REFERENCE AND BELIEF:

SOME PROBLEMS AND THEORIES

by

JACK BERNARD COBETTO

A.B., Yale University
1978

SUBMITTED TO THE DEPARTMENT OF LINGUISTICS AND PHILOSOPHY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

in

PHILOSOPHY

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

June 1985

© Jack Bernard Cobetto 1985

The author hereby grants to M.I.T. permission to reproduce and to distribute publicly copies of this thesis document in whole or in part.

Signature of Author _____
Department of Linguistics and Philosophy
April 30, 1985

Certified by _____
Richard Cartwright
Thesis Supervisor

Accepted by _____
Richard Cartwright
Chair, Department
Graduate Committee

MASS. INST. TECH.
JUL 10 1985
LIBRARIES

REFERENCE AND BELIEF:

SOME PROBLEMS AND THEORIES

by

JACK BERNARD COBETTO

Submitted to the Department of Linguistics and Philosophy on April 30, 1985, in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Philosophy.

ABSTRACT

In Part One, I offer and defend a solution to Saul Kripke's puzzle about belief. I also consider and reject some of the general claims that Kripke makes regarding the philosophical significance of the puzzle. Finally, I examine and evaluate some of the solutions to the puzzle that have been proposed by other philosophers.

In an appendix to Part One, I discuss Kripke's argument that purports to establish that proper names are not synonymous with the definite descriptions with which they are commonly associated. I present another argument for this conclusion, and show that my argument is not open to several objections that have been raised against Kripke's.

In Part Two, I consider questions like "How is the reference of a proper name determined?" and "How do we manage to use proper names to refer to things?" I argue that although these questions are considered the central questions to which any theory of reference must respond, they have not been given a clear sense. I consider several possible interpretations and show that they are all unacceptable because they do not allow us to regard the various prominent theories of reference, such as the causal and description theories, as offering initially plausible but conflicting responses to these questions.

Thesis Supervisor: Richard Cartwright
Professor of Philosophy

ACKNOWLEDGMENTS

I would like to thank Richard Cartwright and Sylvain Bromberger for supervising this thesis. Professor Bromberger provided much of the encouragement that enabled me to finish this dissertation. Professor Cartwright has had an especially strong influence on me, and I hope that this thesis exhibits at least some of his philosophical clarity and integrity.

I have, at one time or another, discussed various aspects of my work with almost all of the faculty and most of the graduate students in the philosophy program, and I thank them all. George Boolos and Paul Horwich have been especially helpful. And I would like to thank Judith Thomson for encouraging and supporting me in so many different ways. Also, I thank Marcia Lind, Elizabeth Prevett, Susan Russinoff, and Linda Wetzel for their friendship and support.

I would also like to thank Susan Waite for her help and encouragement during the final stages of preparation of this thesis. And I thank Carol Levesque for typing so much on such short notice.

Finally, I thank my parents for making it possible for me to undertake and complete this endeavor.

TABLE OF CONTENTS

PART ONE:

Kripke's Belief Paradox

I

In his article, "A Puzzle about Belief", Saul Kripke argues that philosophers are not justified in rejecting the principle of substitutivity of coreferential proper names.[1] He says,

> When we enter into the area exemplified by Jones and Pierre, we enter into an area where our normal practices of interpretation and attribution of belief are subjected to the greatest possible strain, perhaps to the point of breakdown. So is the notion of the content of someone's assertion, the proposition it expresses. In the present state of our knowledge, I think it would be foolish to draw any conclusion positive or negative, about substitutivity.[2]

Central to Kripke's argument for these claims is the now famous puzzle about Pierre. In this paper, I will offer and defend a solution to Kripke's puzzle. In addition, I will discuss some of the conclusions that Kripke and others have inferred from Pierre's predicament.

Let me begin with a short discussion of the background of the present debate concerning substitutivity. Kripke has argued that names are not synonymous with "associated" definite descriptions.[3] Names are "rigid" designators, whereas their "associated" definite descriptions typically are not. Some have taken this to entail that names do not have any "meaning" or "sense" at all. These philosophers have argued that if proper names do not have any "meaning" or "sense" then their only "semantic function" is to refer to whatever they denote. But if this is so, it is argued, then someone

who says "Cicero = Cicero" would be "making the same statement" or "expressing the same proposition" as someone who says "Cicero = Tully," since "Cicero" and "Tully" refer to the same person. After all, how could they be making different statements? The only difference between the two utterances, it appears, is that one person used the name "Cicero" and the other used the name "Tully." But if the only "semantic function" of a name is to refer to its denotation, it would seem that there can not be any "semantic difference" between the two utterances. The two people must have "asserted the same proposition." The notion of the "semantic function" of a proper name is not clear, and I will discuss at length later on what is meant by "expressing the same proposition" and "making the same statement" and other equally familiar expressions. However, the above formulation of the argument is fairly well known and is adequate for my present purposes.

Many philosophers have argued that the conclusion to the above argument is clearly false. Obviously, it is argued, someone who utters (1) is making a different statement or expressing a different proposition than a person who utters (2).

(1) Cicero = Cicero.

(2) Cicero = Tully.

But if it does follow from the supposition that names are rigid designators that people who utter (1) express the same

proposition as those who utter (2), then if they are not in fact expressing the same proposition, it follows that names are not rigid designators. Again, this argument is fairly well known. It seems that Kripke is left with a choice of rejecting one of the argument's assumptions. Either it doesn't follow from the assumption that names are rigid designators that a person who utters (1) asserts the same proposition as one who utters (2), or the two people actually do express the same proposition. In "A Puzzle about Belief," Kripke argues that the most common argument used by philosophers to establish that (1) and (2) express different propositions is no good.

That argument goes as follows. If (1) and (2) do express the same proposition, then a person who believes the proposition expressed by (1), i.e., that Cicero = Cicero, would also believe the proposition expressed by (2), i.e., that Cicero = Tully. In general, this view of proper names would entail that coreferential names are interchangeable, salva veritate, in belief contexts (i.e., the principle of substitutivity of proper names in belief contexts). However, this is clearly not the case. Suppose that (3) and (4) are both true.

(3) Jones believes that Cicero = Cicero.

(4) Jones does not believe that Cicero = Tully.

If substitutivity holds then we can substitute "Tully" for

the second occurence of "Cicero" in (3). The result is (5).

(5) Jones believes that Cicero = Tully

But (5) contradicts (4). So, either one of (3) and (4) is false or the principle of substitutivity is false. It seems crazy to hold that there couldn't be a situation in which both (3) and (4) are true. Thus it seems that we must reject the principle of substitutivity.

In his paper, Kripke argues, in effect, that it just is not clear that there is a situation in which both (3) and (4) are true. Suppose that (3) is true. Why, asks Kripke, do we think that (4) is true? Well, if we ask Jones "Is Cicero = Tully?," he says no. He says, we may suppose, things like "I do not believe that Cicero = Tully." Since it seems natural to assume that Jones ought to know better than anyone else what he believes, we infer that he does not believe that Cicero = Tully. If it did follow from the fact that a person dissents (reflectively, sincerely, etc.) from a particular sentence that he does not believe the proposition which it expresses, then (4) would be true (if Jones responds as he did above). In other words, if what Kripke calls the "strengthened disquotation principle" is true, it seems that one must grant that (4), along with (3), is true, since Jones assents to "Cicero = Cicero" and dissents from "Cicero = Tully." The main point of the puzzle in "A Puzzle about Belief" is to show that it is not <u>clear</u> that this principle and other of

our "normal ways of attributing belief" are true.

But before I turn to Kripke's puzzle, I want to say a few, I hope, uncontroversial words about puzzles and paradoxes in general. Not all paradoxes are puzzling to the same degree. It is a trivial matter to construct a derivation of a contradiction or of an obvious falsehood from a set of premises. Whether such a derivation is paradoxical depends partly on the plausibility of the premises. If one has good reason to believe the premises, then the derivation of a contradiction or of an obvious falsehood will be puzzling. In general, the more plausible or obvious the premises are, the more paradoxical the derivation will be.

Consider the following paradoxes. First, here is a paradox inspired by Russell:

(6) Some barber "B" shaves all and only those barbers who do not shave themselves.

(7) Either B shaves himself or B does not shave himself.

(8) If B shaves himself, then B does not shave himself, since B shaves only those barbers who do not shave themselves.

(9) If B does not shave himself, then B does shave himself, since B shaves all those barbers who do not shave themselves.

(10) B shaves himself if and only if B does not shave himself.

But (10) is obviously false. It is a contradiction. Of course, the solution to this puzzle is that there is no such barber "B."

There simply is no barber who shaves all and only those barbers who do not shave themselves. The "paradox" is just a reductio ad absurdum from premise (6). But why is this reductio considered a paradox at all? Well, it is paradoxical because it is natural to think, before having seen the argument, that there very well could be such a barber (just as it seemed obvious to people before Russell's discovery that every instance of the axiom schema of comprehension was true). So it is surprising that there can be no such barber. But that is all that it is, just a bit surprising. We need not linger over the puzzle any longer.

In contrast, consider the following version of the liar paradox:

> The sentence in the rectangle in Jack Cobetto's doctoral dissertation is not true.

(11) "The sentence in the rectangle in Jack Cobetto's doctoral dissertation is not true" is true if and only if the sentence in the rectangle in Jack Cobetto's doctoral dissertation is not true.

(12) "The sentence in the rectangle in Jack Cobetto's doctoral dissertation is not true" = the sentence in the rectangle in Jack Cobetto's doctoral dissertation.

(13) The sentence in the rectangle in Jack Cobetto's doctoral dissertation is true if and only if the sentence in the rectangle in Jack Cobetto's doctoral dissertation is not true.

(13), like (10), is, it seems, a contradiction. But (13) appears to follow from (11) and (12). So, it seems, either

(11) or (12), or both, must be false. But both (11) and (12) seem undeniably to be true.

I will not discuss this puzzle in great detail. Perhaps the only thing that is obvious about it is that its solution is not obvious. The problem is that even after one has seen the derivation, both of the premises still seem to be true and the seemingly contradictory conclusion still appears to follow. In contrast, in the barber paradox, once one has seen the derivation, one simply recognizes that premise (6) (perhaps somewhat surprisingly) is false. The liar paradox can not be "solved" in this way. It may be that one of the premises (or both) of the liar paradox is false. But it is not enough just to say that (11) or (12) is false in order to solve the puzzle. A real solution will explain why these premises, if in fact false, seem obviously true. No such explanation is necessary in order to solve the barber paradox, for premise (6), though perhaps initially plausible or unobjectionable to an unsuspecting victim of the paradox, does not any longer seem obviously true. There is no problem in simply accepting its falsehood.

What is the situation with Kripke's puzzle? Is it deeply puzzling like the liar paradox or is it simply a bit surprising like the barber paradox? I will argue that it is more like the barber paradox than it is like the liar paradox. Let us now turn to the puzzle.

Kripke actually offers two slightly different versions of the puzzle, one stronger, one weaker. The weaker version, on which Kripke concentrates most of his efforts, employs the "disquotation principle," whereas the stronger version employs the "strengthened disquotation principle." I will begin by considering the stronger version. (I will assume that the reader is familiar with Pierre and the highly unusual, even bizarre, manner in which he learned English and French).

(SDP_E) A normal English speaker who is not reticent will be disposed to sincere reflective assent to "p" if and only if he believes that p.[4]

(T) If a sentence of one language expresses a truth in that language, then any translation of it into any other language also expresses a truth (in that other language).

(SDP_F) [This is obtained by translating "A normal French speaker who is not reticent will be disposed to sincere reflective assent to "p" if and only if he believes that p" into French.]

(14) [This is obtained by translating "Pierre sincerely assents to "Londres est jolie" into French.]

(15) Pierre does not sincerely assent to "London is pretty."

(16) Pierre is a normal French speaker and also a normal English speaker.

(17) [by (15), (16), and (SDP_E)] Pierre does not believe that London is pretty.

(18) [by (14), (16), and (SDP_F)] Pierre croit que Londres est jolie.

(19) [by (18) and (T)] Pierre believes that London is pretty.

But (17) and (19) contradict each other.

Since (17), (18), and (19) follow from the premises indicated, at least one of those premises must be false. (14) and (15) seem undeniable. And since Kripke went to such great lengths to make sure that Pierre can be considered a "normal" speaker of French and English, let us, at least initially make this assumption. That leaves only the two strengthened disquotation principles ((SDP_E) and (SDP_F)) and the principle of translation (T). The two disquotation principles stand or fall together. It would be incredible to suppose that the French principle is true while the English version is false, or vice versa. Let us turn first to the principle of translation (T).

Is this principle true? Does it even make sense to doubt it? The answers to these questions are not obvious. First of all, the translation principle is a universal statement about all sentences, actual and possible. There are very many sentences that can be constructed. As a general rule in philosophy, when one sees a universal statement such as (T), one becomes or should become very cautious. Nevertheless, one might be inclined to think that the translation principle is as obviously true as the statement that all bachelors are unmarried. It would not make any sense to search for counterexamples to this statement. It is part of the meaning of the term "bachelor" that no one is a bachelor unless he is unmarried.

In conversation, several people have given me the following argument to the effect that it does not make sense to search for counterexamples to the principle of translation. "Sentences are true or false only indirectly in the sense that they are used to express true or false propositions. Since any two sentences which are translations of each other express the same proposition, and since a proposition has only one truth value, any sentences which are translations of each other have the same truth value."

This argument, though likely to spring to mind to any advocate of the notion of "propositions," really does not do the job. Consider the premise that any two sentences which are translations of each other express the same proposition. Is it true? Well, it depends on what one means by "proposition." If one introduces the term "proposition" by stipulating that that is what two sentences which are translations of each other have in common (expressing the same proposition), then although one premise of the argument would be true by definition, it would be possible to deny the other premise by maintaining that "propositions" are not the "primary" truth bearers. That is, one could maintain that two sentences may express the same proposition but have different truth values. On the other hand, if the notion of a "proposition" is introduced in another way such that it is not stipulated that two sentences which are translations of each other express the same proposition,

then it makes sense to look for counterexamples to (T). So, the above informal argument is not conclusive, for it contains premises which are just as much in doubt as its conclusion.

The following two sentences seem to suggest not only that (T) is not analytically true, but that it might actually be false.

(20) Dieser satz ist auf Deutsch.

(21) This sentence is in German.

(21) seems to be the English translation of (20). How else would one translate (20)? Certainly the words "Dieser satz" have the same meaning as the English words "This sentence." The same is true of the rest of the sentences. However, (20) seems true, while (21) seems false.

While the principle of translation (T), as stated, is dubious, perhaps it can be modified to escape the above type of counterexample. Suppose that we modify (T) in a way which yields (T^1).

> (T^1) If a sentence, which contains no indexicals, of one language expresses a truth in that language, then any translation of it into any other language also expresses a truth (in that other language).

For reasons mentioned before, I do not think that this modified principle is analytically true. It is not obvious that there are no counterexamples to it. The best that I can do by way of such an attempt is the following:

Let the biconditional (B_E) be true by definition.

> (B_E) A person, A, bluttered, at t, that London is pretty if and only if A uttered, at t, "London is pretty."

Similarly, I would like to introduce a French version of this definition.

> (B_F) [This is obtained by translating (B_E) into French (introducing another word for "blutter"), and substituting "Londres est jolie" wherever "London is pretty" occurs in (B_E).]

Now suppose

> (22) At t, A uttered "Londres est jolie" (and nothing else).

By (22) and (B_F), we have

> (23) A bluttered at t [in French] que Londres est jolie.

By (23) and (T^1) we have

> (24) A bluttered at t that London is pretty.

But by (22) and (B_E) we have

> (25) A did not blutter at t that London is pretty.

(24) and (25) contradict each other.

There is even less room to maneuver here than in Kripke's puzzle. (B_E) and (B_F), unlike (SDP_E) and (SDP_F), are to be true by definition. Also, there would not seem to be any problem in supposing that (22) is true. Thus we are left with the inference from (23) to (24). There are two steps involved in this inference. First, (24) is taken to be a translation of (23). Also (24) is taken to be true. Since the truth of (24) follows from (T^1) (if (24) is a tranlsation of (23) and neither contains any indexicals), we must say either that (24) is not a translation of (23) or that (T^1) is false.

I am strongly inclined to say that (24) just is not a translation of (23). However, this does not seem to me to be obvious. If the expression "Londres est jolie" appeared in quotation marks in (23), then it would be obvious that they are translations of each other. After all, "A uttered at t "London is pretty"" is clearly not a translation of "A a prononcé "Londres est jolie." When translating "A uttered "London is pretty"" into another language, one does not translate the expressions inside the quotation marks.[5] But the expressions in question in (23) and (24) do not appear in quotation marks. They appear in "that" clauses. And, we do typically translate expressions when they appear in such contexts.

So neither (T) or the modified principle (T^1) is obviously true. We have reason to doubt them without ever being aware of (T)'s use in Kripke's puzzle. Thus the situation here is different from the situations in which we find ourselves when we consider different versions of the liar paradox. In those cases, all the premises seem indubitable even after one has seen the derivation of an apparent contradiction. Although Kripke compares his puzzle to the liar paradox, it just does not have the same stature in that not all its premises have the same degree of indubitability.

However, in spite of this, I do not think that the solution to Kripke's puzzle lies in the rejection of (T), or (T^1), or other modifications of these principles. There is a totally obvious premise that will, when combined with the other premises of the puzzle, serve just as well as principle (T) in allowing Kripke to deduce a contradiction. The premise is (T^{11}).

> (T^{11}) If "Pierre croit que Londres est jolie" is true then so is "Pierre believes that London is pretty" (where "Pierre" is taken to refer to the same person in both sentences and "London" and "Londres" are taken to refer to the same city).

By (T^{11}), we can still infer (19) from (18) and get the contradiction. Since (T^{11}) is not a generalization, we need not worry about there being any counterexamples to it. So, none of the "a priori" doubts that accompany most universal statements need worry us. Furthermore, (T^{11}) is no less plausible than (T). Indeed, part of the reason for believing that general principles (T) and (T^1) are true is that the "instances" of those principles, such as (T^{11}), seem to be true.

Consider the liar paradox again. (11), the first premise of the version given in this paper, is an instance of what has come to be known as the Tarski schema. To most, the Tarski schema has seemed obviously true. But (11) does not gain its plausibility because it is an instance of that schema. Rather the schema seems true because all or most of the instances of it which naturally come to mind seem to be true. The version of the liar paradox given above is no less strong because it

uses an instance of the Tarski schema rather than the schema itself. Similarly, the new version of the strong belief puzzle is no less strong because it employes (T^{11}) instead of (T) or (T^{1}).

Given that the puzzle can be fortified in this way, let us turn our attention to the disquotation principles. Since, as I said before, (SDP_E) and (SDP_F) stand or fall together, I will discuss only (SDP_E). This principle seems to be at least initially plausible. For instance, if we ask a person whether he believes that grass is green and he responds, "Yes, grass is green," we do infer (in normal circumstances) that he believes that grass is green. Similarly, if we had asked that person if he believed that grass is purple and he had responded, "No, grass is not purple," we would have inferred that he does not believe that grass is purple. But just as philosophers have become very good at coming up with counterexamples to initially plausible universal statements, they have become equally proficient at refuting "if and only if" statements. It seems to me that this is exactly what Kripke has done in dreaming up Pierre. He has found a counterexample to an initially plausible "if and only if" statement. It may be surprising that the principle turns out to be false, but it should not be viewed as a deeply troubling result.

Compare the strengthened disquoation principle (SDP_E) to another very famous "if and only if" statement in recent

philosphy.

> (K) A person knows that p if and only if he has a justified true belief that p.

Principle (K) is often referred to as "the traditional analysis of knowing that p." This principle is not only initially plausible, it was for a long time accepted as a correct analysis of knowledge. However, Gettier succeeded in constructing counterexamples to it.

I do not want to discuss Gettier's actual example, but rather how philosophers have reacted to it. When it was presented, the counterexample was not taken to reveal anything deeply puzzling about "knowing that p." It was taken as yet another illustration of how difficult it is to formulate a correct "analysis" of major philosophical concepts. Philosophers reacted by saying, "Gee, that's surprising. I wish I had thought of it. Oh well, it's back to the drawing board." I think this is the reaction that philosophers should have to Kripke's Pierre. The puzzle just shows, among other things, that the strengthened disquotation principles (SDP_E) and (SDP_F) are too strong. They must be modified.

A brief look at several possible modifications of these principles will help us see where they go wrong. Consider the following modification of (SDP_E).

> (SDP_E^1) A believes that p if and only if A sincerely and reflectively (etc.) assents to "p."

This principle differs from (SDP_E) in that it does not have any constraints as to whom or to what "A" may refer. In Kripke's puzzle, (SDP_E) has the condition that the believer (or non-believer) be a normal English speaker. (SDP_E^1), which lacks this condition, does not seem at all plausible. For instance, suppose that A is some dog and "p" is replaced by "the yellow bowl is empty." Also, suppose that the dog is usually fed by putting dog food into the yellow bowl and calling the dog. Whenever the dog is called at around 5:00, he goes to the bowl. If the bowl is empty, he puts his paw into the bowl and moves the bowl around. Suppose that, knowing this, we call the dog at 5:00 and do not put any food into the bowl. As expected, the dog puts his paw into it and moves it around. It is awfully tempting to say that the dog believes that the yellow bowl is empty. After all, that is why he put his paw into it. However, according to (SDP_E^1) this is false. The dog does not believe that the yellow bowl is empty because he is not disposed to (sincerely and reflectively) assent to "the yellow bowl is empty." So, if we want to attribute any beliefs to animals or people who do not speak a language, we must modify (SDP_E^1).

These considerations suggest that (SDP_E^1) be modified as follows:

(SDP_E^{11}) A person, A, who speaks a language, believes that p if and only if he sincerely (etc.) assents to "p."

This principle leaves open the possibility that animals and

people who do not speak a language have beliefs even though they do not assent to or dissent from sentences. But this principle is also clearly inadequate. Suppose that we substitute "snow is white" for "p" and take "A" to refer to some person who speaks only German. This person is not disposed to assent to "snow is white." However, he may still believe that snow is white. His belief may be manifested by his willingness to make assertions like "Ich glaube dass der schnee weiss ist." Obviously, people who speak only languages other than English can believe that snow is white.

