actions. When Ivan sees a carrot, he will believe

> that there is a carrot in front of i,

and when Ivan believes this, and is hungry, he will reach out and grab the carrot. In this way, Michael explains, the system of beliefs each person has will be oriented to the objective world, in the way that their perceptions and action have been all along.

"That is a very long list," Penelope says. "Days are getting short, and these people are going to be dropping off like flies if they do not start getting some carrots down at night. I have an idea that will save some time."

Penelope's better idea is very simple: "We choose one special sense, Ivan's, perhaps, and retool everyone so that they form the belief

that there is a carrot in front of i

when they see a carrot. And we retool everyone so that when they have this belief and are hungry, they reach out." She points out that this will make psychology possible, since there will be systematic connections between perceptual states, belief, and types of action, rather than separate link-ups for each person.

Michael is worried. "That is going to knock hell out of reference and truth," he observes. "Everyone but Ivan will be having false beliefs about there being carrots in front of Ivan, when they actually see carrots in front of themselves."

"Granted," Penelope observes. "But their false beliefs will lead systematically to effective action. Those people are starving. We can worry about reference and truth when I get back."

Michael agrees reluctantly. But when Penelope returns, a simple solution to the problem of reference is found. Michael, by decree, makes the sense i no longer stand for anyone. "We will say that the person's belief is true, when they believe

there is a carrot in front of i

if there is a carrot in front of them," he adds, as an afterthought. This change seems to have no effect on the nocturnal carrot-eating, which is now proceeding apace.

Later, pleased with their success, God charges Michael and Penelope with giving humans language. They are apprehensive that their rather rushed solution to the carrot-eating problem will wreak havoc with this project.

The problem is this. Their strategy for language is to have words associated with senses, sentences with propositions. Sentences are true if the associated propositions are.

But there in the middle of the system is a sense with no reference. What is worse, Michael's decree that when someone believes,

that there is a carrot in front of i

they believe truly if and only if there is a carrot in front of them, has somehow to fit in.

Again, Penelope has a promising if rather ad hoc idea. So far, they have been assigning senses to expressions, and letting reference take care of itself. This will not work with i; the word used to express it will not stand for anything, since i does not have a reference. Penelope suggests simply having the word governed by two separate rules,

    (i) *I* expresses i.

(ii) Whenever it is used, I stands for the user.

She explains: "Although i does not refer to anything, and thoughts with i in them are not true or false, it has a definite cognitive role, as a result of all those connections between i, perception, and actions that I built into humans. So humans have to have a word to express i, to explain their actions. Now they can say 'I reached because I believed there was a carrot in front of I'. But your hastily drawn rule says that they believe truly, in such a case, if there *is* a carrot in front of them. When it comes to truth, we need a reference. The same sense cannot have different references; that is clear. So we cannot have I's reference determined by the sense it expresses. We will just have it determined by who uses it, and this is what (ii) does." Again, Michael and Penelope give in to expedience.

But later Penelope reasons as follow.

"Just look at part (ii) of our rule for I. Now suppose a human learns to use the word correctly *just insofar as its use is governed by rule (ii)*. That means that a human will say

> I am planting a carrot patch

if and only if they believe

> that i am planting a carrot patch

and so forth. Rule (i), the rule that allows I to express a certain special cognitive role in human psychology, will be conformed to automatically by anyone who masters rule (ii)!"

"Do you suppose," Michael ruminates, "that if I had started with language and exploited the systematic links God had wired in between perception and action to give words a use, the way we did with I, we could have just worried about reference, and left sense out of it?"[5]