What we want to allow is that someone may believe that snow is white and that this belief may be manifested not only by sincere assent to "snow is white" but also by sincere assent to a sentence in some other language that would typically be taken to indicate that the person believes that snow is white. The following principle is an attempt to capture this:

> (SDP_E^{111}) A person, A, who speaks some language, believes that p if and only if he sincerely (etc.) assents to "p" or to some other sentence which is a translation of "p" or expresses the same proposition as "p."

This principle is not identical to (SDP_E). One cannot derive the contradiction which Kripke does if one substitutes (SDP_E^{111}) for (SDP_E), because (SDP_E^{111}), (15), and (16) do not entail (17). Pierre does believe that London is pretty because he assents to "Londres est jolie" which is just a translation of "London is pretty." His failure to assent to "London is pretty" does not allow us to infer by (SDP_E^{111}) that he does not believe that London is pretty.

Given all of this, it is easy to see why Kripke's principle (SDP_E) does allow us to derive a contradiction. It goes wrong in that it does not allow for the possibility that a person's belief that snow is white (or any other belief) may be inferred from his assent to some sentence in another language which is a translation of "snow is white." I do not mean to imply that Kripke's principle (SDP_E) has no plausibility at all. Indeed, when a person actually dissents from a particular sentence such as "snow is white," it is a pretty safe bet that he does not believe that snow is white. The point is that it is not a <u>sure</u> bet. Discovering that it is not a sure bet is the real importance of Kripke's puzzle. The moral of the puzzle should not be taken to be that it, like the liar paradox, reveals some seemingly unsolvable problem with the notion of belief or with our ways of attributing belief. Rather, the puzzle simply presents a counterexample to the perhaps long accepted idea that one can infer from the fact that a person dissents from a sentence such as "London is pretty" that he does not believe that London is pretty. Just as with the traditional analysis of knowledge, we must now go about modifying (SDP_E) if we want to arrive at a correct analysis of belief. Unlike the premises of the liar paradox, after we have seen the problem, we should not feel that (SDP_E) is still, in spite of this, self evident or indubitable.

But modifying the strengthened disquotation principles in an appropriate manner, if Kripke is correct, does not put an end to our difficulties. Indeed, in his article, Kripke does not spend very much time discussing the strong version of the puzzle. Rather, Kripke's official version of the puzzle employs much weaker premises than (SDP_E) and (SDP_F). I will call these weaker disquotation principles "(DP_E)" and "(DP_F)."

> (DP_E) If a normal English speaker, on reflection, sincerely assents to "p," then he believes that p.
>
> (DP_F) [This is the same as (DP_E) except that it is stated in French and "a normal English speaker" is replaced with the French translation of "a normal French speaker."]

Here is the "official" puzzle about belief:

> (26) [This is obtained by translating "Pierre assents to "Londres est jolie"" into French.]
>
> (27) Pierre assents to "London is not pretty."
>
> (28) Pierre is a normal speaker of both English and French.

By (26) and (DP_F), we have

> (29) Pierre croit que Londres est jolie.

By (29) and (T), we have

> (30) Pierre believes that London is pretty.

But by (27) and (DP_E), we have

> (31) Pierre believes that London is not pretty.

There is no contradiction yet; however, at this point, Kripke introduces another premise. He says,

> We may suppose that Pierre, in spite of the unfortunate situation in which he now finds himself, is a leading philosopher and logician. He would never let contradictory beliefs pass. And surely anyone, leading logician or no, is in principle in a position to notice and correct contradictory beliefs if he has them. Precisely for this reason, we regard individuals who contradict themselves as subject to greater censure than those who merely have false beliefs. But, it is clear that Pierre, as long as he is unaware that the cities he calls "London" and "Londres" are one and the same, is in no position to see, by logic alone, that at least one of his beliefs must be false. He lacks information, not logical acumen. He cannot be convicted of inconsistency: to do so is incorrect. [6]

So Kripke takes the following principle to be a premise of the puzzle:

> (C) If a person has contradictory beliefs, then he is, in principle, in a position to tell through the use of logic alone that he has contradictory beliefs.

The truth of (30) and (31) entails that Pierre has contradictory beliefs. But (C) and the fact that Pierre cannot tell through the use of logic alone that he has contradictory beliefs entail that he does not have contradictory beliefs.

What are we to reject here? (26), (27), and (28) seem, as before, obviously true. And even if (T) is not totally obvious, we could still infer (30) from (29) by using the more specific and indubitable (T^{11}) as a premise. The disquotation principles (DP_E) and (DP_F) stand or fall together and seem true. Nothing that I have said so far has cast any doubt on these weaker principles. Indeed, we could still derive a contradiction using the modified

biconditional disquotation principle (SDP^{III}).

This leaves (C) as an obvious candidate for rejection. I am surprised that Kripke uses it as a premise. It is understandable why someone would think that the disquotation principles in their stronger and weaker forms are true. We do, in fact, often attribute belief or disbelief on the basis of the sentences to which, and from which, a person is disposed to assent or dissent. But premise (C) is not so common. The situation here seems to be the same as it was with respect to (SDP_E) and (SDP_F). Premise (C) may be at least initially plausible, but Kripke has simply come up with a counterexample to it. Pierre just happens to be a person who has contradictory beliefs even though he cannot tell that he does through the use of logic alone.

Perhaps Kripke has some reason other than its initial plausibility for thinking that (C) is true. He does not defend (C) any more fully in places other than the quotation given above. I want now to consider some independent reasons for doubting (C). First, Kripke states (C) in different ways in several places in his paper. Sometimes he uses the expression "inconsistent" instead of the expression "contradictory." But this difference is potentially very important. If "inconsistent" is defined in the same way as it is defined in many elementary logic books, two propositions are inconsistent if and only if it is impossible for them both to be true. If

this means that they are inconsistent if and only if there is no possible world (in Kripke's sense) in which they are both true, then (C) would seem clearly false.

Consider (32), (33), and (34):

(32) Cicero was bald.

(33) Tully was not bald.

(34) Tully was bald.

If Kripke is correct, there is no possible world in which (32) and (33) are both true (since "Cicero" and "Tully" are rigid designations of the same person). However, many if not most philosophers would not deny that a person, Jones, can believe both that Cicero was bald and that Tully was not bald. But it takes much more than logical acumen to determine that (32) and (33) are inconsistent. One must know that "Cicero" and "Tully" are coreferential. Thus, if these philosophers are correct, (C) is clearly false when the term "inconsistent" is used instead of "contradictory" and understood in the manner above.

Even if the term contradictory is used in (C), there are other reasons for doubting (C). I will attempt to bring these out by comparing (C) to several closely related principles. Consider (C^1):

> (C^1) If a person understands two sentences which express contradictory propositions, then that person is in principle in a position to determine through the use of logic alone that they express contradictory propositions.

Surely (C^1) is just as plausible as (C). A logician who understands two sentences should be able to determine through the use of logic alone whether they express contradictory propositions. But consider the following possibility. Suppose that Paul grows up in France, and learns to speak French just like everyone else. In particular, he often hears about some city called "Londres," and for this reason, he understands the sentence "Londres est jolie" perfectly well. Unfortunately, Paul suffers a blow to the head and, while unconscious, is shipped off to America. In the U.S., Paul has to learn English from scratch. That is, he can not find any people who speak both French and English who can help him learn English. Also, he lives in a depressed area, and there are no dictionaries or any other kind of device that could be of any help. In spite of this, he manages to learn English. Furthermore, he occasionally hears about a city oversees called "London." Thus he understands the sentence "London is not pretty." However, being dense like Pierre, Paul does not realize that "Londres" and "London" are names of the same city.

Clearly the two sentences "Londres est jolie" and "London is not pretty" express contradictory propositions (when "London" and "Londres" are taken to refer to the same thing). But even if we suppose that Paul like Pierre, is a leading logician of the time, Paul is just not in a position to tell, through the use of logic alone, that the two sentences

express contradictory propositions. Moreover, I have developed Paul's situation so that the only important difference between his story and Pierre's story is that Kripke has told a slightly longer story about Pierre to insure that Pierre will give his sincere reflective assent to both "Londres est jolie" and "London is not pretty." I made no such provisions in telling Paul's story. Only Paul's understanding with respect to the propositions in question, not whether he <u>believes</u> them, is at issue.

This example suggests that Pierre's belief puzzle is not, in a sense, a puzzle about belief at all. Paul is a counter-example to the principle (C^1). He <u>understands</u> two sentences that express contradictory propositions, yet he is not able to tell through the use of logic alone that they do express contradictory propositions. And, given that (C^1) is false, is there any reason to think that (C) is true and that we can not modify Paul's story slightly in order to get him to believe the two propositions which he understands but fails to realize are contradictory? What is it about belief in contrast to mere understanding that insures that (C) is true even though (C^1) is false? Unless one can offer some reason for there being a difference, one is not justified in maintaining (C) (and rejecting (C^1)) in the face of examples like Paul and Pierre.

Let me make this point in a somewhat different way. Here is a variation on Kripke's puzzle (Peter, we shall assume

is in the same sort of difficulty as Paul and Pierre with respect to "London" and "Londres").

(A_E) If a person (in appropriate circumstances) uttered "London is not pretty" at t, then he, at t, asserted (or said) that London is not pretty.

(A_F) [This is obtained by translating "If a person (in appropriate circumstances) uttered "Londres est jolie" at t, then he, at t, asserted (or said) that London is pretty" into French.]

(T) [This is the same as before.]

(35) [This is obtained by translating "At t, Peter uttered "Londres est jolie"" into French.]

(36) At t, Peter also uttered "London is not pretty."[7]

(37) Peter asserted, at t, that London is not pretty.

By (35) and (A_F), we have:

(38) [This is obtained by translating "Peter asserted, at t, that London is pretty" into French.]

By (38) and (T), we have:

(39) Peter, at t, asserted (or said) that London is pretty.

Add to this the following principle:

(C^{11}) If a person at some time, t, has contradicted himself, then he is, in principle, in a position to tell through the use of logic alone that he has.

But from (39) and (37), we have:

(40) Peter, at t, asserted both that London is pretty, and that London is not pretty.

(40) clearly entails that:

(41) Peter, at t, contradicted himself.

However, since we have assumed that Peter is just like Pierre, it is clear that:

(42) Peter is not in a position to tell through the use of logic alone that he has contradicted himself.

But by (42) and (C^{11}), we have:

(43) Peter, at t, did not contradict himself.

However, (43) contradicts (41).

What are we to reject? Surely (A_E) and (A_F) are true, since uttering sentences in "appropriate" circumstances is clearly one of the ways in which we make assertions. Even if (T) is somewhat doubtful, we could replace it with (T^{111}) which is clearly true.

(T^{111}) If the French translation of "Peter, at t, asserted that London is pretty" is true, then so is "Peter, at t, asserted that London is pretty" (where "Peter" and "London" are understood to refer to the same person and city).

Moreover, (C^{11}) is just as plausible as (C) or (C^{1}). Thus we might call this variation of Kripke's puzzle "a puzzle about assertion."

Of course, my solution to this paradox will not surprise the reader. I think we must simply reject (C^{11}). It may have been plausible initially; but it turns out to have counter-examples. People like Peter just can, in some rather unusual circumstances, contradict themselves without being able to tell through the use of logic alone that they have done so. Once we recognize this potential problem, the paradox disappears.

Given that both (C^{1}) and (C^{11}) are false and given obvious similarities between the counterexamples which I have offered to them, what reason is there to think that Pierre is <u>not</u> just

a counterexample to (C)? In light of my examples of Peter and Paul, I think it is possible to see what is, at root, really surprising about all three situations. The problem is that Peter, Paul, and Pierre are all under the mistaken impression that the two sentences "Londres est jolie" and "London is pretty" are used to express different propositions. They do not realize that one is just a translation of the other. It is a bit surprising that this kind of situation could arise. However, once we have gotten over our initial surprise, we should not be shocked that principles like (C^{11}), (C^{1}) and (C), which ignore the possibility of such unusual people as Peter, Paul and Pierre, all turn out to be false. The solution to the puzzle about belief, just as the solutions to the puzzles about understanding and assertion, lies in the simple recognition that some principles which formerly seemed plausible must, in the light of the counterexamples above, be modified or rejected.

II

I have answered Kripke's prominent question "Does Pierre believe that London is pretty?" and given a solution to his puzzle. This is all that he actually asks for in his article. But Kripke also says that the mere discovery of the puzzle forces us to reconsider several widely held views in the philosophy of language. Primary among these is the view discussed at the beginning of this paper, that belief contexts are not "Shakespearean," i.e, that codesignative proper names cannot be substituted for each other, salva veritate, in belief contexts. In the remainder of this paper, I will evaluate and argue against some of the general claims that Kripke makes about the puzzle. Also, I will discuss several solutions which have recently been offered to the puzzle and argue that these solutions support my claims about the general significance of Kripke's puzzle.

Let me begin with the main conclusion which Kripke draws and the one with which I am, to some extent, in agreement. Given our present state of knowledge, the simple conclusion that coreferential proper names are <u>not</u> interchangeable, salva veritate, in belief contexts is not justified. Kripke correctly maintains that his puzzle undermines the most common argument for that seemingly obvious conclusion. Consider sentences (44) and (45).

(44) Jones believes that Cicero was bald.

(45) Jones believes that Tully was bald.

It has seemed to most philosophers that (44) can be true while (45) is false. Suppose that Jones does not know that "Cicero" and "Tully" are coreferential. It is easy to see (especially after one has gotten used to these types of examples) that Jones might assent to "Cicero was bald," but dissent from "Tully was bald." If one asks Jones what he believes, he will respond that he believes that Cicero was bald but not that Tully was bald. But who should know better about Jones' beliefs than Jones himself? As Kripke points out, we infer by the strong disquotation principle (SDP_E) that (44) is true while (45) is false. Since (45) can be obtained from (44) simply by substituting the name "Tully" for the coreferential name "Cicero," it follows that co-referential proper names are not always interchangeable, salve veritate, in belief contexts. Kripke's puzzle undermines this argument by casting doubt on the strong disquotation principle (and other similar principles). Indeed, I've argued that puzzling Pierre simply shows that this principle is false. Either way, the above argument is seen to rely on a premise, (SDP_E), that is, at best, doubtful. Thus we are not justified in accepting the argument's conclusion.

However, while I agree that the argument is not conclusive, it is not clear to what extent it has been weakened by the discovery of the puzzle. The strength of the argument depends on the strength of our inference that Jones does not believe that Tully was bald from the fact that he does not assent to "Tully was bald." While the strong disquotation principle is false or at least doubtful, might there not be some other principle just as plausible as (SDP_E) which not only enables us to make the desired inference but is true as well? Suppose that we modify (SDP_E) as follows (again, for the sake of simplicity I will give instances of the intended principles):

> (MDP) A normal English speaker, who speaks no other language, believes that Tully was bald if and only if he gives his sincere (reflective, etc.) assent to "Tully was bald."

Obviously, the idea behind the modification is precisely to exclude Pierre and other bilingual counterexamples to the original strong disquotation principle. Since Jones, we may suppose, speaks only English, it is valid to infer from (MDP) that Jones does not believe that Tully was bald, and that (44) is true while (45) is false.

The important question is whether any of the considerations Kripke has raised with his puzzle cast any doubt on (MDP) and other instances of the intended

principle behind it. Here is what Kripke has to say about this:

> Jones' situation strikingly resembles Pierre's . . . Intuitively, Jones' assent to both "Cicero was bald" and "Tully was not bald" [and his dissent from "Tully was bald"] arises from sources of just the same kind as Pierre's assent to both "Londres est jolie" and "London is not pretty" [and his dissent from "London is pretty]."[8]

But is Jones' case really "just the same" as Pierre's? Certainly there are striking similarities. There are also striking differences. It does seem true that both Jones and Pierre behave (with respect to the sentences to which and from which they will assent and dissent) in the unusual ways in which they do because they are unaware that two proper names are coreferential. If Pierre knew that "Londres" and "London" both denote the same city, he would not assent to both "Londres est jolie" and "London is not pretty," and dissent from "London is pretty." Similarly, if Jones knew that "Cicero" and "Tully" are coreferential, he would not assent to both "Cicero was bald" and "Tully was not bald," and dissent from "Tully was bald." However, the cases are different in that the names about which Jones is mistaken are both names in the English language, whereas the names about which Pierre is mistaken are not both names in the English language. I do not know and will not discuss whether

this difference is important or perhaps even crucial. My point is only that Kripke is giving an argument from analogy. The strength of this argument against the claim that one can not substitute coreferential proper names, salva veritate, in belief contexts can not be assessed until the strength of this analogy is determined.

Let us now turn to some other conclusions which Kripke feels are warranted in light of puzzling Pierre. Consider the following passage from his article:

> When we enter into the area exemplified by Jones and Pierre, we enter into an area where our normal practices of interpretation and attribution of belief are subjected to the greatest possible strain, perhaps to the point of breakdown. So is the notion of the content of someone's assertion, the proposition it expresses.[9]

This brings the notions of a "proposition" and "propositional content" into the discussion for the first time. Kripke seems to be saying that his puzzle causes a problem for these notions. Indeed, this impression is reinforced when we look at a similar quotation at the end of the preface to the book Naming and Necessity. He says,

> How this relates to the question of what "propositions" are expressed by these sentences, whether these "propositions" are objects of knowledge and belief, and in general, how to treat names in epistemic contexts, are vexing questions. I have no "official doctrine" concerning them, and in fact I am unsure that the apparatus of "propositions" does not break down in this area.[10]

I want now to argue, at some length, that this suggestion is misguided. There is a clear and natural way of understanding the term "proposition" which is in no way threatened by any of the doubts which are, in fact, justified by Kripke's discovery of the puzzle.

Here is one natural way of understanding the term "proposition." Suppose that the following exchange takes place between three people, A, B, and C, during a serious discussion of the state of the government:

A: "The president is wise."

B: "That's true. I, too, believe that the president is wise."

C: "Yes, the president is wise."

Given only these sincere utterances for information, it seems clear that we can make the following inferences.

(46) A asserted that the president is wise.

(47) C asserted that the president is wise.

(48) B believes that the president is wise.

It seems obvious that (46)-(48) entail, respectively, (49), (50), and (51).

(49) A asserted something.

(50) C asserted something.

(51) B believes something.

It also seems clear that (52) and (53) are in some sense, true.

(52) A and C asserted the same thing.

(53) B believes the same thing that both A and C asserted.

But given that (46)-(53) are all true, are not (54)-(58) true as well?

(54) There is something which A asserted (i.e., that the president is wise).

(55) There is something which C asserted.

(56) There is something which B believes.

(57) There is something which both A and C asserted.

(58) There is something which B believes and which both A and C asserted.

Are not these sentences used to assert the same things that (49)-(53) are used to assert? (54)-(58) do sound a bit awkward, but there may be circumstance in which it would be just as natural to utter them as it would be to utter (49)-(53). And even if there are no such circumstances they may still be true. Let us assume they are.

(54)-(58) grammatically resemble the following sentences:

(59) There is someone whom A loves.

(60) There is someone whom C loves.

(61) There is someone whom B hates.

(62) There is someone whom both A and C love.

(63) There is someone whom B hates and whom both A and C love.

Furthermore, certainly no one would object to representing (in semi-logical notation) (59)-(63) in the following way:

(64) (∃x)(A loves x).

(65) (∃x)(C loves x).

(66) (∃x)(B hates x).

(67) (∃x)(A and C love x).

(68) (∃x)(B hates x and both A and C love x).

But if (59)-(63) are correctly represented by (64)-(68), are not (54)-(58) correctly represented by (69)-(73)?

(69) (∃x)(A asserted x).

(70) (∃x)(C asserted x).

(71) (∃x)(B believes x).

(72) (∃x)(A and C asserted x).

(73) (∃x)(B believes x and both A and C asserted x).

For the moment, let us assume that they are.

It seems obvious that other "propositional attitude" sentences similar to (46)-(53) will entail sentences analogous to (69)-(73) (if (46)-(53) entail (54)-(58)). So, on the assumptions we have been making, there is a class of objects which satisfy open sentences like "A asserted x," "B believes x," "A asserted x and B believes x and C knows x," and others derived in an obvious way from what have come to be known as "proposi-

tional attitude" sentences. Let us call these objects "propositions." Roughly, an object is a proposition if and only if it satisfies or could satisfy an open sentence obtained in the above obvious manner from a "propositional attitude" sentence. This definition is sloppy, but I trust that any philosopher who is at all acquainted with the many discussions involving the term "proposition" will understand what is intended.

Before continuing, let me address two objections that might be raised at this point. First, while introducing the term "proposition" in this way makes our reasons for thinking that there are such things relatively clear, (they are the same reasons we have for thinking that sentences like (69)-(73) are true, i.e., they seem inferable from propositional attitude sentences like (46)-(53).), these reasons may not be <u>good</u> reasons. Indeed, the inferences from (54)-(58) to (69)-(73) are especially suspect. Philosophers of language are fond of pointing out how "mere" grammatical similarities between sentences can mislead if one is not careful. Accordingly, we must be very cautious in inferring anything from the fact that (54)-(58) are grammatically similar to (59)-(63). I will not discuss this issue in detail. In this paper, I will be concerned only with

what properties propositions have, assuming that they exist, and whether this notion of a "proposition" is, as Kripke claims, strained to the point of breaking down in the cases of Pierre and Jones.

The second objection which I want to consider is a bit more difficult to respond to. It would be a simple task to introduce a notion of a "proposition" unrelated to the one which Kripke intends and then show that his claim that the notion of a proposition, in my sense, is "strained to the breaking point" when we enter the realm of cases like Pierre and Jones, is incorrect. It remains for me to establish that Kripke and I have the same notion of "proposition" in mind. Unfortunately, this is not possible. As I mentioned earlier, Kripke does not indicate precisely what he means by the term "proposition." In lieu of such an explanation, my argument will be that my notion of a "proposition" captures an important use of the term "proposition" in the recent history of the philosophy of language and is not threatened or strained in cases like Pierre's and Jones'. To this end, let us first examine some of the claims that are often made about propositions.

Beginning philosophy students are typically introduced to the term "proposition" through the use of pairs

of sentences such as (74a) and (74b), and (75a) and (75b).

(74a) London is pretty.

(74b) Londres est jolie.

(75a) John is a bachelor.

(75b) John is an unmarried male adult.

Philosophers, like most people, have noticed that the two sentences in each of these pairs seem to have something very important in common. These are paradigm examples of two sentences "expressing the same proposition." A person uttering both sentences in one of the above pairs is often said to have "expressed the same proposition." Students quickly catch on to this usage of the term "proposition" and are usually able to point out other pairs of sentences which seem related to each other in an analogous way.

In recent years, however, philosophers have come more and more to the realization that the relationship between the sentences in each of the above pairs is a bit more complex. Most importantly, these sentences are now widely viewed as having <u>several</u> things in common. On the one hand, both of the sentences in each of the pairs can be used to assert the same thing, i.e., that London is pretty or that John is a bachelor. But both

of the sentences in the above pairs may also be said to have the same meaning. This distinction, between what is asserted by a person using a particular sentence and the meaning of that sentence is central to several of the most important recent articles which investigate the notion of a "proposition."

In his famous article "On referring," Strawson says:

> Generally, as against Russell, I shall say this. Meaning (in at least one important sense) is a function of the sentence or expression; mentioning and referring and truth and falsity, are functions of the use of the the sentence or expression. To give the meaning of an expression (in this sense in which I am using the word) is to give general directions for its use to refer to or mention particular objects or persons; to give the meaning of a sentence is to give general directions for its use in making true and false assertions. It is not to talk about any particular occasion of the use of the sentence or expression. The meaning of an expression cannot be identified with the object it is used, on a particular occasion, to refer to. The meaning of a sentence cannot be identified with the assertion it is used, on a particular occasion, to make. For to talk about the meaning of an expression or sentence is not to talk about its use on a particular occasion, but about the rules, habits, conventions governing its correct use, on all occasions, to refer or to assert.[11]

While Strawson's discussion is general and intuitive, others have discussed this distinction (and others) in greater detail. In his article "Propositions," Richard Cartwright carefully distinguishes what is asserted on a

particular occasion by uttering a sentence from several other things with which it has been confused. Here is how Cartwright argues for the claim that what is asserted in uttering a sentence is not always identical with the meaning of that sentence.

> Consider, for this purpose, the words "It's raining." These are words, in the uttering of which, people often (though not always) assert something. But of course what is asserted varies from one occasion of their utterance to another. A person who utters them one day does not (normally) make the same statement as one who utters them the next; and one who utters them in Oberlin does not usually assert what is asserted by one who utters them in Detroit. But these variations in what is asserted are not accompanied by corresponding changes in meaning. The words "It's raining" retain the same meaning throughout.[12]

Also, Cartwright produces examples to show that sameness of meaning of sentence uttered is not even in all cases a necessary condition for sameness of assertion made. The following example, though it is not Cartwright's, will suffice to establish this. Suppose that I assert that I am hot by uttering the sentence "I am hot." Another person could assert the very same thing, i.e., that I am hot, by pointing at me and uttering the sentence "You are hot." Yet, the two sentences "I am hot" and "You are hot" do not have the same meaning. The words "I" and "You" have different meanings and consequently the sentences of which they are constituents differ in meaning as well.

Cartwright even presents an argument to establish that the meaning of a sentence is never identical to what the sentence has been used to assert. He says,

> If what someone asserts, on some occasion, is itself the meaning which the words he utters have, on that occasion of their utterance, then anything predicable of what he asserts must also be predicable of the meaning of his words. But it is obvious on very little reflection that ever so many things predicable of what is asserted cannot (on pain of nonsense) be predicated of the meaning of a sentence. And the fundamental point to be noticed in this connection is that although we may predicate of something asserted that it is (or was) asserted, this cannot be predicated of the meaning of a sentence. It simply makes no sense to say that someone asserted the meaning of a sentence -- any more than it makes sense to say he said it.[13]

In light of these arguments, I will, from now on, assume that Strawson and Cartwright are correct about the need to distinguish, in sentences like (74) through (77), the meanings of the sentences from what they can be used to assert.

Cartwright and Strawson are not the only ones who have noticed this distinction. While they use similar terminology, other philosophers who have discussed the distinction use somewhat different language. For example, in an article entitled "Sentences, Statements, and Propositions," E.J. Lemmon uses the term "proposition" rather differently.[14] While both Strawson and Cartwright

use the word "proposition" to apply to what is asserted, Lemmon uses it, I will argue, in connection with the meaning of a sentence. He says that when we utter (in "appropriate" circumstances) a sentence we both "express a proposition" and "make a statement." For Lemmon, "the statement made" denotes the same thing that Cartwright and Strawson denote with the term "proposition." Unfortunately Lemmon's use of the term "proposition" is a bit problematic. He says that a proposition not only can have a truth value; but unlike a statement, it can even change in truth value over time.

To see what Lemmon has in mind, consider again the sentence "It's raining." Lemmon holds that a person who utters (in "appropriate" circumstances) this sentence today has expressed a proposition which is, we may suppose, true. But a person could utter the sentence tomorrow and, according to Lemmon, express the same proposition. Yet, that proposition will be false tomorrow even though it is true today. What has remained constant in the two contexts of utterance that would justify us in saying that the same proposition has been expressed? Clearly different statements, in Cartwright's and Strawson's sense, have been made. Thus Lemmon is not using the term "proposition" to denote what has been asserted by uttering the sentence.

It would seem that we have two choices. The proposition is either the sentence "It's raining" itself or the meaning of the sentence (which has not changed). Lemmon gives examples that make it clear that he does not intend the phrase "proposition expressed" to denote the sentence which has been uttered. He says,

> If I say, "I am hot," and you, being French, say "J'ai chaud," then we have neither uttered the same sentence nor made the same statement; but there is still a sense in which we have said the same thing, namely, expressed the same proposition.[15]

And since "J'ai chaud" and "I am hot" (if uttered by different people) can not be used to make the same statement or assertion, it seems that Lemmon is using the term "proposition expressed" to refer to the meaning of the two sentences, which has indeed remained constant throughout both contexts of utterance.[16]

Although one might object that too much time has been spent discussing mere differences of terminology. But I think the discussion confirms the claims that I have made regarding my use of the term "proposition." Clearly what I call a "proposition" has been noticed and discussed by many other philosophers. Cartwright refers to it more frequently as "what is asserted," while Lemmon calls it a "statement." Thus my usage of the term "proposition" is a natural one.[15] I have not constructed a "straw man"

concept in order to refute Kripke's claim that the notion of a proposition is stretched to the breaking point in cases like Pierre's and Jones'. With this in mind, let us finally turn to this problem.

How is the above notion of a proposition "threatened" in a situation like Jones' and Pierre's? Again, the general argument in "A Puzzle About Belief" runs as follows. It is <u>not</u> clear that coreferential proper names are not interchangeable, salva veritate, in belief contexts. The typical arguments purporting to show that they are not interchangeable rely on the strong disquotation principle. This principle is one of what Kripke calls "our normal methods of attributing belief." It codifies our actual practices of attributing belief. However, while these practices may work in most cases (says Kripke), they "break down" in cases like Pierre's (and perhaps even in situations like Jones'). In Pierre's case, we simply can't attribute belief or disbelief accurately on the basis of these principles. While we can usually infer correctly from a person's dissent from the sentence "London is pretty" that he does not believe that London is pretty, we can not correctly do this in Pierre's case (again, Kripke only argues for the weaker claim that we <u>may</u> not be able to correctly infer that Pierre doesn't

believe that London is pretty). Thus arguments that depend on these principles do not establish that co-referential proper names are not substitutable, salva veritate, in belief contexts. As I said earlier, I agree with all of this. But how do whatever conclusions we might come to concerning the disquotation principle or the question of the substitutivity of proper names in belief contexts "threaten" the very notion of a proposition? The disquotation principles and perhaps other methods of attributing belief seem to be what are threatened, not the notion of a proposition itself. Perhaps the best procedure is to look a bit more closely at the relationship between my notion of a proposition and questions concerning the substitutivity of coreferential proper names in belief contexts.

Earlier I compared sentences like (76) to grammatically similar sentences like (77).

(76) $(\exists x)$(Jones believes x).

(77) $(\exists x)$(Jones loves x).

The second sentence is true if and only if there is at least one object that satisfies the open sentence (78).

(78) Jones loves x.

The "x" in (78) is transparent in the sense that if we replace it with any expression which denotes an object

that satisfies (78), we get a true sentence. For example, suppose that Mary satisfies (78). By replacing "x" in (78) with "Mary," we get the true sentence (79).

(79) Jones loves Mary.

And, if "the tallest woman in the country" also denotes Mary, we can get the true sentence (80) by replacing "x" with "the tallest woman in the country."

(80) Jones loves the tallest woman in the country.

My notion of a "proposition" involves treating the "x" in (81) as transparent in the above sense, and hence, taking "that" clauses to denote the things, the propositions, which Jones believes.

(81) Jones believes x.

So if "that Cicero was bald" denotes an object (proposition) which satisfies (81), we can obtain the true sentence (82) by replacing "x" in (81) with "that Cicero was bald."

(82) Jones believes that Cicero was bald.

And if "my favorite proposition" denotes the same thing as "that Cicero was bald," then (83) is also a true sentence.

(83) Jones believes my favorite proposition.

It is not easy to see precisely how the nction of a proposition is related to the question of whether co-referential proper names are interchangeable, salva veritate, in belief contexts. For example, consider the

two "that" clauses "that Cicero was bald" and "that Tully was bald." If we assume that these two clauses denote the same proposition, it follows that the occurance of "Cicero" in (82) may be replaced, salva veritate, with the coreferential proper name "Tully." Replacing "Cicero" with "Tully" yields the sentence (84).

(84) Jones believes that Tully was bald.

Since we have assumed both that the two "that" clauses in (82) and (84) denote the same proposition and that the "x" in (81) is transparent, it follows that (82) and (84) have the same truth value. But is the reverse true? Does it follow from the fact (supposing that it is a fact) that "Cicero" and "Tully" are interchangeable, salve veritate, in belief contexts that the corresponding "that" clauses denote the same proposition? There is some reason to think that Kripke believes that it does follow. In the preface to Naming and Necessity, he says,

> A final issue: Some critics of my doctrines, and some sympathizers, seem to have read them as asserting, or at least implying, a doctrine of the universal substitutivity of proper names. This can be taken as saying that a sentence with "Cicero" in it expresses the same "proposition" as the corresponding one with "Tully," that to believe the proposition expressed by the one is to believe the proposition expressed by the other, or that they are equivalent for all semantic purposes.[18]

However, it is just not obvious that the assumption that coreferential proper names are interchangeable, salve veritate, in belief contexts entails that a sentence with "Cicero" in it expresses the same "proposition" as the corresponding one with "Tully" (or, in my terminology, that the corresponding "that" clauses denote the same proposition). To see this, let's suppose that "Cicero" and "Tully" are indeed interchangeable, salva veritate, in belief contexts. If (82) is true, then so is (84). On my account of propositions, it does follow that "that Cicero was bald" and "that Tully was bald" both denote objects, i.e., propositions, which satisfy the open sentence (81) (if 82 and 84 are true). But nothing I have said in introducing the term "proposition" entails that the two "that" clauses denote the same proposition. It is consistent to maintain that they denote different propositions, both of which happen to satisfy (81). Thus we are not justified, in lieu of some further arguments, in assuming that if coreferential proper names such as "Cicero" and "Tully" are interchangeable, salve veritate, in beliefs contexts, then the "that" clauses containing them denote the same proposition.

Should we infer from this that the notion of a proposition employed in this discussion is simply not

what Kripke has in mind in the above quotation? As I said before, this is possible. But I don't think that supposing that he means something else will help. Suppose that Kripke is using the expression "proposition," not for what is asserted or believed (i.e., in my sense), but for the other possible candidate which I have mentioned, the meaning of a sentence. If so, Kripke's claim in the quotation can be rendered as follows:

> (M_1) "Cicero" and "Tully" are interchangeable, salva veritate, in propositional attitude contexts if and only if "Cicero was bald" and "Tully was bald" have the same meaning.

Since the only difference between the two sentences "Cicero was bald" and "Tully was bald" is that one has "Cicero" where the other has "Tully," it would seem plausible to attribute any difference in meaning between the sentences to a difference in meaning between "Cicero" and "Tully." If this is correct, then (M_2) is true (if (M_1) is).

> (M_2) "Cicero" and "Tully" are interchangeable, salva veritate, in propositional attitude contexts if and only if "Cicero" and "Tully" have the same meaning.

(M_2) is an interesting claim. In fact, I will discuss some aspects of it in the Appendix to this paper. It is widely taken for granted that if two expressions are synonymous, they are interchangeable, salva veritate, in all propositional attitude contexts. It is

not so clear, however, that if two expressions are interchangeable, salva veritate, in propositional attitude contexts they have the same meaning. It is not obviously inconsistent to maintain both that "Cicero" and "Tully" are interchangeable, salva veritate, in propositional attitude contexts, and also that they are not synonymous.

However, regardless of whether (M_2) is true, there is a more serious problem with attributing to Kripke the position that propositions are the meanings of sentences. In the quotation I've been discussing, he says that if "Cicero" and "Tully" are everywhere interchangeable, then "to believe the proposition expressed by one ("Cicero was bald") is to believe the proposition expressed by the other ("Tully was bald")." Unfortunately, as we saw earlier, it does not make sense to say that Jones, or anyone else, believes (or knows, asserts, or says) the meaning of the sentence "Cicero was bald." Meanings of sentences can not be, at least in this sense, objects of belief or assertion or any of the other propositional attitudes.[19]

It is tempting to respond to all of this by conceding that Kripke is using the word "proposition" in an unclear and perhaps confused manner, and that he has failed to make the rather important distinction between what is asserted or believed, etc. and the meaning of a

sentence. But even if we grant that Kripke is using the expression "proposition" loosely, it is still difficult to see why he thinks that his arguments concerning the substitutivity of coreferential proper names show that the notion of a proposition is in danger of breaking down in cases like Jones' and Pierre's. The problem is that no matter whether we can or can not substitute coreferential proper names, salva veritate, in belief contexts, this doesn't throw any additional doubt on the truth of sentences like "(∃x)(Jones believes x)." In arguing for the truth of this sentence, I took no position on the question of the substitutivity of coreferential proper names in belief or any other propositional attitude contexts. To illustrate this, suppose that one can indeed substitute "Cicero" and "Tully," salva veritate, in all propositional attitude contexts. Even if this is so, "that Cicero was bald" and "that Tully was bald" might denote different propositions. Nevertheless, it is granted that both the "that" clauses denote some proposition that satisfies the open sentence "Jones believes x." Similarly, suppose that one can not substitute "Cicero" and "Tully" in all propositional attitude contexts. This entails that "that Cicero was bald" and "that Tully was bald" denote different propositions. But this is to grant that they do denote some proposition; that

is, that (assuming that Jones believes that Cicero was bald) there is at least one object, i.e., proposition, that satisfies the open sentence "Jones believes x." Again, we are led to the conclusion that there are things which I, and others, call "propositions" and which are the objects of belief.

So, it is possible to maintain that "(∃x)(Jones believes x)" is true regardless of the position one takes on the question of whether "Cicero" and "Tully" are interchangeable, salva veritate, in belief contexts. Consequently, Kripke's argument that we don't know whether or not such substitution is possible does not, in itself, give us any reason to doubt that there are such things as propositions which are the objects of belief and the other propositional attitudes. In this sense, Kripke has given no reason to think that the notion of a proposition is in danger of breaking down in cases like Pierre's and Jones'.

Here is another way of viewing the situation. My argument for the existence of things which are the "objects" of the propositional attitudes, weak as it may be, began with the assumption that sentences like "Jones believes that Cicero was bald" are sometimes true. What Kripke has done by throwing doubt on the disquotation principles and, thus, showing that we don't know whether

coreferential proper names are interchangeable, salva veritate, in belief contexts, is to show that it is not clear under what circumstances sentences like the one above are true. Is it true if Jones says (sincerely, etc.) "I believe that Cicero was bald"? Is it true if he says "I believe that Tully was bald but not that Cicero was bald"? Is it true if he points to a nearby statue of Cicero and says "I believe that he was bald"? We may grant that these are difficult and "vexing" questions. But to argue that it isn't clear under what circumstances sentences like "Jones believes that Cicero was bald" are true is not to argue that they are not at least sometimes true. If Kripke had argued this, his views would be quite a bit less popular. And, if they are sometimes true, then my argument, which makes no use of any disquotation principles and takes no position on questions of substitutivity, can proceed unaffected in any obvious way by Kripke's worries. As far as Kripke's problems are concerned, it doesn't matter whether we even try to introduce the term "proposition" in the first place. The questions concerning the substitutivity of coreferential proper names and the disquotation principles are extremely difficult, and remain.

Of course, this does not mean that there are no problems with my rather simpleminded way of introducing

the term "propositions." And, while I mentioned several potential difficulties earlier, an even more notorious problem remains. I have said nothing about how individual propositions are to be identified and distinguished from each other. In other words, I have not supplied a set of "identity conditions" for the notion. However, providing a satisfactory set of identity conditions has proved to be an extraordinarily difficult task, and I will not attempt to do so. Furthermore, while my introduction of the notion of a "proposition" is perhaps incomplete and inadequate in this sense, Kripke has not explained how his problems concerning Pierre might affect the search for a set of such conditions. So even if we interpret his claim that the notion of a "proposition" breaks down in cases like Pierre's and Jones' extremely generally to mean that something about those cases shows that it will be more difficult or impossible to provide a principle for individuating various propositions, it is still unclear how anything about Kripke's puzzle itself makes the notion of a "proposition" incoherent, or untenable in some other way.

Thus I conclude that the notion of a "proposition" which I have introduced is not "threatened" or "strained" in any obvious way by anything that Kripke infers from

his puzzle. However, while I think that he is incorrect in drawing this general moral from the puzzle, several of Kripke's other remarks suggest some interesting points about propositions. Before I go on to discuss what other philosophers have had to say about the puzzle, I want rather briefly to make some, I hope, not uninteresting points about my introduction of the term "proposition" and some of the uses which might be made of that notion.

III

Kripke says that just as any theory of truth must take the Liar paradox into account, so must any theory of belief take his puzzle into account. He also says, in the passage from which I quoted earlier, that the questions of what propositions are expressed by "simple" sentences containing proper names, and whether these propositions are "objects of belief" are vexing questions, especially when we consider puzzling Pierre. These remarks, and others, strongly suggest that the questions of whether there are such things as "propositions" and of whether these "propositions" are the "objects of belief" are distinct. Kripke's comments suggest that he thinks that there may be such things as propositions and yet these things might fail to be "objects of belief." This is where theories of belief would seem to be important. Apparently, Kripke believes that a theory of belief must, among other things, at least tell us whether propositions (assuming that they exist) are "objects of belief" and, if they are not, what are. I want now to argue that this way of thinking about propositions, though common, is misguided.

First of all, what is a theory of belief?[20] Of course, a theory of belief might simply be taken to be a set of sentences which, in some sense, are about belief.

In modern philosophical practice, however, theories of belief have usually taken the following form:

(T_B) John believes that p if and only if ____.

Here "p" is taken to be a schematic letter and is to be replaced with sentences of English. Different theories of belief will fill in the blank after the "if and only if" in different ways. For instance, one theory would be the following.

> (B_1) John believes that grass is green if and only if John gives his sincere, reflective (etc.) assent to "grass is green."[21]

This theory, which is similar to, but not identical with, the strong disquotation principle, is false. If John speaks only German, he will not give his sincere, reflective assent to "grass is green." Nevertheless, John, like many other people who speak languages other than English, might well believe that grass is green.

Here's another theory.

> (B_2) John believes that grass is green if and only if John bears some relation, call it "R," to an object which is the proposition that grass is green.

I think that this theory may well be true. If there are such things as propositions, then the relation R will just be that of belief, i.e., of <u>believing</u> the proposition that grass is green.

It often seems to be taken for granted that there can only be one true theory of belief. However, it is not at all clear that there are not several (perhaps even many) correct ways of filling in the blank in (T_B) resulting in several true theories of belief. To illustrate this, let me briefly discuss a debate which took place some years ago between Alonzo Church and Rudolph Carnap. In his article, "On Carnap's Analysis of Statements of Assertion and Belief," Church argues that it will not be possible to give a theory (or analysis) of belief that fills in the blank in (T_B) with some description of a relation between a person, Jones, and a sentence which replaces the schematic letter p. He says:

> For statements such as (1) Seneca said that man is a rational animal and (A) Columbus believed the world to be round, the most obvious analysis makes them statements about certain abstract entities which we call "propositions" . . . namely the proposition that man is a rational animal and the proposition that the world is round; and that these propositions are taken as having been respectively the object of an assertion by Seneca and the object of a belief by Columbus. We shall not discuss this obvious analysis here . . . our purpose is to point out what we believe may be an insuperable objection against alternative analyses that undertake to do away with propositions in favour of such more concrete things as sentences.[22]

In "On Belief Sentences--Reply to Alonzo Church," Carnap has the following to say in response:

> Church entertains the view that a belief must be construed as a relation between a person and a proposition, not a sentence, and that therefore only the first form, like (i) [John believes that the earth is round ,] is adequate, not the second, like (v) [John has the relation B to "the earth is round" as a sentence of English.] I do not reject the first form, but regard both forms as possible. I do not think that the arguments offered by Church so far show the impossibility of the second form.[23]

Thus Carnap grants that Church has discovered a problem with his analysis, but feels that it isn't an insurmountable difficulty. Carnap still thinks that an analysis (theory) of belief in terms of sentences is possible.[24]

What is interesting from the perspective of this paper is that Carnap explicitly states that he believes that both analyses (or theories), one in terms of propositions and one in terms of sentences, are possible. Furthermore, Church himself, while he does not think that a correct analysis of belief in terms of "such concrete things as sentences" is possible, gives no reason to doubt that there might be other true theories of belief. For example, later in this paper, I will discuss the view of Ruth Marcus that "belief is a relation between a person and a state of affairs." Perhaps Church

would find "states of affairs," whatever they may be, to be acceptable in a theory of belief. In contrast, most philosophers talk as if we have to make a choice. Either belief is a relation between a person and a proposition, or it is a relation between a person and a sentence, or it is a relation between a person and something else, perhaps a "state of affairs." But this need not be so. It may be possible to analyze belief as a relation between people and propositions and as a relation between people and sentences and perhaps even as a relation between people and states of affairs. It is misguided for philosophers, a priori, to conceive of their task as a search for "the way" in which a person is related to one particular kind of object. As Carnap quite perceptively points out, several correct analyses or theories of belief may be possible.

This isn't the only manner in which this common way of thinking about belief and propositions is confused. Another closely related misconception is the view that there may be such things as propositions and yet those "mysterious abstract entities" are not the "objects of belief." Kripke himself suggests that this view has some plausibility when he says that it is a "vexing question" whether propositions (assuming that they exist)

are the objects of belief. In order to see how ill conceived this view is, let's examine the expression "object of belief." It has been understood in at least two ways. Some consider something to be an object of belief just in case belief is correctly analyzed as a relation between a person and that object. For example, sentences have been called the true objects of belief (rather than propositions) by people who hold that it is possible to give a correct analysis of belief in terms of sentences. However, there is another natural way of understanding "object of belief." This is to take something to be an object of belief just in case it is believed (or perhaps, could be believed). This is analogous to maintaining that something is an object of my desire just in case I desire it. In this sense, something will be an object of my desire just in case it satisfies the open sentence "Jack Cobetto desires x." Of course, this is reminiscent of the way in which I introduced the term "proposition" in the first place. We called something a "proposition" just in case it is (or could be) believed, in the sense that it satisfied (or could satisfy) open sentences like "John believes x."

What is important to notice is that on either reading of "object of belief," it is obviously true that propositions (in my sense) are indeed objects of belief.

On the second reading, it is simply tautologous to say that propositions are objects of belief. Something is an object of belief in that sense if and only if it is (or could be) believed in the sense that it satisfies (or could satisfy) open sentences like "John believes x." But that is how we defined the term "proposition." As for the first reading of "object of belief," it is a simple matter to correctly analyze belief as a relation between a person and a proposition. Consider the following form of an analysis of belief.

> (BF) John believes that grass is green if and only if John bears R to the proposition that grass is green.

If R is simply taken to be the relation of belief, then the resulting analysis is clearly correct. It is true that John believes that grass is green if and only if John believes (bears R to) the proposition that grass is green. Perhaps this is why Church calls the analysis of belief in terms of propositions "obvious." Of course, while it may be obvious that propositions are objects of belief in this sense, it must be remembered that it is possible that other things, such as sentences, may also be objects of belief in this sense.

I would like to conclude this discussion with one further example of how the "theory" that propositions

are the objects of belief can be and has been [illegible] stood. It is widely held that the following "[illegible] tional attitude" sentences are ambiguous:

(85) John believes that the president is wise.

(86) John asserted that the president is wise.

(87) John knows that the president is wise.

If the definite description "the president" is read with small scope, the sentences receive what is called their "de dicto" interpretations. If it is read with large scope, the sentences receive their so-called "de re" interpretations. It is often asserted that the de dicto interpretations of these sentences assert or express a relation between a person and a proposition, i.e., the proposition that the president is wise.

I do not want to dispute this. However, the notion of a proposition is often used to explain the de dicto--de re distinction. For example, if a person, say Alfred, did not know anything about Russell's notion of scope and, as a perhaps not too ignorant layman, had not noticed that the propositional attitude sentences are ambiguous, it is thought that one could make him understand the ambiguity by appealing to the notion of a proposition. In fact, Kripke does something suggestive of this in the beginning pages of "A Puzzle About Belief." He says,

> The contrast, according to the Millian view, must come in the de dicto or "small scope" reading, which is the only reading, for belief contexts as well as modal contexts, that will concern us in this paper. If we wish, we can emphasize that this is our reading in various ways. Say, "It is necessary that: Cicero was bald" or, more explicitly, "The following proposition is necessarily true: Cicero was bald . . ."[25]

Kripke may be correct in holding that "John believes the proposition: the president is wise" emphasizes the de dicto reading of (85). However, appealing to the notion of a proposition (in my sense) will not help a person, like Alfred, who doesn't already grasp the ambiguities in (85)-(87) to understand them. Such an appeal will not enable Alfred to recognize the kinds of circumstances in which each of the readings of (85)-(87) is true. To see this, suppose that we inform Alfred that the de dicto interpretation of (85) is true if and only if John believes the proposition: that the president is wise. Since Alfred, we may suppose, does not know how the term "proposition" is being used, he will ask for an explanation. We will then inform him of the manner in which we introduced and defined the term. However, since we inferred the existence of propositions from the truth of sentences like (85)-(87), the explanation we are giving to Alfred employs the very kinds of sentences about which he is unclear. If he is alert, he will respond that he isn't

sure that he correctly understands some of the sentences used in our explanation, i.e., (85)-(87).

Thus the notion of a proposition, in my sense, can not be used in the above manner to explain the de dicto-de re distinction. Again, this doesn't mean that the sentence "John believes the proposition that the president is wise" can not be used to emphasize or suggest the de dicto interpretation of (85). Furthermore, this doesn't rule out the possibility that some other notion of a proposition might enable a person to grasp the de dicto-de re ambiguity. Finally, I don't intend to assert that my notion of a proposition and the claim that propositions are objects of belief possess no explanatory power. My only point is that it is not clear what uses (if there are any) it might have.

IV

Several philosophers have attempted to solve Kripke's puzzle. In particular, Tom McKay, Igal Kvart, William Lycan, and Ruth Marcus have all proposed different theories of belief in their attempts to resolve the paradox. McKay takes belief to be a relation between a person and a proposition. Both Kvart and Lycan offer (distinct) theories which analyze belief as a relation between a person and a particular sentence. Marcus claims that belief is an "epistemological attitude towards" a "state of affairs." However, if what I've said in the preceding section of this paper is correct, it is not clear that these "theories" are incompatible. Furthermore, while all of the solutions which I will discuss advocate <u>different</u> theories of belief, it is interesting that they all reject the <u>same</u> premises of Kripke's paradox. With this in mind, let's examine these purported solutions.

In his article "On Proper Names in Belief Ascriptions," McKay offers a solution which is nearly identical to mine.[26] He begins his paper with a discussion of indexical expressions such as "I," "you," "he," "this," and "that." After examining the ways in which these expressions are used in various contexts, he argues for the same distinction insisted upon by Cartwright, Strawson,

Lemmon, and (says McKay) Kaplan. He concludes that we must distinguish what is asserted by uttering a sentence in a particular context from the meaning of the sentence which was uttered. McKay uses the term "sentential meaning" for the meaning of a sentence and the term "content" for what I have called a "proposition," i.e., what is asserted.

McKay then goes on to point out that, interestingly, two sentences can have the same sentential meaning and can be used to assert or indicate belief in the same proposition even though a person who understands both sentences may not realize this. For instance, "Londres est jolie" and "London is pretty" have the same meaning and can be used to make the same assertion or to indicate a belief in the same proposition, i.e., that London is pretty. But Pierre doesn't realize this. He believes that London is pretty, yet he is not willing to indicate his belief in that proposition by assenting to "London is pretty."

Thus McKay contends that the strong disquotation principle is false, and, since Pierre also believes that London is not pretty, that Pierre has contradictory beliefs. In the concluding paragraph of his paper, McKay states,

> Language is a great resource for entertaining propositions, yet in its very virtue, that this can lead us to new beliefs, lies our problem. Language can lead us to contradictory beliefs without providing the immediate resources for removing the contradiction. Our grasp of a name on a particular occasion can be a sufficient basis for a belief about its referent and yet be an insufficient basis for determining all of the co-reference relations of that use with other uses of names that we grasp.[27]

Clearly McKay is rejecting the premise of the puzzle (Principle C) that states that if a person has contradictory beliefs, then he is, in principle, in a position to tell through the use of logic alone that he has contradictory beliefs. Thus McKay solves both the stronger version and the weaker version of the puzzle in the same way that I solve them. Both the strong disquotation principle and principle C must be rejected.

However, while McKay and I agree about which premises of the puzzle must be rejected, I want to mention several of McKay's positions with which I do not agree. First, McKay fully accepts and defends the "direct reference" theory of names which states that a proper name "has no connotation, but refers directly, without any semantic contribution from associated properties." Thus he accepts the view that two people uttering the sentences "Cicero was bald" and "Tully was bald" make assertions that have the same content (in my terminology,

that "that Cicero was bald" and "that Tully was bald" denote the same proposition). He accepts the view that coreferential proper names are indeed interchangeable, salva veritate, in belief contexts. I want to emphasize that I do not take this strong position and, furthermore, that it is not necessary to take a position on this question to solve the puzzle. Solving the puzzle in the way in which both McKay and I advocate (rejecting the strong disquotation principle) may indeed considerably weaken the most influential argument purporting to refute the "direct reference" theory of names. But that does mean that there are no problems with the theory and that we ought to be in a hurry to embrace it.

Also, McKay seems to misunderstand the general significance of the puzzle. He says,

> And Saul Kripke, the leading promoter of the view that names lack connotation, has found the direct reference view contributing to a puzzle about belief.[28]

and also,

> In "A Puzzle About Belief" Kripke considers a slightly different argument against the conclusion that proper names are devices of pure reference in belief contexts.[29]

In fact, this simply isn't the way in which Kripke views the puzzle. He does not consider it to pose a problem for the "direct reference" theory of proper names. Indeed,

as I argued earlier, Kripke believes that the puzzle supports the "direct reference" theory. He take the puzzle to show that a widely (almost universally) accepted view about belief, which is seemingly inconsistent with the "direct reference" theory, is unjustified. The discovery of the puzzle, Kripke says repeatedly, shows that the simple conclusion that coreferential proper names are not interchangeable, salva veritate, in belief contexts, is "unwarranted." These alleged failures of substitutivity have sometimes been taken to establish or at least to suggest that the "direct reference" theory is incorrect. Kripke's argument rejects the threatening conclusion about substitutivity and, therefore, purports to support the "direct reference" theory of proper names.

The solutions proposed by Igal Kvart and William Lycan both analyze belief as a relation between a person and a sentence. Here is an excerpt from Kvart's article "Kripke's Belief Puzzle."

> I shall conduct my discussion against a set-up in which the notion of 'p' being a belief of r (for a believer r) will play a major role. My main working hypothesis will involve taking beliefs as linguistic representations. At a given time t a cognizer can be said to be in a certain belief-state, reflecting whatever beliefs he has at the time. Such belief-states would be classifiable via particular beliefs: if 'p' is a belief of our cognizer r (as I shall call

> our believer from now on) at time t, then r could be said to be in a 'p'-belief state. (I will remind the reader that I limit my discussion in this paper to cognizers with minimal logical, linguistic and conceptual acumen).
>
> Of course, having particular beliefs is a constitutent of the causal order: the acquisition of beliefs can be caused by various stimuli, and r's possession of certain beliefs can cause modes and dispositions of behavior, not the least among them would be his verbal dispositions to assert or to assent to various sentences, or to refrain from such assents. Beliefs, being linguistic representations, that is, sentences, are in one language or another: a cognizer r may have 'the king is bald,' but not 'le roi est chauve,' as a belief of his, which will be attested to by his disposition to assent to the first but not to the second in 'appropriate' circumstances (which will happen, for instance, when he knows no French). If he would assent to both in 'appropriate' circumstances, they would constitute two distinct beliefs of his. The framework I am working with here should be contrasted with the theory which takes beliefs to be propositions.[30]

Although Kvart makes some other interesting claims about belief, the passage above provides sufficient background to enable one to understand his solution to Kripke's puzzle.

Here is the analysis of belief that Kvart proposes:

> (KB) r believes that p if and only if for some "q," such that "p" is an adequate paraphrase of "q," "q" is of belief of r.

In order to facilitate a discussion of this analysis, let me replace the above (intended) schema with the following

instance which is of special interest to us.

> (KB^1) Pierre believes that London is pretty if and only if for some sentence p such that "London is pretty" is an adequate paraphrase of p, sentence p is a belief of Pierre.

Before this analysis can be understood, we must know both what Kvart means by the phrase "is an adequate paraphrase of" and when a sentence is a belief of Pierre. Since Kvart claims that "verbal disposition to assent or not to assent to certain sentences are thus key indicators for the latter being beliefs," it is plausible to suppose that Kvart maintains that the sentence "London is pretty" is a belief of Pierre if and only if Pierre assents to it. Unfortunately, Kvart is even less clear about what he means by "is an adequate paraphrase of." However, since he does consider "Londres est jolie" to be an adequate paraphrase of "London is pretty" and says also that every sentence is "of course" an adequate paraphrase of itself, it is natural to assume that two sentences are "adequate paraphrases" of each other just in case they can be used to assert the same proposition. Fortunately, we need not settle this because the two assertions just mentioned make it possible to see how Kvart's solution works.

Let's examine how Kvart's analysis handles two relatively straightforward and uncontroversial cases. First, suppose that John is a normal English speaker who speaks no other language and who assents to the sentence "London is pretty." On Kvart's view, we can infer that "London is pretty" is a belief of John. Moreover, since there is some sentence which is an adequate paraphrase of "London is pretty" (that sentence itself) and which is a belief of John, it follows that John believes that London is pretty. This result seems to be correct. Now suppose that John speaks only French and that he assents to the French sentence "Londres est jolie." Since there is some sentence p which is an adequate paraphrase of "London is pretty" (namely "Londres est jolie") and which is also a belief of John, it follows, again, that John believes that London is pretty.

Thus Kvart's analysis of belief gives the correct answer to the question "Does John believe that London is pretty?" in these two unproblematic cases. But what does it say about the more complicated case of Pierre? Kvart says,

> In a nutshell, my resolution of Kripke's puzzle is as follows: The consistency of a believer resides in his sorting out purely logical contradiction from among his beliefs, given unlimited logical acumen, as Kripke would agree. 'Londres est jolie' and 'London is [not] pretty' are two distinct beliefs, [an obvious mistatement] and no logical acumen can tell that they are contradictory. (This is not the case if the beliefs are taken to be propositions rather than sentences.) Furthermore: it is possible for 'Londres est jolie' to be a belief of r without 'London is pretty' being a belief of his (despite his logical and linguistic acumen). Given our analysis of belief-sentences in terms of beliefs, in this case 'r believes that London is pretty' will be true without qualification, since some adequate paraphrase of 'p' being a belief of r is sufficient to make 'r believes that p' true, while no one adequate paraphrase of 'p' which is not a belief of r is sufficient to make 'r believes that p' false. Hence non-assent to 'p' does not entail that r does not believe that p, which makes Kripke's Biconditional Form of the disquotational Principle false. Without this principle, no contradiction follows.[31]

In this passage, Kvart is rejecting the strong disquotation principle. He claims that while Pierre does not assent to "London is pretty," Pierre still believes that London is pretty because he does assent to "Londres est jolie" which is both an "adequate paraphrase" of "London is pretty" and is one of his beliefs. Thus Kvart, as I did, is simply citing Pierre's case as a counterexample to the claim that if a normal English speaker believes that London is pretty, then he will assent to the sentence "London is pretty."

Kvart's solution to the weaker version of the puzzle is a bit more problematic. He agrees with me that Pierre believes both that London is pretty and that London is not pretty since Pierre assents to and has as beliefs both "Londres est jolie" and "London is not pretty," which are adequate paraphrases of "London is pretty" and "London is not pretty." But despite this, Kvart maintains that Pierre does not have contradictory beliefs! He says,

> Now the notion of a believer being consistent or not resides in whether his beliefs yield a contradiction. It resides in whether the use of logical tools alone can allow him to derive a contradiction from the beliefs he possesses, which is tantamount to whether he has contradictory beliefs. Thus, the subject matter of his being consistent or not lies at the level of his beliefs, in whether there is a set of beliefs of his such as {'p', '-p'}; or, more generally, a set of beliefs of his {'p_1', . . . , 'p_n'} which is self-contradictory.[32]

Kvart seems to think that a person has inconsistent beliefs if and only if he can tell by using logical tools alone that those beliefs (i.e. sentences) lead to a contradiction. While Pierre has the two sentences "Londres est jolie" and "London is not pretty" as beliefs, he could not logically infer the sentence "London is pretty" from his "belief-set" of sentences. Thus he can not determine by logical means that he has inconsistent beliefs.

This account of inconsistency is troubling in several respects. Suppose that a person, Peter, has the

sentences "London is pretty" and "London is not pretty" as beliefs. Can he tell "through the use of logic alone" that these sentences (beliefs) are contradictory? If he judges them to be contradictory then he is obviously assuming that "London" refers to the same city in both sentences. (If he thought it referred to different cities, he wouldn't judge the sentences to be inconsistent.) But isn't the assumption that "London" refers to the same city in both sentences, in some sense, "extra-logical"? After all, logic does not tell Peter that it refers to the same city in both sentences. Now suppose that the person in question is Pierre and the two suspect beliefs are "Londres est jolie" and "London is not pretty." Kvart says that if Pierre is to derive a contradiction from these two sentences, he must make the "extra-logical" assumption that "Londres" and "London" refer to the same city. But if this assumption is "extra-logical," why isnt' Peter's assumption that "London" refers to the same city in both of his beliefs also "extra-logical"? It just is not very clear, on Kvart's account of inconsistency, when we are entitled to conclude that a person has deduced or recognized a contradiction "on the basis of logic alone."

More importantly, even if Kvart is correct in claiming that Pierre's beliefs "Londres est jolie" and

"London is not pretty" are consistent in the sense that Pierre can't derive a contradiction from them through logic alone, there is still an important sense in which Pierre has inconsistent beliefs. Kvart's analysis of belief entails that Pierre believes that London is pretty and also that Pierre believes that London is not pretty. When Kvart argues that, in spite of this, Pierre does not have contradictory beliefs, I am strongly tempted to respond that when I say that a person has contradictory beliefs, I simply mean that, for some sentence p, that person believes that p and also believes that not p. In other words, I define "has contradictory beliefs" not in terms of a person's behavior with respect to particular sentences in his or her "belief-set" (as Kvart does), but rather in terms of sentences like "Pierre believes that London is pretty" and "Pierre believes that London is not pretty." While Pierre may not have contradictory beliefs in Kvart's sense, he clearly has contradictory beliefs in my sense.

I will not here discuss the questions of whether other senses of "inconsistent" and "contradictory" exist (or might be introduced) and of whether any of these senses are (or would be) preferable. If I hesitate to adopt Kvart's definitions of "has inconsistent beliefs"

and "has contradictory beliefs," it is because they appeal to his analysis of belief in terms of particular sentences. In contrast, my definitions of these expressions make no appeal to any theory or analysis of belief. While this seems to me to be a more natural and preferable approach, it is not important for the purposes of this paper that we choose one or the other.

However, it is important that our difference in terminology does not obscure that Kvart and I essentially agree on how Kripke's puzzle, in both its stronger and weaker forms, is to be solved. We both solve the stronger version by rejecting the strong disquotation principle. As regards the weaker version, we both claim that Pierre believes both that London is pretty and that London is not pretty. If one uses "has contradictory beliefs" in Kvart's sense, then Pierre does not have contradictory beliefs, and the premise of the weaker version of the puzzle which claims that he does must be rejected. On the other hand, if one uses "has contradictory beliefs" in my sense (which is also, I think, the sense intended by Kripke), then Pierre does have contradictory beliefs, and the puzzle is solved by realizing that a person can have contradictory beliefs without being in a position to determine this through the use of logic alone.

The other "sentential" solution to the puzzle is proposed by William Lycan. I won't try to describe his complicated theory of belief in all of its detail. Indeed his article, "Toward a Homuncular Theory of Believing," contains many expressions and claims that cry out for explanation and elucidation.[33] Instead, I'll include the following passage in which he lays out, in broad outline, his view:

> Homunctionally: To judge or believe occurrently that P is to have a storage-and-playback mechanism that in a certain distinctive way harbors a representation whose syntactic/semantic structure is analogous to that of the sentence that replaces "P." What makes such a state of affairs a case of believing <u>that P</u> is the syntactic/semantic properties the representation shares with the (here, English) sentence in question; what makes it a case of <u>believing</u> that P is the distinctive mode of "harboring" or storage-playback. This distinctive mode of storage is what we might otherwise call the <u>type of functional role</u> played by a belief <u>qua</u> belief--the characteristic contribution that a "believed" representation makes to the believer's ongoing institutional order of business. Thus, it is this mode of storage or type of functional role that will distinguish beliefs from other propositional attitudes; and a full specification of the mode of storage would contain parameters whose values would determine such interesting features as belief <u>strength</u> (I take belief strength to be a matter of the belief's use in explanatory inference, the amount and type of causal sustenance that it receives from its basing reason(s), and its authority in interacting with other beliefs and desires in determining action).[34]

If this were all that Lycan said about his theory, it would obviously be quite difficult to extract a clear analysis of belief sentences. Fortunatley, Lycan adopts a "semantics for belief-ascription" offered by other philosophers which is a bit more manageable. Here is the essential passage.

> The sentential account squares with a plausible semantics for belief-ascriptions. Sellars, Hill, and also Davidson (1968) have argued that the sentential complement of a belief-ascription serves as a sort of exemplar of what is said to be believed, the semantical function of the complementizer "that" being to ostend or demonstrate this exemplar. Thus:
>
> 1. Jones believes that broccoli causes erysipelas
>
> is to be understood along the lines of
>
> 2. Jones believes some •Broccoli causes erysipelas•,
>
> where the Sellarsian dot quotes are common-noun-forming operators that also serve to ostend the linquistic token that they enclose. A slight variation would be to express the force of (1) as
>
> 3. Jones believes one of those Broccoli causes erysipelas.
>
> Thus, in this approach, belief is construed as a dyadic relation that a person bears to a linguistic or quasi-linguistic token that falls into a certain category. I shall not here rehearse the virtues of this semantical hypothesis, as its proponents have already touted them at length.
>
> Now, how are we to determine the extension of the predicate "is a •Broccoli causes ersipelas•"? Alternatively, how are we to tell when some linguistic or quasi-linguistic token of some quite other shape is "one of those"?

> Differing individuative schemes are possible here. Davidson merely invokes an unexplicated relation of "samesaying." Sellars (1963) offers a more elaborate scheme: For him, an item will count as a ·Broccoli causes erysipelas· just in case that item plays approximately the same inferential role within its own surrounding conceptual framework that the sentence "Broccoli causes erysipelas" plays within ours. Other possibilities are available. We might count a thing as a ·Broccoli causes erysipelas· if the thing has the same truth-condition as does our sentence "Broccoli causes erysipelas," or if the thing has the same truth-condition computed according to the same recursive procedure. Later on I shall make special use of this availability of alternative methods of individuation.[35]

Thus Lycan's analysis of belief is that

> (L_B) Jones believes that broccoli causes erysipelas if and only if Jones believes some token of ·Broccoli causes erysipelas·.

In order to properly understand this theory of belief, one must take note of the suprising way in which Lycan uses the expression "·Broccoli causes erysipelas·." While it is tempting to assume that this expression is being used to denote or name a sentence, Lycan is really using it as a predicate expression. Moreover, this expression may have in its extension some sentence tokens which are <u>not</u> tokens of the sentence "Broccoli causes erysipelas." (T_1) and (T_2) are the two definitions of the predicate which Lycan offers in the passage above.

> (T_1) (x)(x <u>is a token of ·Broccoli causes erysipelas·</u> if and only if x has the same truth-conditions as "Broccoli causes erysipelas").

> (T_2) (x)(x is a token of ·Broccoli causes erysipelas· if and only if x plays the same inferential role as "Broccoli causes erysipelas").

Interestingly, Lycan claims that (T_1) and (T_2) correspond to an ambiguity in belief sentences. Moreover, he claims that Kripke's puzzle is solved once this ambiguity is recognized.

However, before I evaluate this claim, I want to say a word about (T_1) and (T_2). (T_1) is relatively unproblematic. There is a long tradition of using the term "truth conditions" in the following manner. Two sentences have the same truth conditions just in case they attribute the same property or properties to the same object or objects. For example, "Cicero was bald" and "Tully was bald" have the same truth conditions because they both attribute the same property (baldness) to the same object (Cicero, i.e., Tully). Furthermore, Lycan's use of "truth conditions" appears to be consistent with this tradition.

In contrast, there is no long tradition of use surrounding the expression "plays the same inferential role as." It would seem that to talk about the "inferential role" of a sentence would be to talk about the inferences which different people might make from that sentence. For instance, from the sentence "John is a

lonely bachelor," we might infer the sentence "John is lonely" or the sentence "John is not married." However, we would not infer the sentence "John is hungry" from any of the above sentences. Thus we might conclude that "John is hungry" plays a different "inferential role" than the other sentences. Unfortunately, while this example may serve to illustrate what Lycan intends, Lycan does not offer a general definition of the term. With this in mind, let's now consider how Lycan's theory is supposed to solve Kripke's puzzle.[36]

Lycan says that there should be no "unequivocal" answer to Kripke's question "Does Pierre believe that London is pretty?" He says that the question "has a strong" yes and no "feel to it" and that a solution to the puzzle must explain this "yes and no" feeling. Lycan argues that the ambiguity of belief sentences which he posits explains this feeling. Consider (L_1) and (L_2), which are obtained by combining (T_1) and (T_2) with Lycan's original analysis of belief sentences, (L_B).

> (L_1) Pierre believes$_1$ that London is pretty if and only if he believes some sentence token with the same truth conditions as "London is pretty."
>
> (L_2) Pierre believes$_2$ that London is pretty if and only if he believes some sentence token which plays the same inferential role as "London is pretty.[37]

Since Pierre believes the sentence "Londres est jolie" and it has the same truth conditions as "London is pretty," (L_1) entails that Pierre believes$_1$ that London is pretty. In contrast, while Pierre believes the sentence "Londres est jolie," Lycan claims that neither that sentence nor any other sentence which Pierre believes plays the same inferential role for Pierre as "London is pretty." Consequently, if Lycan is correct about this, (L_2) entails that Pierre does not believe$_2$ that London is pretty. Thus the alleged "yes and no" feeling to Kripke's question is explained by the fact that the answer is "yes" for the first sense of believe and "no" for the second sense of belief.

Nevertheless, this purported solution to the puzzle is inadequate. To see this, suppose that Kripke had asked the question "Does Pierre believe that London is not pretty?" Clearly Kripke might have done this. After all, the whole point in telling Pierre's elaborate story is to give us apparently equal reason to assert either "that Pierre believes" that London is pretty or "that Pierre believes" that London is not pretty. Accordingly, if the question "Does Pierre believe that London is pretty?" has a "yes and no" feeling which must be explained then so does the question "Does Pierre believe that London is not pretty?"

However, Lycan's two senses of belief can not explain the "yes and no" feeling of the second question. (L_1) forces us to conclude that Pierre does believe$_1$ that London is not pretty because he believes a sentence that has the same truth conditions as "London is not pretty." In fact, he believes that very sentence. Furthermore, although the term "inferential role" is not fully explained, it is obvious that a token of "London is not pretty" plays the same inferential role as itself. Therefore, since Pierre believes the sentence "London is not pretty," (L_2) entails that Pierre also belives$_2$ that London is not pretty. But this leaves the "yes and no" feeling of the question "Does Pierre believe that London is not pretty?" unexplained, because the answer is "yes" on either sense of "belief."

It is interesting to note that, while Lycan's solution clearly does not explain the "yes and no feeling" to both of Kripke's questions, there is a sense in which my solution does explain it. In my view, Pierre believes both that London is pretty and that London is not pretty. Accordingly, if one is asked whether Pierre believes that London is pretty, it would be very misleading to say simply "yes." While the resonse may, strictly speaking, be true, it would have an air of deception in that "the whole truth" was not communicated.

Similarly, if one responded to the question "Does Pierre believe that London is not pretty?" with a simple "yes," an analogous air of deception would arise. While this explanation may not be as elegant as a solution which posits an ambiguity in the term "believes," at least it, unlike Lycan's solution, explains the "yes and no" feeling to both of Kripke's questions.

Furthermore, if it is not clear that my solution to Kripke's puzzle succeeds in explaining the "yes and no" feeling of Kripke's questions, the problem may well lie not in my explanation, but in what is being explained. The expression "the yes and no feeling" is extremely vague. Lycan does not explain it. He merely introduces it and then claims that his solution to the puzzle succeeds in explaining it. Thus until we are given a better idea of just what it is we are to explain, it would be more productive to evaluate solutions to Kripke's puzzle on the basis of whether they succeed in resolving the specific contradictions of the puzzle. I want now to show that neither of Lycan's analyses of belief (L_1) and (L_2), can, in itself, produce an adequate solution to the puzzle.

(L_1) seems to entail a solution to the puzzle which is very similar to mine. As we saw above, (L_1) entails that Pierre believes both that London is pretty and that

London is not pretty. Consequently, (L_1) requires that we give up principle (C) which states that a person who has contradictory beliefs is in principle in a position to tell through the use of logic alone that this is so. Also, (L_1) entails that the strong disquotation principle is false. Pierre believes that London is pretty even though he doesn't assent to the sentence "London is pretty." Instead he believes and assents to the sentence "Londres est jolie" which has the same truth conditions as "London is pretty."

However, in spite of this similarity, (L_1) has other consequences which render it unsuitable to serve as a basis of a solution to Kripke's puzzle. Consider my earlier example about Jones and the sentences "Cicero was bald" and "Tully was bald." Since these two sentences have the same truth conditions, (L_1) entails that, if Jones believes that Cicero was bald, he also believes that Tully was bald. Moreover, since two sentences have the same truth conditions if they attribute the same property or properties to the same object or objects regardless of the manner in which the objects are designated, (L_1) entails that coreferential expressions are interchangeable, salva veritate, in belief contexts. I have argued that, even in light of Kripke's puzzle, we should not rush to conclude that coreferential proper

names are interchangeable, salva veritate, in belief contexts. However, the view that any coreferential expressions are interchangeable, salva veritate, in belief contexts would be rejected by most philosophers, even those who are ardent supporters of the "direct reference" theory of proper names.

Lycan's other analysis of belief is just as implausible. As we have seen, (L_2) entails both that Pierre believes that London is not pretty and also that Pierre does not believe that London is pretty. But to accept this would be to respect Pierre's behavior as an English speaker (he assents to "London is not pretty") and to reject his behavior as a French speaker (he assents to "Londres est jolie.") While I will not repeat his argument here, Kripke argues powerfully that Pierre's behavior as a French speaker and Pierre's behavior as an English speaker deserve equal respect. That is, there is no more reason to conclude that Pierre believes that London is not pretty than there is to conclude that he believes that London is pretty. In so far as a proponent of (L_2), without giving any explanation, only respects Pierre's behavior as an English speaker, he has failed to resolve Kripke's puzzle.

Here's another way of stating this objection. Surely, if (L_2) is true, there will be an analogous

theory, stated in French, which is also true. We can obtain this theory, call it "(L_2F)," by translating (L_2) into French but replacing "London is pretty" with "Londres est jolie." (This is just what Kripke does by insisting that a "French" principle analogous to the strong disquotation principle will be true if the English principle is true.) If we translate (L_2F) into English, we get that Pierre believes that London is pretty if and only if he believes some sentence token that plays the same inferential role as "Londres est jolie." Since Pierre believes the sentence "Londres est jolie" and it has the same inferential role as "Londres est jolie" (i.e., itself), it follows that Pierre does believe that London is pretty. But this contradicts the consequence of (L_2) that Pierre does not believe that London is pretty. Therefore, if Lycan wants to escape from the contradiction, he must, given that he accepts (L_2), hold that (L_2F) is false. But this is absurd. Surely if (L_2) is true, then the analogous French principle is also true. English has no such special position in the universe of languages.

In "A Proposed Solution to a Puzzle About Belief," Ruth Marcus offers a solution which is distinctive in several respects.[38] First, she takes the rather uncommon position that belief "is a relation between a person and

a state of affairs." Secondly, she maintains a very controversial view which entails, for example, that the sentence "John believes that 2 + 2 = 5" is never true, regardless of to whom "John" refers and how mathematially inept that person may be. Since this second position is directly related to one of the premises of Kripke's puzzle, I'll begin with it.

Marcus argues that belief and knowledge are related in an interesting and suprising way. She begins by pointing out that if a person claims that he or she knows that Cicero $\neq$ Tully, that person is mistaken. This is not controversial, since, in general, one can't know something that is false. However, Marcus goes on to claim that just as someone who claims to know something that is false is mistaken, anyone who claims to believe something that is impossible is also mistaken. Here is the passage containing her intuitive argument for that conclusion.

> There is an intuition about belief which I have (as do others) but which is not so widely shared. That intuition suggests a modification of the disquotation principle. Suppose that someone were to claim that he believes Hesperus is not identical with Phosphorus or that Tully is not identical with Cicero, or that Londres is not the same as London where in those contexts of use the names of the "pairs" in question do, on the theory of direct reference, refer to the same thing. It is my (non post-hoc)

> intuition that on discovery that those identities hold, and consequently that the associated name pairs name the same thing, I would not say that I had changed my belief or acquired a new belief to replace the old, but that I was mistaken in claiming that I had those beliefs to begin with. After all, if I had believed that Tully is not identical with Cicero, I would have been believing that something is not the same as itself and I surely did not believe that, a blatant impossibility, so I was mistaken in claiming to have the belief. Nor am I insisting that I did not have any belief, but only that it was not the belief that Hesperus is not the same as Phosphorus, that Tully is not the same as Cicero, that Londres is not the same as London. 39

Needless to say, Marcus has not won many converts. The consequences of the view that one can't believe a necessary falsehood are hard to swallow. For example, if Marcus is correct, no mathematician ever really believed any of the false axioms or conjectures that have been proposed over the years (since mathematical falsehoods are necessarily false). However, I will not discuss further whether Marcus is right about this. Instead, I want to consider how this view affects Kripke's puzzle.

Marcus states above that her intuitive arguments suggest that the weaker disquotation principle in Kripke's puzzle has to be modified. Here is what she suggests.

> The analogy between this intuition about belief claims and the more universally accepted ones about knowledge is close. Just as a condition for knowing that p is that p obtains, so a condition for believing is

C. If x believes that p, then possible p.

The link between belief and possibility also suggests a modification of the disquotation principle A as follows.

D. Again assuming that assent is sincere and reflective, if (i) a normal English speaker assents to 'p' and (ii) 'p' is a sentence of English and (iii) p is possible, then he believes that p.

It follows from C and D given all the assumptions, that

E. If a speaker assents to 'p,' then he believes that p if and only if p is possible.[40]

Interestingly, these modifications do not affect the puzzle. To see this, suppose that Marcus' modified principle E is true. Since Pierre assents to the sentence "Londres est jolie" and this sentence expresses a proposition which is true in some possible world, it follows that Pierre believes that London is pretty. Similarly, since Pierre assents to "London is not pretty" and, in some possible world, London is not pretty, it follows that Pierre believes that London is not pretty. But this is the same situation we were in before we used Kripke's ("unmodified") disquotation principle.

In fact, Marcus' solution to the puzzle is the same as the solution which I and several others have given. That is, she rejects the same two premises of the puzzle. Regarding the weaker version of the puzzle,

her modified disquotation principle E entails that Pierre has contradictory beliefs. Hence, since Pierre can't tell that he does have contradictory beliefs through the use of logic alone, principle (C) must be given up. Regarding the stronger version of the puzzle, Marcus, too, says that the strong disquotation principle is to be rejected. The fact that some English speaker believes that London is pretty simply does not entail that that speaker will assent to "London is pretty." Marcus, just as I do, explicitly cites Pierre as a counter-example to this principle.[41]

Let's turn to Marcus' other rather uncommon view. Here is the passage in which she introduces her theory that belief "is a relation between a person and a state of affairs."

> Knowing and believing have been characterized as "propositional attitudes." The vagaries of the many uses of 'proposition' have been a considerable source of epistemological confusion. There is a seemingly naive as well as much maligned view, to which I subscribe, Russell's for example, where knowing and believing are attitudes towards states of affairs (not necessarily actual), which may have individuals and attributes as constituents. The "propositional content" of a sentence on an occasion of use is (are) the (those) state(s) of affairs that would make that sentence true. States of affairs may be actual, not actual, possible, necessary, even impossible.[42]

Notice that nothing in this passage rules out the possibility that states of affairs are themselves propositions (in my sense of "proposition"). Indeed, the passage suggests that Marcus does believe that states of affairs satisfy (or could satisfy) open sentences like "John believes x" and "John knows x." If so, then states of affairs are propositions.

While this may seem rather innocuous, if one is not careful, one could be led into difficulties. Marcus says that states of affairs have actual individuals and attributes as constituents. Thus the states of affairs denoted by "that Cicero was bald" and "that Tully was bald" have the actual man, Cicero, as a constituent. Now some might be tempted to suppose that the state of affairs denoted by "that the greatest Roman orator was bald" has Cicero as a constituent as well. If so, it is also very tempting to infer that all three of these "that" clauses denote the same state of affairs, because these "that" clauses don't seem to differ in any other significant respect (they all attribute baldness to Cicero). However, since the "x" in "John believes x" is transparent (in the sense I explained earlier), it follows that (88),

(89), and (90), have the same truth value.[43]

(88) John believes that Cicero was bald.

(89) John believes that Tully was bald.

(90) John believes that the greatest Roman orator was bald.

And, if we take the objects referred to by all denoting expressions in "that" clauses to be constituents of the states ofaffairs denoted by the "that" clauses, then all coreferential expressions will be interchangeable, salva veritate, in belief contexts. Of course, these consequences would not be accepted by most philosophers.

There are several possible responses. First, given that Marcus is such a strong supporter of the "direct reference" theory of proper names, she would probably be willing to assert, like McKay, that "Cicero" and "Tully" are interchangeable, salva veritate, in belief contexts. She just wouldn't regard this consequence as a difficulty for her theory of belief. However, even Marcus would probably not be willing to assert that all coreferential expressions, including definite descriptions, are interchangeable, salva veritate, in belief contexts. As we saw, this probelm arises only if the individuals referred to by definite descriptions in "that" clauses are taken to be constituents of the states of affairs that the "that" clauses denote. Interestingly, it was

Russell, the very philosopher from whom Marcus derives her view that states of affairs are objects of belief, who showed how to avoid this position. Briefly, he asserted that the states of affairs (propositions) denoted by "that" clauses containing definite descriptions have as constituents not the denotations of the descriptions, but rather certain "propositional functions." Fortunately, Russell's views about definite descriptions are so well-known that it is not necessary to repeat them here.

Appendix

As I indicated briefly at the beginning of this paper, the views which Kripke advocates in "A Puzzle about Belief" are intimately related to his neo-Millian view that proper names refer "directly" to their referents and are not synonymous with "associated" definite descriptions. Some philosophers have argued that this view entails the obvious falsehood that coreferential proper names are interchangeable, salva veritate, in belief contexts. Although Kripke denies that this is entailed by his views, his main purpose in "A Puzzle about Belief" is to show that it just is not clear that coreferential proper names are not interchangeable, salva veritate, in belief contexts. In this appendix, I want to examine Kripke's claim that proper names are not synonymous with "associated" definite descriptions. As is well known, his argument for this view centers on a difference in the manner in which definite descriptions and proper names designate their referents. Proper names, says Kripke, designate their referents "rigidly," whereas, in general, definite descriptions designate their referents "accidently" or "nonrigidly." Although I will explain and make some, I hope, interesting points about the rigidity argument, my main purpose is to point out another difference in the manner in which names and some definite descriptions designate their referents; and I will consider

some of the advantages and disadvantages in using this difference to establish the non-synonymy of names and "associated" definite descriptions.

Precisely how does Kripke's argument that names are not synonymous with "associated" definite descriptions go? As I've said above, it has something to do with the alleged fact that names are "rigid" designators while definite descriptions, typically, are "nonrigid" designators. Here is the passage in Naming and Necessity in which Kripke first introduces the concept of rigidity.

> I wish at this point to introduce something which I need in the methodology of discussing the theory of names that I'm talking about. We need the notion 'identity across possible worlds' as it's usually and, as I think, somewhat misleadingly called to explicate one distinction between asking whether it's necessary that 9 is greater than 7 or whether it's necessary that the number of planets is greater than 7? Why does one show anything more about essence than the other? The answer to this might be intuitively "Well, look, the number of planets might have been different from what it in fact is. It doesn't make any sense, though, to say that nine might have been different from what it in fact is.' Let's use some terms quasi-technically. Let's call something a rigid designator if in every possible world it designates the same object, a nonrigid or accidental designator if that is not the case. Of course we don't require that the objects exist in all possible worlds. Certainly Nixon might not have existed if his parents had not gotten married, in the normal course of things. When we think of a property as essential to an object we usually mean that it is true of that object in any case where it would have existed. A rigid designator of a necessary existent can be called strongly rigid.

> One of the intuitive theses I will maintain in these talks is that names are rigid designators. Certainly they seem to satisfy the intuitive test mentioned above: although someone other than the U.S. President in 1970 might have been the U.S. President in 1970 (e.g., Humphrey might have), no one other than Nixon might have been Nixon. In the same way, a designator rigidly designates a certain object if it designates that object wherever the object exists; if, in addition, the object is a necessary existent, the designator can be called strongly rigid. For example, 'the President of the U.S. in 1970' designates a certain man, Nixon; but someone else (e.g., Humphrey) might have been the President in 1970, and Nixon might not have; so this designator is not rigid.[44]

As several philosophers, in particular, Michael Slote, Hugh Chandler, and George Smith, have pointed out, these passages and others seem to suggest several non-equivalent definitions of the term "rigid designator."[45] The test for rigidity suggested by the second of the above paragraphs seems to be the following:

> A designating expression α is rigid if and only if
>
> (T_1) ⌜Somebody other than the person (or thing) who (which) is α might have been α⌝
>
> expresses a falsehood.

In contrast, the first of the above paragraphs seems to suggest the following test:

> A designating expression α is rigid if and only if
>
> (T_2) ⌜α might not have been α⌝
>
> expresses a falsehood.[46]

We can see that these tests, (T_1) and (T_2), are not equivalent by considering, for example, the designators "the father of W.A. Mozart" and "the son of Leopold Mozart." Kripke would argue that it is an essential property of W.A. Mozart that he came from the very sperm and egg that he did. If he is right about this (and I will assume that he is), it is false to say that somebody other than the person who was the father of W.A. Mozart might have been the father of W.A. Mozart. However, it seems true to say that the father of W.A. Mozart might not have been the father of W.A. Mozart. Surely Leopold didn't have to have any children at all. He might never have married. Similarly, it is true that someone other than the person who was the son of Leopold Mozart might have been the son of Leopold Mozart. Wolfgang's mother might have lost him through some accident before he was born; and his parents might have had a different baby boy at some later date. But it is false to say that the son of Leopold Mozart might not have been the son of Leopold Mozart. As I've been assuming, Wolfgang couldn't have had different parents. Thus, according to (T_1), "the father of W.A. Mozart" is a rigid designator, but "the son of Leopold Mozart" is not; however, according to (T_2), "the father of W.A. Mozart" is <u>not</u> a rigid designator, but "the son of Leopold Mozart" is.

Interestingly, these same problems arise when we consider another way in which Kripke attempts to explain the notion of

rigidity. Sometimes he defines rigidity by making explicit reference to the notion of "possible worlds"; and Chandler has pointed out that Kripke might intend either of the following definitions:

> (P_1) A designating expression is a rigid designator if and only if that expression designates the same object in any possible world in which it designates any object at all.
>
> (P_2) A designating expression is a rigid designator if and only if that expression designates the object which is its referent in the actual world in any possible world in which that object exists.[47]

As with (T_1) and (T_2), (P_1) entails that "the father of W.A. Mozart" is a rigid designator; but (P_2) entails that it is not. If "the father of W.A. Mozart" designates anything in some possible world, it designates the person who is, in that possible world, the father of W.A. Mozart. But, according to Kripke, that can only be the person who is, in the actual world, the father of W.A. Mozart. Thus, according to (P_1), "the father of W.A. Mozart" is rigid. In contrast, although that expression designates Leopold Mozart in the actual world, there are lots of possible worlds in which Leopold exists but does not have Wolfgang as a son. Since the description "the father of W.A. Mozart" does not, in those possible worlds, denote Leopold, "the father of W.A. Mozart" is not, according to (P_2), rigid.

Given that Kripke has offered these different, non-equivalent definitions of rigidity, we would seem to have

several options. First, we could treat all of the definitions as equals, and accept that there is not one concept of rigidity, but several. Alternatively, we could take the view that there is one clear concept of rigidity which is intended by Kripke but is not captured by any of the definitions given above. This last position is taken by George Smith in his in-depth study of rigid designation.[48] Or, finally, we might take one of the definitions to be "correct," in the sense that it really does capture what Kripke intends by the term "rigid designator," and hold that the other definitions were offered only because they are so similar and Kripke did not notice that they were not equivalent.

This last alternative might seem especially unlikely since it, unlike the other two, requires us to reject one of the definitions in favor of another. However, even though Kripke doesn't seem to give any reason for preferring one of the definitions over the others, I think it is at least plausible to argue that he, in fact, regards (T_1) as the true test of rigidity. In his most recent statement of the rigidity thesis, the preface to the book Naming And Necessity (which was published after the articles by Slote and Chandler), Kripke gives the following account or rigidity.[49] He asks us to consider the following two sentences:

(1) Aristotle was fond of dogs.

(2) The last great philosopher of antiguity was fond of dogs.

Kripke says that (1) "truly describes a counterfactual situation" if and only if the same person denoted by "Aristotle" in the actual world would have been fond of dogs had that situation obtained. In contrast, (2), says Kripke, truly describes a counterfactual situation if and only if the person who satisfies the definite description "the last great philosopher of antiquity" in that counterfactual situation would have been fond of dogs if that situation had obtained. Kripke takes this difference to illustrate that "Aristotle" is a rigid designator, whereas "the last great philosopher of antiquity" is not. And, in general, he regards a designator as rigid if, when it occurs in a simple sentence like (1) or (2), in deciding whether that sentence truly describes a particular counterfactual situation, we consider only what properties the person who is denoted by the designator in the actual world has in that counterfactual situation.

On this, Kripke's latest elucidation of the notion of rigidity, it seems to follow, just as (T_1) entailed, that "the father of W.A. Mozart" is rigid, but "the son of Leopold Mozart" is not. Consider sentences (3) and (4).

(3) The father of W.A. Mozart played the piano.

(4) The son of Leopold Mozart played the piano.

We have been assuming, with Kripke, that Wolfgang's father in the actual world is also his father in every possible world. Thus (3) would correctly describe a counterfactual situation if and only if the person who is Wolfgang's father in the actual world would have played the piano had that situation obtained. In contrast, Wolfgang's mother might have decided to abort him and have another son later on. Thus (4) would correctly describe a counterfactual situation if and only if the person who satisfies "the son of Leopold Mozart" in that counterfactual situation (who might not be Wolfgang) would have played the piano had that situation obtained. Therefore, since on Kripke's latest account of rigidity, as with (T_1), "the father of W.A. Mozart" is rigid and "the son of Leopold Mozart" is not rigid, we have some evidence that Kripke really does prefer the definition offered by (T_1) over the one offered by (T_2).

However, as I will now try to make clear, regardless of how one feels about each of the alternatives concerning the different tests of rigidity, both (T_1) and (T_2) are equally useful in establishing that proper names are not synonymous with their "associated" definite descriptions. Why does the fact, if it is a fact, that the proper name "Gorbachev" is a rigid designator, while the "associated" definite description "the leader of the Soviet Union" is not, show that they are not synonymous? There are lots of properties that two synonymous expressions can fail to share. Why isn't rigidity one of those

properties? I think we can see why if we consider the intuitive tests (T_1) and (T_2). By substituting "Gorbachev" for α in (T_1), we get:

> (5) Somebody other than the person who is Gorbachev might have been Gorbachev.

Similarly, by substituting "the leader of the Soviet Union" for α we get:

> (6) Somebody other than the person who is the leader of the Soviet Union might have been the leader of the Soviet Union.

When these two sentences are understood taking the first occurence of the expression replacing α to have large scope, we see that they have different truth values.[50] Sentence (5), understood "de re," is false, while sentence (6), understood "de re," is true. Someone other than Gorbachev might have succeeded Chernenko. But (6) can be obtained from (5) simply by replacing the occurences of "Gorbachev" with "the leader of the Soviet Union." So all we have done is to replace one expression with another expression which is supposed (by the description theorists) to be synonymous with it; and the result is a change in truth value. But how can this be? Surely, if two expressions have the very same meaning, then they ought to be interchangeable, salva veritate, in sentences like (5) and (6). Thus it would seem that the name "Gorbachev" can not be synonymous with the description "the leader of the Soviet Union" after all.

This conclusion seems to me to be correct; but more importantly for the purposes of this appendix, the above argument reveals a basic assumption of Kripke's argument for the view that names are not synonymous with definite descriptions. Let me use another example, one which Kripke himself uses, in order to highlight this assumption. Kripke argues that (7)

> (7) Somebody other than Aristotle might have been Aristotle.

is false, whereas (8)

> (8) Somebody other than Aristotle might have been the teacher of Alexander.

is true.[51] Again, it seems to follow that "Aristotle" is not synonymous with "the teacher of Alexander." However, it is clear that we make this inference only because we are assuming that if two expressions are synonymous, then they are interchangeable, salva veritate, in modal sentences like (5) - (8). Thus it is the fact that we can't substitute names and their "associated" definite descriptions in these modal contexts that makes Kripke's argument so compelling.

Once we realize this, it is clear that, regardless of what significance is given to the fact that Kripke offers both (T_1) and (T_2) as tests of rigidity, (T_1) and (T_2) serve equally well in the attack on the view that names are synonymous with "associated" definite descriptions. Consider the sentences we obtain by replacing α in (T_2) with "Aristotle" and "the teacher of Alexander."

(9) Aristotle might not have been Aristotle.

(10) The teacher of Alexander might not have been the teacher of Alexander.

While sentence (9) is false, sentence (10) is true. But (10) can be obtained from (9) simply by replacing occurences of "Aristotle" with "the teacher of Alexander." Thus those two designating expressions are not synonymous. This argument is just as strong as the analogous argument which uses instances of (T_1) instead of instances of (T_2). It does not matter that Kripke has offered both (T_1) and (T_2) as tests for <u>rigidity</u> or even that he might actually prefer (T_1) as a test for rigidity. The above argument does not even mention the expression "rigid designator."

Furthermore, there are even situations in which one of the two tests is useful in showing that a particular proper name is not synonymous with an "associated" definite description, but the other test is not useful. Take the name "Leopold Mozart" and the definite description "the father of Wolfgang Amadeus Mozart." Since most people who have ever heard of Leopold Mozart think of him as Wolfgang's father, the description "the father of W.A. Mozart" is a plausible one to be considered (by a description theorist) as synonymous with "Leopold Mozart." Now consider sentences (11) and (12).

(11) Somebody other than the person who was Leopold Mozart might have been Leopold Mozart.

(12) Somebody other than the person who was the father of W.A. Mozart might have been the father of W.A. Mozart.

These sentences are obtained from (T_1) by replacing L with "Leopold Mozart" and "the father of W.A. Mozart." But if Kripke is correct in holding that W.A. Mozart could not have had different parents from the ones he actually had, both (11) and (12) are false. Thus we are not yet in a position to claim that "Leopold Mozart" is not synonymous with "the father of W.A. Mozart."

However, let's apply the other test for rigidity. Consider the sentences we get by substituting "Leopold Mozart" and "the father of W.A. Mozart" for α in (T_2).

(13) Leopold Mozart might not have been Leopold Mozart.

(14) The father of W.A. Mozart might not have been the father of W.A. Mozart.

Sentence (13) is false; but what about (14)? It seems to assert that Wolfgang's father, Leopold, might not have been the father of W.A. Mozart. But this seems to be true. Leopold might have decided not to have any children, let alone Wolfgang. Thus here we do have a failure of substitutivity; and, unlike when we used (T_1), we are able to conclude that "Leopold Mozart" is not synonymous with "the father of W.A. Mozart."

In contrast, here's a situation in which (T_1), but not (T_2), is useful. Most people who have heard of Caroline Kennedy probably think of her as the daughter of John F. Kennedy. Thus a description theorist would probably consider

"Caroline Kennedy" to be synonymous with "the daughter of John F. Kennedy." What does (T_2) say about this? Sentences (15) and (16) are the appropriate instances of (T_2).

> (15) Caroline Kennedy might not have been Caroline Kennedy.
>
> (16) The daughter of John F. Kennedy might not have been the daughter of John F. Kennedy.

Sentence (15) is obviously false; and since Caroline could not have had different parents, (16) is false as well. Thus we are not yet able to conclude that "Caroline Kennedy" is not synonymous with "the daughter of John F. Kennedy."

Fortunately, (T_1) does enable us to draw this conclusion. Consider (17) and (18).

> (17) Somebody other than the person who is Caroline Kennedy might have been Caroline Kennedy.
>
> (18) Somebody other than the person who is the daughter of John F. Kennedy might have been the daughter of John F. Kennedy.

Surely (18) is true. President and Mrs. Kennedy might have decided to abort Caroline and have another daughter at a later date. Therefore, since (17) is obviously false, (T_1) does allow us to conclude that "Caroline Kennedy" is not synonymous with "the daughter of John F. Kennedy."

Keeping all of this in mind, let me now turn to the main purpose of this appendix. The fact that Kripke's arguments depend on the assumption that synonymous expressions are interchangeable, salva veritate, in <u>modal</u> sentences raises an interesting question. Could we prove that proper names are

not synonymous with their "associated" definite descriptions by pointing out other kinds of sentences in which the names and their descriptions are not interchangeable, salva veritate? I think we can; and the sentences which I want to consider reveal another difference in the manner in which names and many definite descriptions denote their referents. Moreover, I will argue that, although the notion I introduce is, in some ways, less powerful than Kripke's notion of rigidity, it is immune to several of the objections that have been raised concerning Kripke's non-synonomy arguments. And, before I introduce this notion, let me say what several of those objections are.

First, the arguments which I used to establish that names are not synonymous with their "associated" definite descriptions depend on our being able to give a large scope, or "de re," reading to the various instances of Kripke's intuitive tests, (T_1) and (T_2). For example, we concluded that "Aristotle" is not synonymous with "the teacher of "Alexander" because (9), read "de re," is false, while (10) read "de re," is true. However, some philosophers have argued that sentences like (9) and (10) can not be given a sensible "de re" interpretation. Here is how Kripke voices their concerns.

> Some philosophers have distinguished between essentialism, the belief in modality de re, and a mere advocacy of necessity, the belief in modality de dicto. Now, some people say: Let's give you the concept of necessity. A

much worse thing, something creating great additional problems, is whether we can say of any particular that it has necessary or contingent properties, even make the distinction between necessary and contingent properties. Look, it's only a statement or a state of affairs that can be either necessary or contingent! Whether a particular necessarily or contingently has a certain property depends on the ways it's described. This is perhaps closely related to the view that the way we refer to particular things is by a description. What is Quine's famous example? If we consider the number 9, does it have the property of necessary oddness? Has that number got to be odd in all possible worlds? Certainly it's true in all possible worlds, let's say, it couldn't have been otherwise, that nine is odd. Of course, 9 could also be equally well picked out as the number of planets. It is not necessary, not true in all possible worlds, that the number of planets is odd. For example if there had been eight planets, the number of planets would not have been odd. And so it's thought: Was it necessary or contingent that Nixon won the election? (It might seem contingent, unless one has some view of some inexorable processes . . .) But this is a contingent property of Nixon only relative to our referring to him as "Nixon" (assuming "Nixon" doesn't mean "the man who won the election at such and such a time"). But if we designate Nixon as "the man who won the election in 1968," then it will be a necessary truth, of course, that the man who won the election in 1968, won the election in 1968. Similarly, whether an object has the same property in all possible worlds depends not just on the object itself but on how it is described. So it's argued.

It is even suggested in the literature, that though a notion of necessity may have some sort of intuition behind it (we do think some things could have been otherwise; other things we don't think could have been otherwise), this notion [of a distinction between necessary and contingent properties] is just a doctrine made up by some bad philosopher, who (I guess) didn't realize that there are several ways of referring to the same thing.[52]

Having stated this objection, Kripke goes on to say the following about why these philosophers think that the doctrine of modality "de re," i.e., essentialism, is incoherent.

> Why have they thought this? While there are many motivations for people thinking this, one is this: The question of essential properties so-called is supposed to be equivalent (and it is equivalent) to the question of "identity across possible worlds." Suppose we have someone, Nixon, and there's another possible world where there is no one with all the properties Nixon has in the actual world. Which one of these other people, if any, is Nixon? Surely you must give some criterion of identity here! If you have a criterion of identity, then you just look in the other possible worlds at the man who is Nixon; and the question whether, in that other possible world, Nixon has certain properties, is well defined. It is also supposed to be well defined, in terms of such notions, whether it's true in all possible worlds, or there are some possible worlds in which Nixon didn't win the election. But, it's said, the problems of giving such criteria of identity are very difficult.[53]

Kripke's response to these objections is well known. He argues that people have thought that there is a problem of "identity across possible worlds" because they have been thinking about "possible worlds" in a misguided fashion. He says that we must not think of them as if we "discover" them by "looking through powerful telescopes." Rather they are stipulated by us; and, because we can stipulate what properties Nixon is to have in some possible world, there is no problem about identifying him. Although I find this response convincing, I will not go into this any deeper since my purpose is only to indicate that the doctrine of "de re"

necessity, i.e., essentialism, is somewhat controversial.

There is yet another criticism of sentences obtained from (T_1) and (T_2) by replacing α with various proper names. I have claimed, with Kripke, that these sentences are all, when read "de re," obviously false. However, some philosophers who are willing to grant that "de re" necessity at least makes sense, actually deny that it is clear that these sentences are false. They point out that Kripke doesn't give any arguments for the falsity of these sentences. He merely relies on our "intuitions." But the people voicing this objection claim that they just do not have any clear intuitions about what is and what is not metaphysically possible and, thus, they just do not know whether these sentences are true or false.

A particularly good example of this type of objection is given by Douglas Cannon in his doctoral dissertation.[54] He asks us to consider the sentence "Socrates might not have been Socrates." Although Cannon agrees that it makes sense to give this sentence a "de re" interpretation, he argues that it just is not clear that this sentence, even taking the first occurence of "Socrates" to have large scope, is false. To show this he asks us to suppose that (and we don't know that this isn't what really happened) Socrates actually had an identical twin brother. Unfortunately, after the original zygotes divided, the two resulting zygotes did not both survive. While the zygotes from which Socrates developed successfully implanted

itself in the uterine wall, the other zygote did not implant itself and died several hours later. Now, supposing that things really did happen this way, Cannon asks us to imagine the following counterfactual situation: The zygote which, in fact, was Socrates' twin brother successfully implanted itself in the uterine wall and was later born, while the zygote from which Socrates, in fact, developed did not implant itself and died a few hours later. Moreover, just as in the actual world, the baby's parents named it "Socrates" (or gave it the name which later on became "Socrates") and the baby grew up to be a philosopher, drank the hemlock, and did all the other things we, in fact, attribute to Socrates.

Cannon asks whether this counterfactual situation is one in which Socrates would not have been Socrates, i.e., whether somebody other than the person who actually was Socrates would have been Socrates. Although Cannon initially argues that the twin would have been Socrates, he backs off this claim and says:

> I am ready now to retreat a bit. Actually I do not believe that if Socrates' twin had not died but had lived and done everything that Socrates in fact did, then he would be Socrates. I have no idea who in that case would be Socrates. And I have been trying to shake your confidence that you know who would be. I believe that there is no answering that question.[55]

Is Cannon correct? Is there no answering the question whether the above counterfactual situation is one in which someone other than the person who actually was Socrates

would have been Socrates? I must confess that I do not find this example very convincing. It seems to me that this is obviously not a situation in which Socrates' actual twin brother would have been Socrates. Rather, it is a situation in which Socrates does not exist and somebody else, the twin, is merely being called "Socrates." But perhaps I'm wrong about this. Perhaps the reader will have different "intuitions" about this than I have. Indeed, it is often very puzzling and difficult to see what grounds we have for asserting either that a particular property is an essential property of some object or that it is only a contingent property of that object. But again, since my purpose is only to point out some features of Kripke's argument which some philosophers find objectionable, I will not attempt to settle this dispute.

Both this objection and the earlier one concerning the possibility of giving modal sentences a "de re" interpretation are bypassed by the notion I want to introduce now. Kripke points out that while proper names denote the same object in all possible worlds in which they denote anything at all, definite descriptions often denote different objects in different possible worlds. However, there is another way in which these two kinds of expressions differ. Consider the name "Socrates" and the description "the President of the U.S." The name "Socrates" has always and will always denote the same person, the man who (we think) taught Plato and drank the hemlock. In

contrast, "the President of the U.S.," although it presently denotes Reagan, used to denote Carter and many others.[56] Thus, while proper names do not denote different objects at different times, many definite descriptions do. Let me define (as Kripke does, "quasi-technically") the notion of an "eternal" designator as follows: A designating expression is an eternal designator if and only if it denotes the same object at all times at which it denotes anything at all. A designator that is not an eternal designator is a temporary designator.

I claim that all proper names are eternal designators, while, in many cases, their "associated" definite descriptions are only temporary designators. Moreover, just as Kripke does, I will give the following two "intuitive tests" which can be used to determine whether a designator is eternal or temporary:

(T_3) ⌜Somebody other than the person who is currently α was or will be α⌝

(T_4) ⌜α was not always or will not always be α⌝

If we substitute "Gorbachev" for α in (T_3) and (T_4), we get sentences (19) and (20).

(19) Somebody other than the person who is currently Gorbachev was or will be Gorbachev.

(20) Gorbachev was not always or will not always be Gorbachev.

These two sentences, understood "de re," are clearly false. In contrast, if we replace α with "the leader of the Soviet Union" (the description which I presume most people associate

with "Gorbachev", we get (21) and (22).

> (21) Somebody other than the person who is currently the leader of the Soviet Union was or will be the leader of the Soviet Union.
>
> (22) The leader of the Soviet Union was not always or will not always be the leader of the Soviet Union.

Clearly (21) and (22), interpreted "de re," are true! Gorbachev was not always the leader of the Soviet Union (both Stalin and Breshnev came before him); and he will not be the leader of the Soviet Union forever. Thus "Gorbachev" is an eternal designator, and "the leader of the Soviet Union" is a temporary designator.

It is interesting to note that the two tests, (T_3) and (T_4), are similar to Kripke's tests for rigidity, (T_1) and (T_2), in another respect. In certain situations, they can give different results. For example, if we replace α in (T_3) and (T_4) with "the biological mother of Caroline Kennedy," we get (23) and (24).

> (23) Somebody other than the person who is currently the biological mother of Caroline Kennedy was or will be the biological mother of Caroline Kennedy.
>
> (24) The biological mother of Caroline Kennedy was not always or will not always be the biological mother of Caroline Kennedy.

According to (T_3), "the biological mother of Caroline Kennedy" is an eternal designator, since (23) is obviously false. In contrast, according to (T_4), "the biological mother of Caroline Kennedy" is a temporary designator. Since Mrs. Kennedy did not

have Caroline until some time after she married John F. Kennedy, she was not always Caroline's mother.

However, just as I argued in the case of rigidity, even if (T_3) and (T_4) sometimes give different results, they are equally useful for the purpose of establishing that proper names are not synonymous with their "associated" definite descriptions. To see this, consider sentences (19) - (22) again. (21) and (22) can be obtained from (19) and (20), respectively, by substituting "the leader of the Soviet Union" for occurences of "Gorbachev" in (19) and (20). But (21) and (22) are true, while both (19) and (20) are false. Therefore, since these expressions are not interchangeable, salva veritate, in temporal sentences like (19) - (22), "Gorbachev" and its "associated" definite description "the leader of the Soviet Union" are not synonymous.[57] After all, there is surely no reason to think that a failure of substitution in modal sentences <u>does</u> show that the two expressions are not synonymous, but a failure of substitution in temporal sentences does <u>not</u>. My argument, which is based on a failure of substitution in temporal sentences, is just as strong a proof that "Gorbachev" is not synonymous with "the leader of the Soviet Union" as Kripke's.

In fact, temporal sentences give us an even more elegant way of establishing non-synonomy. Until now, I've been using sentences obtained from the Kripke-style tests, (T_3) and (T_4), in order to establish that "Gorbachev" is not synonymous with

"the leader of the Soviet Union." But those sentences are a bit awkward (especially the phrase "was not always and will not always be"); and there is no reason why we have to use them. Others will do just as well. For example, consider the following sentences:

(25) In 1975, Gorbachev was not Gorbachev.

(26) In 1975, Gorbachev was not the leader of the Soviet Union.

Interestingly, while (25) is obviously false, (26) is obviously true. Breshnev, not Gorbachev, was the leader of the Soviet Union in 1975. However, (26) is obtained from (25) merely by substituting "the leader of the Soviet Union" for the second occurence of "Gorbachev" in (25). Thus we have here another failure of substitutivity of the name and its "associated" definite description; and, hence, they can not be synonymous.

Are there any advantages or drawbacks to using temporal sentences, rather than modal sentences, in arguing that names are not synonymous with "associated" definite descriptions? I must admit that Kripke's notion of a "rigid" designator is, in at least one way, more powerful than my notion of an "eternal" designator. For there are lots of definite descriptions which are eternal designators, but not rigid designators. For example, the description "the inventor of bifocals" is an eternal designator. From the moment it came to denote anything at all, it has designaged, and will continue to designate, the same person,

(i.e., Ben Franklin). However, since Ben Franklin might not have invented bifocals, this description is not a rigid designator. Thus we will not be able to argue that the name "Ben Franklin" is not synonymous with "the inventor of bifocals" on the grounds that the name is an eternal designator while the description is a temporary designator. But we can argue that they are not synonymous on the grounds that the name is a rigid designator while the description is not rigid.

However, even if Kripke's modal sentences can prove non-synonomy in more cases than my temporal sentences, I'm not sure what should be made of this. After all, even Kripke can't prove non-synonomy in every case he would like. For example, suppose that we introduce the name "Harry" by stipulating that it is to denote the number 5. I suspect that Kripke would like to assert that the proper name "Harry" does not have the same meaning as the definite description "the square root of 25." He would probably say that the description has a "sense" but the proper name refers "directly" to its referent. But Kripke can not prove this by appealing to his notion of rigidity. Both expressions are, according to both (T_1) and (T_2), rigid designators. Furthermore, the description theorists hold, for various theoretical reasons, that proper names must be synonymous with definite descriptions, and, hence, that all proper names

are synonymous with their "associated" definite descriptions. My temporal argument shows that this view is mistaken. But once this is granted, it isn't clear what reason there would be to maintain that there are some cases in which the proper name is synonymous with its "associated" definite description. Thus, even if Kripke is able to prove non-synonomy in more cases than I can, it isn't clear what the theoretical interest of this is.

More importantly, however, even if Kripke has a few points on me in the above sense, my arguments involving temporal sentences are not open to either of the two objections I mentioned earlier which have been raised against Kripke's modal arguments. The first objection was that one can not give modal sentences "de re" interpretations. That is, one just can not make sense of the idea that an object has some of its properties essentially, and others contingently, independent of the manner in which the object is designated. But look again at sentence (26).

(26) In 1975, Gorbachev was not the leader of the Soviet Union.

Surely no one would claim that (26) does not make sense! The problems concerning "possible worlds" and "transworld identity" are not relevant here. All that (26) asserts is that, in 1975, a particular man did not have a particular property. In general, all that my temporal sentences, when understood "de re," would seem to commit us to is the

view that some objects endure through time while retaining some of their properties and losing others. While there may be philosophers who object to this, many of the philosophers who raise objections about "identity across possible worlds" and question the coherence of essentialism would not challenge these seemingly less controversial commitments.

How does my argument fare against Cannon's objection? He claims that Kripke hasn't established that names are rigid designators because it just is not clear that no one other than the person who actually was Socrates could have been Socrates. However, while I will admit that it isn't always clear what our grounds are for holding that a particular property (being Socrates) is essential to an object, this type of objection has no effect on my temporal arguments. All that I need to do in order to establish that "Gorbachev" is not synonymous with its "associated" definite description, "the leader of the Soviet Union," is to show that (25) differs in truth value from (26). But this is simple. Surely no one (unless he has an almost unbelievable distrust of the Soviet Union's leadership) would claim that Gorbachev, not Breshnev, was really running things in 1975. This is a well established empirical fact. Similarly, nothing could be more obvious than that (25) is false. Even philosophers who get confused over objects being called by different names at different times and in

different counterfactual situations would have no complaint here. Gorbachev, in fact, had the same name in 1975. To show that (25) is not false, one would have to show that the person who is at present Gorbachev was not Gorbachev in 1975. But it would be a ridiculous waste of time to attempt to do so.

Notes

[1]S.A. Kripke, "A Puzzle about Belief," in Meaning and Use, ed. A. Margalit, (Dordrecht: Reidel, 1979), pp. 239-283.

[2]Ibid., p. 269.

[3]S.A. Kripke, Naming and Necessity, (Cambridge, MA: Harvard University Press, 1980); originally in Semantics of Natural Language, ed. D. Davidson and G. Harman, (Dordrecht: Reidel, 1972), pp. 253-355.

[4]This is Kripke's actual statement of the principle. Of course, it can not, as it stands, be said to be true or false because the letter "p" is being used as a schematic letter. Kripke is here, as elsewhere, rather sloppy in his statements of the principles he has in mind. However, I will, in most cases, state Kripke's principles in the manner in which he introduces them, since it is usually easy enough to tell what he has in mind.

[5]I do not mean to assert that when we translate a sentence such as (22), which refers to a particular utterance, we never translate the sentence or expression inside the quotation marks. For example, suppose that at some point in a novel an author uses the following sentence:

(Q) John then said "London is pretty."

A person translating the novel into French might well put the sentence "Londres est jolie" into the translation of (Q), even though, according to the novel, John didn't really say the words "Londres est jolie." However, since we, in the context of this paper, do care about the actual words which are being uttered, we should not use "Londres est jolie" when translating (Q) into French.

[6]Kripke, "A Puzzle about Belief," p. 257

[7]Of course, in supposing that Peter uttered both "Londres est jolie" and "London is not pretty" at the same time, t, I do not mean to assert that the two sentences could be uttered at the very same instant of time. I mean that Peter uttered one of the sentences and then, almost immediately afterwards, uttered the other. Thus the expression "at t" is intended to refer, not to the time during which the first utterance was made, but to the time during which both utterances were made.

[8]Ibid., p. 268.

[9]Ibid., p. 269.

[10]Kripke, Naming and Necessity, p. 21.

[11]P.F. Strawson, "On Referring," Mind, LIX, No. 235 (1950), reprinted in Readings in the Philosophy of Language, ed. J. Rosenberg and C. Travis, (Englewood Cliffs, N.J.: Prentice-Hall, 1971), p. 181.

[12]R. Cartwright, "Propositions," in Analytic Philosophy Vol. 1. ed. R. Butler, (Oxford: Basil Blackwell, 1962), p. 92 and pp. 93-94.

[13]Ibid., p. 101.

[14]E.J. Lemmon, "Sentences, Statements and Propositions," originally in British Analytic Philosophy, ed. B. Williams and A. Montifiore, (New York: Humanities Press, Inc.,; London: Routledge & Kegan Paul, Ltd., 1966), Chapter 4, pp. 87-107; reprinted in Readings in the Philosophy of Language, ed. J. Rosenberg and C. Travis, (Englewood Cliffs, N.J.: Prentice Hall, 1971), pp. 233-250.

[15]Ibid., p. 242.

[16]I might add that this usage is unfortunate. Just as it is nonsense to say that someone asserted or said the meaning of a sentence, it is also nonsense to say either that the meaning of a sentence is true or that it is false. Yet, Lemmon says that propositions do, in fact, have truth values. To be fair, he does say that they have truth values in the following "derivative" sense: "A proposition may be said to be true (false) relative to context of

utterance, if there is a sentence, true (false) relative to that context, which expresses it." (p. 241)

[17]Also, I do not intend to rule out in this discussion the possibility that there are distinctions to be made in these contexts other than the notions of a sentence, the proposition (in Lemmon's terminology) or meaning which that sentence expresses or posseses, and the statement or proposition (in Strawson's and Cartwright's terminology) which is asserted by uttering it. In fact, in his recent work on demonstratives, Kaplan has drawn the distinction between what he calls the "content" of an assertion (or belief) and the "character" of that assertion (or belief). While the notion of the content of an assertion is precisely Cartwright's and Strawson's notion of a proposition (and Lemmon's notion of a statement), the character of an assertion is probably not to be identified with the meaning of a sentence used to make an assertion or indicate a belief. Kaplan says that if a sentence like "that's a penny" is used on two occasions, to make the same assertion (assuming that the speaker is pointing at the same penny) those assertions may still have a different "character" if the speaker is mistaken about the penny to which he is pointing (i.e., if he believes that he is pointing to different pennies). But since the meaning of the sentence which he has uttered on both occasions has not changed, it seems that the meaning of the sentence he uses can not be the character of the assertion. Again, I don't mean to attack or even fully explicate Kaplan's notion of the character of an assertion. I only want to emphasize that distinctions other than those discussed so far may be possible.

[18]Kripke, Naming and Necessity, p. 20.

[19]As I will discuss shortly, the expression "object of belief" can be understood in more than one way.

[20]I think this question is especially interesting in the sense that it isn't always clear why philosophers propose different theories of belief. Is there some phenomenon which must be explained? Is there some particular question to be answered or problem to be solved? Or, is an "analysis of belief" all that is desired? If so, what kinds of analyses are to be desired? I discuss similar questions with respect to theories of reference in the other paper included in this thesis, and I think many of the points made there are applicable here as well.

[21]Here, as elsewhere, I have made the theory specific to one sentence (in this case, "grass is green") for the sake of simplicity. Of course, the theory will only be true if all other such instances of it are true.

[22]A. Church, "On Carnap's Analysis of Statements of Assertion and Belief," Analysis, 10, (1950), p. 97.

[23]R. Carnap, "On Belief Sentences-Reply to Alonzo Church," in Philosophy and Analysis: A selection of articles published in Analysis between 1933-40 and 1947-53, ed. with an introduction by M. MacDonald, (Oxford: Basil Blackwell, 1954), (viii 296 pp.), pp. 128-131.

[24]It is not important for the purposes of this paper to determine whether it is possible to correctly analyze belief in terms of sentences. Thus, I will not enter into the details of the Church-Carnap debate.

[25]Kripke, "A Puzzle about Belief," p. 242.

[26]T. McKay, "On Proper Names in Belief Ascriptions," Philosophical Studies, 39, (1981), pp. 287-303.

[27]Ibid., p. 301.

[28]Ibid., p. 298.

[29]Ibid., p. 297.

[30]I. Kvart, "Kripke's Belief Puzzle," unpublished manuscript, p. 5.

[31]Ibid., p. 14.

[32]Ibid., p. 24.

[33]W. Lycan, "Toward a Homuncular Theory of Believing," Cognition and Brain Theory 4, pp. 139-159.

[34]Ibid., pp. 141-142.

[35]Ibid., pp. 142-143.

[36]Lycan thinks that Pierre is in essentially the same predicament with respect to London as Jones is in with respect to Cicero. Consequently, in discussing how his theory solves the puzzle, he talks about Jones, not Pierre. However, for the sake of clarity, I have restated Lycan's solution in terms of Pierre and London. In either case, the solution works equally well (which, I will argue, is not at all).

[37]Again, I've simplified these analyses by referring to the sentence "London is pretty." Lycan also does this with respect to the sentence "Cicero was bald." (page 147 of his paper).

Also, Lycan, just as Kvart, considers Pierre's assent to a sentence, say "London is not pretty," to be extremely good evidence that Pierre believes a token of London is not pretty.

[38]R. Marcus, "A Proposed Solution to a Puzzle about Belief," in Midwest Studies in Philosophy, Volume VI, The Foundations of Analytic Philosophy, ed. P. French, T. Uehling, Jr., and H. Wettstein, (Minneapolis: Univ. of Minnesota Press, 1981), pp. 501-510.

[39]Ibid., p. 505.

[40]Ibid., p. 505.

[41]Ibid., pp. 508-509.

[42]Ibid., p. 504.

[43]See above, pp. 51-61.

[44]Kripke, Naming and Necessity, pp. 47-49.

[45]M.A. Slote, Metaphysics and Essence, (New York: New York University Press, 1975), pp. 72-75.

H.S. Chandler, "Rigid Designation," The Journal of Philosophy, LXXII, (July 17, 1975), pp. 363-369.

G.E. Smith, "Rigid Designation, Scope and Modality," Diss. M.I.T., 1979.

[46]Of course, the sentences obtained by replacing α with a designating expression are intended to be given a "de re" reading. The "de dicto" readings of those sentences are obviously false and don't show anything about the rigidity or non-rigidity of the designator contained in them. One gets the "de re" reading of these sentences by taking the first occurrence of the designating expression to have large scope.

[47]Chandler, "Rigid Designation," pp. 363-369.

[48]G.E. Smith, "Rigid Designation, Scope and Modality," Diss. M.I.T., 1979.

[49]Kripke, Naming and Necessity, pp. 6-15.

[50]From now on, when I talk about a sentence obtained by substituting a designating expression for α in any of the schemata I will introduce, I intend that that sentence be given its "de re" interpretation.

[51]Kripke, Naming and Necessity, p. 74.

[52]Ibid., pp. 39-41.

[53]Ibid., p. 42.

[54]D. Cannon, "With Reference to Truth: Studies in Referential Semantics," Diss. M.I.T., 1982, pp. 33-40.

[55]Ibid., p. 40.

[56]Of course, in claiming that a particular name or description has denoted and will always denote the same object, I do not mean to assert that we could not, at some point in the future (or did not in the past) use that expression, say "the square root of 25," to refer to a different object. However, if we did this, we would be changing our usage (i.e., the meaning) of that expression. This is the very same assumption that Kripke makes when he claims that "the square root of 25" and "Socrates" are rigid designators even though there are lots of possible worlds in which we use these expressions (with a difference in meaning) to refer to objects other than their referents in the actual world.

[57]Also, while I won't go through this again, it will be possible to construct examples analogous to my earlier examples with "the father of W.A. Mozart" and "the daughter of John F. Kennedy" that show that, in some cases, one of the tests, but not the other, will be useful in proving non-synonomy.

PART TWO:

Fundamental Questions about Reference

What is the nature of the referential connection between words and objects?

How do proper names hook up with the world?

What is the nature of the glue that holds words and their referents together?

How do we manage to refer to things by using proper names?

How do we manage to refer to anything at all?

These questions are familiar to any student of the philosophy of language. Indeed many philosophers have regarded them as the fundamental questions to which any theory of reference must respond. In this paper, I will argue that, even though these questions are constantly asked, no clear sense has been given to them. Also, while it is relatively easy to suggest different ways in which they could be understood, it is extremely difficult to interpret these questions in ways that allow us to view the various well-known theories of reference as providing plausible but conflicting responses to them.

By way of introduction, consider the following passages. The first is from Kripke's work, Naming and Necessity. The next two are from Searle's "Proper Names."

> Let me give an example of some of the arguments which seem conclusive in favor of the view of Frege and Russell. The basic problem for any view such as Mill's is how we can determine what the referent of a name, as used by a given speaker, is. According to the description view, the answer is clear. If 'Joe Doakes' is just short for 'the man who corrupted

> Hadleyburg uniquely', then whoever corrupted Hadleyburg uniquely is the referent of the name 'Joe Doakes'. However, if there is not such a descriptive content to the name, then how do people ever use names to refer to things at all? Well, they may be in a position to point to some things and thus determine the references of certain names ostensively. This was Russell's doctrine of acquaintance, which he thought the so-called genuine or proper names satisfied. But of course ordinary names refer to all sorts of people, like Walter Scott, to whom we can't possibly point. And our reference here seems to be determined by our knowledge of them. Whatever we know about them determines the referent of the name as the unique thing satisfying those properties. For example, if I use the name 'Napoleon', and someone asks, "To whom are you referring? I will answer something like, 'Napoleon was the emperor of the French in the early part of the nineteenth century; he was eventually defeated at Waterloo', thus giving a uniquely identifying description to determine the referent of the name. Frege and Russell, then, appear to give the natural account of how reference is determined here; Mill appears to give none.[1]

> So now it seems as if the rules for a proper name must somehow be logically tied to particular characteristics of the object in such a way that the name has a sense as well as a reference; indeed, it seems it could not have a reference unless it did have a sense, for how, unless the name has a sense, is it to be correlated with the object?[2]

> But as a proper name does not in general specify any characteristics of the object referred to, how then does it bring the reference off? How is a connection between name and object ever set up? This, which seems the crucial question,. . .[3]

Not only do these passages explicitly ask several of the questions I want to examine, they also capture the spirit of much of the referential debate exceedingly well. In them, two competing theories of reference, Mill's and the "description" theory, are evaluated on the basis of how they

respond to questions like "How is the reference of a proper name determined?" or "How does a proper name come to be correlated with its referent?" Moreover, all of the passages suggest that the Millian view, i.e., that proper names have a sense, can not be correct because it leaves us with no answer to these questions. With this background material in mind, let's examine some of the ways in which these questions can be interpreted.

I

Perhaps the best way to try to understand these questions is to consider the views of Frege. After all, the phrase "sense determines reference" is best known as a shorthand statement of one of his most important doctrines. Frege originally developed the notion of "sense" in order to solve a philosophical puzzle. He claimed that we had to maintain that proper names have both a reference and a sense in order to explain how identity statements could be both true and informative. For example, "Cicero = Cicero" and "Cicero = Tully" are both true, and are about the same man. Yet, one is trivial and uninteresting while the other is or could be quite informative. According to Frege, we can only explain why this is so if we suppose that "Cicero" and "Tully" have different senses.[4]

However, to claim that proper names have both a sense and a reference would seem to be to claim something weaker than that the reference of a proper name is <u>determined</u> by its sense. What does Frege have in mind when he goes on to make this stronger claim? Here's one possibility.[5] For the sake of precision, we may view the relation between sense and reference as a set or ordered pairs whose first members are the senses of particular proper names and whose second members are the referents of those names. Frege's assertion that sense determines reference can be taken to mean that this relation is also a function. That is, no two

ordered pairs in the relation have the same sense as their first member.

However, on this understanding of "determine," while it may be true that sense determines reference, sense won't be the only thing that determines the reference of a proper name. The problem is that there are infinitely many functions which map ever so many kinds of things into the set of referents of proper names. For example, to obtain one such function, simply pair up each of the referents with a natural number, taking care not to pair up one numeber with two referents. To obtain a different function, just pair up the referents with different numbers. Obviously we can in this way generate an infinite number of these types of functions. Furthermore, we could in this way pair up different kinds of things with the referents of proper names. We could pair up the referent of "Nixon" with an atom of hydrogen, the referent of "Carter" with a different atom of hydrogen, and so on. In fact, we don't even have to pair up things of <u>one kind</u> with the referents. We could, for example, pair up the referent of "Reagan" with the number 5, the referent of "Carter" with an atom of hydrogen, and the referent of "Nixon" with a chair.

In addition, interpreting "determines" in this way does not help us to see why Frege felt compelled to assert that sense determines reference. Frege thought that sentences like "a = b" could be both true and informative only if

"a" and "b" both have a sense as well as a reference, and have different senses. However, consider two proper names, "c" and "d," which are not coreferential. In this case, Frege is not forced to assert that "c" and "d" have different senses in order to explain why "c = d" is both true and informative because "c = d" is not true. Thus Frege would have been free to assert that "c" and "d" have the same sense. But to assert this is to assert that sense does not determine (in the above set-theoretic sense of "determine") reference, since the set of ordered pairs whose first members are the senses of proper names and whose second members are the referents of those names (i.e., the relation between sense and reference) would have two ordered pairs with the same first member (the sense of both "c" and "d"). Thus, since the identity puzzle doesn't force Frege to assert that sense determines reference, it is not clear why he did make this stronger claim.

Hence, under this purely set-theoretic interpretation of "determine," it is not plausible to think that Frege asserted that sense determines reference in order to respond to the question "What determines the reference of a proper name?" And, as we have seen, this interpretation leaves it somewhat mysterious why Frege went so far as to assert that the sense of a proper name determines its referent. Consequently, we must search for other interpretations of "determine."

II

Another interpretation of our question is suggested by some of Kripke's remarks in Naming and Necessity. In that work, Kripke frequently asks not how the reference of a proper name is determined, but how it is "fixed."[6] In fact, he seems to use the terms "fix" and "determine" interchangeably, that is, synonymously.[7] Thus let's examine what Kripke means by expressions like "reference fixing."

The expressions in our language have meaning; moreover, we are constantly introducing new expressions. Often we must explain the meaning of an expression to someone. Wittgenstein says that these explanations can be divided up, roughly, into two kinds: verbal definitions and ostensive definitions. A verbal definition, for example, of the word "bachelor" may be given as follows. I say to some person "A bachelor is an unmarried male adult." In contrast, one may give an ostensive definition of, for instance, the word "red" by pointing to different red things and saying "This is red and this is red and this is red." In this case, my gestures enable the person to understand the term being defined. Also, sometimes when we give verbal or ostensive definitions we are not trying to explain what some expression means. Rather we are introducing a new expression and stipulating what that expression is to mean. For example, I might introduce the terms "grue" as follows: Something is grue if and only if it has been observed and is green or it has not

been observed and is blue. This expression has no meaning prior to my giving this verbal definition (or, more accurately, prior to Goodman's giving the definition).

Kripke makes an important point about these definitions. When we offer a verbal definition to explain or stipulate the meaning of a term, we don't always intend for the two expressions or phrases involved to have the same meaning. Kripke discusses at length the definition "A meter is the length of stick S" (for the sake of simplicity, I have made no reference to time and temperature in the definition).[8] He claims that even though we use the phrase "the length of stick S" in giving the definition, it is not synonymous with the term "meter." Although I won't go into detail here, the broad outlines of Kripke's argument for this claim are well-known. "Meter" is a rigid designator, whereas, "the length of stick S" is not.[9]

However, if "meter" and "the length of stick S" do not have the same meaning, in what sense is the statement that a meter is the length of stick S a definition at all? In response, Kripke draws a distinction between definitions which <u>give the meaning</u> of an expression and definitions which merely <u>fix the reference</u> of an expression. Although when we say that a meter is the length of stick S we don't intend the two expressions to be regarded as synonymous, we do indicate to what the expression "meter" is to refer. And by

showing to what the term refers we enable people to use it correctly. Similarly, when we introduce new expressions we sometimes give definitions which serve only to fix the reference of the expression. For example, by stipulating that "Glunk" refers to the largest football player on the Pittsburgh Steeler team, we have not supplied an expression which is to be regarded as synonymous with "Glunk." However, we have stipulated what "Glunk" is to refer to and can now proceed to use that term as a name of the player who satisfies the description.[10]

Unfortunately, this notion of "reference fixing" can not supply us with a plausible interpretation of the question "How is the reference of a proper name determined?" First if we take "determined" to mean "fixed," this question appears to ask what sort of definition was originally used to introduce the name into the language. Indeed, Kripke seems to understand the question in this way. In footnote 33 to Naming and Necessity, he says,

> An even better case of determining the reference of a name by description, as opposed to ostension, is the discovery of the planet Neptune. Neptune was hypothesized as the planet which caused such and such discrepancies in the orbits of certain other planets. If Leverrier indeed gave the name "Neptune" to the planet before it was ever seen, then he fixed the reference of "Neptune" by means of the description just mentioned.

This passage clearly suggests that the question "How was the reference of the proper name "Neptune" determined (fixed)?"

is adequately answered by indicating what kind of definition, verbal or ostensive, was used to introduce the name "Neptune."

However, this question, when asked about any particular proper name, is simply an empirical question concerning the manner in which that name was introduced into the language. To answer it, all we have to do is find out whether a verbal or ostensive definition was used to introduce the name. Although this might be difficult to do in practice, in principle it is a relatively simple matter. Thus it is hard to see why philosophers would bother to propose elaborate theories of reference, such as the causal and description theories, to answer it.

In fact, if it is held that all we have to do to answer the question "How is the reference of a proper name determined or fixed?" is to find out what sort of definition was used to introduce it, Kripke's own causal theory of reference would seem to be largely superfluous. That theory states that a name is introduced at an initial "baptism" ceremony and is then passed from "link to link" along a "causal chain" to present speakers. The referent of any proper name which we use today is the object which was "baptized" at the beginning of the "causal chain." However, if a theory of reference is only required to describe how the original baptism occured, i.e., through an ostensive or verbal (descriptive) definition, there would seem to be no point in going on to say anything about

causal chains or descriptions which might currently be associated with the name.

Moreover, if Kripke's question "How is the reference of a proper name determined (fixed)?" is only meant to ask how the name was originally introduced, why does Kripke use the present tense of the verb "to be"? Since most of the names we use have been in the language for some time, it would be more appropriate to ask how the reference of a particular proper name was determined. For example, if we want to know how the name "Aristotle" was originally introduced it would surely be rather strange to ask "How is the reference of 'Aristotle' determined?" [11]

For the above reasons, it is clear that Kripke's use of the notion of reference "fixing"has not furnished us with an adequate interpretation of the question "How is the reference of a proper name determined?" It simply is not plausible to suppose that so many philosophers would find it necessary to work out such elaborate theories of reference in order to respond to such a theoretically uninteresting, empirical, and apparently ill-stated question. Consequently, we are forced to look elsewhere for a reasonable interpretation of our question. Perhaps a good way to proceed is to consider how the expression "to determine" is used in other non-philosophical contexts.

III

Consider sentences (1) - (5).

1. How will he determine who broke the window?

2. If we don't stay and watch, how will we determine who won the Super Bowl while we are on the road?

3. How did you determine that he was the murderer?

4. Do you think that we could determine who Jack the Ripper really was after all of this time?

5. Is it possible to determine whether or not God exists?

These questions could be restated as follows:

6. How will he discover (or find out) who broke the window?

7. If we don't stay and watch, how will we find out who won the Super Bowl while we are one the road?

8. How did you discover that he was the murderer?

9. Do you think that we could find out who Jack the Ripper really was after all of this time?

10. Is it possible to discover whether or not God exists?

Sentences (6) - (10) clearly ask the same questions as sentences (1) - (5). Thus, in ordinary usage, "to determine" often means simply "to find out" or "to come to know."

In fact, there is, at least initially, some reason to think that this epistemic sense of "to determine" is intended by the question "How is the reference of a proper name determined?" Consider the following remarks from Naming and Necessity:

> The basic problem for any view such as Mill's is how we can determine what the referent of a name, as used by a given speaker, is.[12]

> The picture associated with the description theory is that only by giving some unique properties can you know who someone is and thus know what the reference of your name is.[13]

Both of these passages suggest that Kripke thinks that the question which any theory of reference must address is how we can learn what the referent of a proper name is.

Unfortunately, while the question is relatively clear when understood in this way, this interpretation must, for several reasons, be rejected. First, on this reading, there doesn't seem to be any philosophical reason for asking this question. To see this, suppose that someone were to say to me "Reagan is 74 years old." I, like most people, would not have any reason to ask how I could find out what the referent of the proper name "Reagan" is. I already know who Reagan is. In contrast, if someone uttered a sentence to me that contained a proper name which I had never heard of, I might well have to ask someone

how I could find out what that name referred to so that I could understand the earlier statement. In general, people ordinarily ask how they can find out what the referent of a proper name is only in order to understand some statement which made use of that name.[14] But clearly philosophers have no such practical reasons for asking how one can determine what the referent of a proper name is.

Also, and more importantly, even if there were some philosophical motivation for asking this question, neither the causal nor the description theory of reference can be viewed as even initially plausible answers. The question seems to ask for a description or list of the methods we actually employ to find out what the referents of particular proper names are. However, the causal and description theories don't seem to be concerned with providing any sort of description or list. The causal theory states only that the referent of a proper name is the person or object which lies at the beginning, or "baptism" point, of a "causal chain" of utterances of that name. The description theory says only that the referent of a name is the object which satisfies the description (or "cluster" of descriptions) with which the name is synonymous.

Still, while neither theory makes any explicit

statement about how one might actually go about finding out what the referent of a particular proper name is, we might try to view these theories as making some general suggestions in this regard. The description theory, for instance, might be taken as suggesting that in order to find out what the referent of some proper name is we must find out what descriptions are synonymous with that name and what object satisfies those descriptions. Similarly, the causal theory might seem to suggest (and this is especially general and vague) that we trace the causal chain associated with a proper name in order to find out what object was baptized at the beginning of the chain.

Initially these suggestions might seem, in a very general way, to be in accord with the methods we actually use to find out to what a proper name refers. To see this, consider how a person, Fred, who doesn't know to what the proper name "Ronald Reagan" refers, might go about finding out. Fred might ask the person who first uttered the name to him either to point out Reagan or to tell him where he could find someone who could point out Reagan. This kind of attempt seems to correspond to the causal theory in that Fred seems, to some degree, to be tracing the "causal chain" associated with"Ronald Reagan." At least he is not asking anyone to describe Reagan in any way. Alternatively, Fred might consult

other sources of information such as newspapers, dictionaries, or encyclopedias. If he does this, he will probably find out rather quickly that the name "Ronald Reagan" is associated with various descriptions. In particular, Fred will find out "Ronald Reagan" refers to the person who satisfies the definite description "the President of the United States." In this way, Fred could learn to what the name Ronald Reagan refers without bothering to have someone point Reagan out to him. Thus this kind of attempt would seem to be suggested, not by the causal theory, but by the description theory.

However, even if the two theories seem relevant and helpful in this very general way, it is still clear that their proponents did not propose them in order to answer the question "How can we find out what the referent of a proper name is?" The philosophers who have proposed the causal and description theories consider them to be complete in themselves and incompatible with each other. That is, each theory is intended to give a full account of how the reference of a proper name is determined; and it is not considered possible that both of the theories are true. However, when interpreted in the very general way above, both theories do, in fact, seem to accurately describe at least some of the ways in which we find out to what proper names refer. Moreover, each of the theories is obviously incomplete in that it fails to suggest the

methods which the other theory suggests. It may well be true that we can find out to what a proper name refers by getting someone to point out, without describing, the referent of the proper name. But we may equally well be able to find out to what a name refers, not by having someone point out the referent, but by finding some descriptions commonly associated with the name.

IV

Since investigating different senses of the expression "determine" has not enabled us to find a satisfactory interpretation of the question "How is the reference of a proper name determined?," perhaps it would be helpful to examine some other questions which are often raised in the same contexts. How do we manage to use proper names to refer to anything at all? Or, how is it possible to use a proper name to refer to something? Both of these questions are raised, in slightly different wording, by Kripke and Searle in the passages quoted at the beginning of this paper. But what is the problem? Isn't that just what we use proper names for, to refer to things? Let's see if we can give a clear sense to these questions.

In philosophy, questions containing the expressions "how is it possible to" and "how do we manage to" are often understood in a special way. Consider the following familiar questions.

11. How is motion possible?

12. How are true negative existential statements possible?

13. How do we manage to make identity statements which are both true and informative?

These questions are all asked for the same reason. In each case, there is a philosophical argument that the task involved is not possible. For example, the argument corresponding to question (11) goes as follows:

(14) In order to travel any distance, one must first travel half of that distance, and then half of the remaining distance, and then half of the remaining distance, and so on, ad infinitum.

∴ (15) In order to travel any distance, one must complete an infinite number of tasks (sub-journeys).

(16) One can't complete an infinite number of tasks in a finite amount of time.

∴ (17) One can't travel any distance. That is, motion is impossible.

Of course, this is just one rather rough formulation of Zeno's paradox. Other formulations are equally familiar. Similarly, the arguments associated with questions (12) and (13) are among the most famous in the philosophy of language.

Do questions like "How do we manage to use proper names to refer to things?" get their sense in this manner? Do the philosophers who ask them desire a response to some argument purporting to prove that it is impossible to use a proper name to refer to something? Interestingly, in recent years, an argument, or family of arguments, has been put forward which purports to show that it is indeed impossible to use any expression to mean something or to refer to something. These arguments can be found in Kripke's recent work on Wittgenstein, Quine's views on the "indeterminacy in translation" and the "inscrutability of reference," and Putnam's recent work on "metaphysical realism" and the correspondence theory of truth. These arguments are very complex

and difficult. Since I'm interested in their general structure, not their specific contents, and Kripke's discussion is the clearest of the three, I will concentrate on his statement of the argument.

In Wittgenstein on Rules and Private Language, Kripke presents what he considers to be the central argument of Wittgenstein's Philosophical Investigations in the form of a "skeptical paradox."[15] He asks us to imagine the following situation. For every person, there is some number such that it is larger than any number which he has previously added to another number. For the sake of simplicity, let's suppose that for some particular person, Jones, that number is 57. Also, suppose that some "skeptic" asks Jones what the sum of 57 and 68 is. Jones, after doing the calculation for the first time, responds that 57 plus 68 is 125. At this point the skeptic makes a rather incredible claim. He says that while Jones' answer is correct, Jones has actually changed his usage of the terms "plus," "sum," and "addition." In the past, says the skeptic, Jones used the addition sign, "+," to denote a different function, the "quaddition" function. That function is defined as follows.

(18) $(x)(y)$ $(x \text{ quus } y = \begin{cases} x \text{ plus } y & \text{if } x, y < 57 \\ 5 & \text{otherwise} \end{cases})$

The skeptic maintains that if Jones was using the words "plus," "sum," and "addition" (etc.) in the same way as he did in the past, he should have answered "5," and not "125."

Jones, of course, is incredulous and objects vehemently that he is absolutely certain that he used to mean addition by "plus" and not quaddition. In response, the skeptic admits that Jones feels certain about this, but points out that Jones had a bad LSD trip last night and that might account for his sincere but mistaken belief that he has not changed his usage of the terms "plus," "sum," and "addition." The skeptic challenges Jones to give some hard proof that, in the past, he meant "plus" rather than "quus."

Kripke then goes through a number of ways in which Jones might try to prove that in the past he really was computing the addition function rather than the quaddition function. The skeptic, in turn, shows that all of the evidence which Jones cites is inadequate. The skeptic rejects the suggestion that something present in Jones' mind or brain, such as mental pictures or images of mathematical rules, shows Jones meant "plus," rather than "quus." Also, nothing about Jones' past behavior or even his past dispositions to behave in particular ways establishes that he meant "plus." Nor would claiming that the hypothesis that Jones meant "plus" is more simple than the hypothesis that he meant "quus" show that Jones meant "plus."

In short, the skeptic claims that there is no reason to say that Jones used to mean "plus" rather than "quus." Moreover, if it is impossible to justify the claim that Jones,

in the past, meant "plus" rather than "quus," it would also seem impossible to justify the claim that Jones (or anyone else) is presently computing the addition function, and not some other function. Just as we are presently able to construct a skeptical hypothesis that Jones, in the past, was really computing the quaddition function, we could in the future construct a similar hypothesis that Jones is not really, at present, computing the addition function. And just as we can not refute the skeptic concerning Jones' past usage of terms like "plus," "sum," and "addition," we can not refute the future skeptic's claim concerning Jones' present usage of those terms.

This then is the "skeptical paradox." It is an argument purporting to establish that there is no reason to think that we are now computing the addition function, when we give the sums of various pairs of numbers, rather than some other function. It is, to say the least, a highly unusual argument. Readers unfamiliar with Kripke's presentation of this argument will undoubtedly have many questions. However, Kripke's actual discussion of the paradox is much too long and complex to include here in any detail. Fortunately, complete mastery of this difficult argument is not essential for understanding the general points which I will make about it.

Before I comment on the argument, however, I want to show, briefly, how some of the views of Quine and Putnam are related

to this skeptical paradox. Quine raises some of the same worries in his now famous work on the "indeterminacy of translation" and the "inscrutability of reference." In Word and Object, he asks how someone, without translation manuals or similar help, might translate the expression "gavagai" as used by the natives of a different culture.[16] Quine argues that there is no "fact of the matter" as to whether the expression should be translated as "rabbit" or as "undetached rabbit parts." He claims that there is nothing in the natives' behavior or in their behavioral dispositions which would justify translating "gavagai" one way rather than the other. As Kripke points out, this is slightly different from the skeptic's approach in that the skeptic is willing to consider any type of evidence, including, as Quine would not, mental images and other "inner states" that someone might cite. However, while Quine places some restrictions on the kinds of evidence he will consider, his argument strongly resembles the skeptical paradox in that both argue that we are not justified in making a particular claim about meaning. Kripke's skeptic argues that there is no evidence that shows that Jones meant "plus" rather than "quus." Quine argues that there is no evidence that establishes that "gavagai" means "rabbit" rather than "undetached rabbit parts."

Putnam, in his recent work, also expresses concern that

various interpretations of the expressions of our language are possible.[17] In contrast to Kripke and Quine, Putnam does not concentrate on any particular term. Rather, he argues that our language, viewed as a whole, has many "admissible" interpretations or models. He even gives a "proof" that there is an "admissible" interpretation of our language in which the word "cat" refers not to cats, but to cherries:[18] These bizarre interpretations are "admissible" in the sense that they meet certain "operational and theoretical" constraints. Putnam argues that it is impossible to justify any additional constraints that would single out one of these "admissible" interpretations as the "intended" interpretation. That is, there is no reason to say that one of the "admissible" interpretations is in fact the real or true interpretation of our language.[19] Again, this claim is analogous to the claim that Kripke's skeptic makes when he argues that there is no way to rule out either of the two possible interpretations ("plus" and "quus") of Jones' past usage of the terms "plus," "sum," and "addition."

From these brief descriptions of the arguments of Kripke, Quine, and Putnam, it is possible to see the general form of the argument that meaning something (and, hence, referring to something) is impossible. It goes as follows:

(19) There is no reason to say that, in using any expression, we mean one thing rather than something else.[20]

(20) If there is no good reason to say we mean one thing rather than another by our expressions, then it is false to say that we mean one thing rather than another. (It would be false to say that we mean "plus" and are, at present, computing the addition function).

∴ (21) In using expressions, we don't mean one thing rather than another. That is, meaning something is impossible.

Interestingly, this argument can be used to give a clear sense not only to the question "How do we manage to use proper names to refer to things?," but also to the question "How is the reference of a proper name determined?" When Kripke, Quine, and Putnam discuss their concerns about the notions of meaning and reference, they often make use of the expression "to determine" and its near relatives. For example, instead of asking how it is possible for an expression to mean one thing rather than another, they often ask what "determines" whether an expression (say "gavagai") means one thing ("rabbit") rather than another ("undetached rabbit parts"). Putnam even uses the term "to fix" in these contexts. Instead of asking how it is possible to single out one of the "admissible" interpretations of our language as the "intended" model, he will occasionally ask what "fixes" one particular interpretation of our language as the "intended" or "correct" model.

Since philosophers who offer the various theories of reference in response to questions like "How is the reference

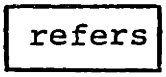
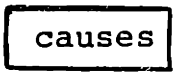
of a proper name determined?" and "How do we manage to use proper names to refer to anything?" obviously don't think that meaning and reference are impossible, it seems we must regard the causal and description theories as attempts to show what is wrong with the "skeptical" argument.[21] Moreover, since the argument is pretty clearly valid, they must be regarded as attempts to refute one of the premises. The second premise, (20), seems to be a formulation of the principle of sufficient reason. This kind of inference (or, in this case, premise) occurs occasionally in philosophical arguments. Whatever one feels about its legitimacy, it is obvious that the description and causal theories of reference are not intended to show that inferences relying on this principle are unjustified. Thus, if these theories are to be regarded as responding to the "skeptical" argument, we must take them to be directed against the first premise of the argument. They must be viewed as offering a way for us to rule out the various skeptical hypotheses by giving us some evidence establishing that one interpretation of an expression, rather than some other "skeptical" interpretation, is correct or "intended."

However, the description theory and the causal theory do <u>not</u> offer plausible reasons for rejecting "unintended" skeptical hypotheses. To see this, suppose that the skeptic proposes the following hypothesis about the proper name

"Aristotle." He says that just as Jones is mistaken in believing that he used to use "plus" to mean addition rather than quaddition, we are mistaken in claiming that we use the proper name "Aristotle" to refer to Aristotle rather than Caesar. We might, using the description theory, respond that the name "Aristotle" is synonymous with "the teacher of Alexander" and that description denotes Aristotle, not Caesar. However, anyone who has played the skeptic's game for a while could see how he would respond. The skeptic would deny that "the teacher of Alexander" denotes Aristotle. He would come up with a new hypothesis that the description really refers to Caesar, not Aristotle. Alternatively, he might even claim that the description denotes neither Aristotle nor Caesar, and offer this as evidence that the name and the description are not really synonymous. The problem is that the skeptic's skepticism is general. On his view, it is impossible for a proper name, a definite description, or any expression to have meaning or reference. Thus saying that one expression means the same thing as another expression does not say why either of them should be regarded as meaning or referring to one thing rather than something else. This response is so obvious that it is implausible to suppose that a philosopher would offer the description theory in response to the skeptic's argument.

Likewise, the causal theory does not offer a plausible way of singling out one interpretation of an expression or of

our language as a whole as the "intended" interpretation. We have this "from the horses' mouths" so to speak. Both Kripke and Putnam make it clear that their causal theories are not intended as responses to the skeptic. In "Realism and Reason," Putnam says

> Notice that a 'causal' theory of reference is not (would not be) of any help here: for how 'causes' can uniquely refer is as much of a puzzle as how 'cat' can on the metaphysical realist picture.
>
> The problem, in a way, is traceable back to Occam. Occam introduced the idea that concepts are (mental) particulars. If concepts are particulars ('signs'), then any concept we may have of the relation between a sign and its object is another sign. But it is unintelligible, from my point of view, how the sort of relation the metaphysical realist envisages as holding between a sign and its object can be singled out either by holding up the sign itself, thus
>
> 
>
>
> or by holding up yet another sign, thus
>
> 
>
>
> or perhaps
>
> 
>
>
> If concepts are not particulars, on the other hand, the obvious possibility is that (in so far as they are 'in the head') they are ways of using signs. But a 'use' theory, while intelligible (and, I believe, correct) as an account of what understanding the signs consists in, does not single out a unique relation between the terms of T_1 and the 'real objects'. If we don't think concepts are either particulars (signs) or ways of using signs, then, I think we are going to be led back to direct (and mysterious) grasps of Forms.[22]

Putnam seems to be making a similar point in the following passage from Reason, Truth and History:

> Suppose there is a possible naturalistic or physicalistic definition of reference, as Field contends. Suppose
>
> (1) X refers to y if and only if x bears R to y is true . . .
>
> the reference of 'X bears R to y' is itself indeterminate, and so knowing that (1) is true will not help. Each admissible model of our object language will correspond to a model of our meta-language in which (1) holds; the interpretation of 'X bears R to y' will fix the interpretation of 'X refers to y'. But this will only be a relation in each admissible model; it will not serve to cut down the number of admissible models at all.[23]

While both of these passages are difficult and unclear in several respects, it is clear that Putnam is claiming that causal theories of reference don't give us a way to rule out any of the bizarre but "admissible" interpretations as being "unintended."

Similarly, it seems equally obvious that Kripke doesn't think the causal theory which he put forward in Naming and Necessity would be of any use in rebutting the skeptic's challenge concerning "plus" and "quus." After all, although Kripke considers all sorts of responses which he finds plausible, he doesn't even mention his causal theory of reference in this context. Surely, if Kripke first offered the causal theory in order to respond to skeptical problems about how meaning and reference are possible, he would at least mention the theory.

So, while the skeptical paradox about meaning reference may well be, as Kripke says, one of the most profound of philosophical paradoxes, and may also be a good statement of what has been bothering many philosophers about meaning and reference, it doesn't solve the problem of this paper. We have been searching for an understanding of questions like "How is the reference of a proper name determined?" and "How do we manage to use a proper name to refer to something?" which allows us to regard the causal and description theories of reference as plausible responses. However, as we have just seen, these theories aren't intended to solve the problems presented by the skeptical paradox. Thus while the paradox enables us to give a clear sense to the questions with which we are concerned, we still don't understand what problem the causal and description theories of reference are supposed to address.

V

Our ordinary usage of questions containing phrases like "How is it possible to" and "how does one manage to" suggests several other ways in which we could understand the question "How do we manage to use proper names to refer to things?" I want now to examine a couple of these interpretations. First, consider questions (22) - (25).

(22) How did the baby manage to get out of the crib?

(23) How did Einstein manage to think of the theory of relativity?

(24) How did the Mets manage to win the 1969 World Series?

(25) How was it possible for Mark Spitz to win seven gold medals in the Olympics?

Unlike the questions ((11) - (13)) which we considered in Section IV, questions (22) - (25) and other similar everyday questions often have sense because there is some reason to think that the actions and feats performed are, not impossible, but difficult. No one ever won seven gold medals in an olympics before, and the Mets were one of the worst teams in the league the year before they won the World Series. To adequately respond to these questions one must explain how the difficulties in performing the feats involved were overcome.

Unfortunately, although questions containing phrases like "How is it possible to" and "how do we manage to" often get their sense in this way, we still are not in a position to understand the question "How do we manage to use proper names

to refer to things?" Suppose that someone were to ask how a particular person, Bob, manages to stand up. If Bob is a perfectly healthy adult, it's hard to see how we are to respond. But if we are told that both of Bob's legs are broken or that he has just drunk two full bottles of whiskey, the question makes sense. Similarly, it would make sense to ask how a person could manage to use an automobile to refer to something since words and certain gestures are usually used for this purpose. But since philosophers who ask how we manage to use proper names to refer to things haven't given any clear reasons for thinking it would be difficult to use proper names to refer, it isn't clear how we are to respond.

However, I think it is possible to interpret "how is it possible to" questions in another way which does not depend on our being able to give some reason for thinking that the task involved would be difficult to perform. Again, consider the question "How did Bob manage to stand up?" This time, however, imagine the following context: Professor Smith is teaching a course in anatomy and voluntary motor function at the local medical school. The entire course has been spent investigating the ways in which our bodies operate when we perform various everyday activities such as running and walking. One day, the professor asks one of his students, Bob, to stand up. After Bob has done this, Smith asks the entire class "How was it possible for Bob to stand up?" Even though no one has any

reason to think that Bob would have any difficulty in standing up, the professor's question is clear to all of the students. Smith is asking for a description of the causal relationships between the parts of the body involved in the act of standing up. He wants to know about the causally related sequence of events that took place in Bob's body from the time at which he decided to stand up to the time when he was on his feet.

Thus when philosophers ask how we manage to use proper names to refer to things, perhaps they are asking for a physical description of some causal mechanism involved in referring to something by uttering a proper name. In fact, this notion of a causal mechanism or sequence could also be used to give sense to the question "How is the reference of a proper name determined?" We often talk of one event causally determining another. For example, suppose that we have a machine rather like the ones that give change for dollar bills. This machine, however, is rather irregular. Sometimes it gives out four quarters in exchange for a dollar, and sometimes it gives out two quarters and five dimes. In this context, it makes perfect sense to ask what determines the output of the machine. What is desired is a description of the causally related sequence of events which take place inside the machine and which result in the machine giving out either four quarters or two quarters and five dimes. In other words, we want to know how the machine works.

Moreover, I have had many conversations about how the reference of a proper name is determined in which people have given me the impression that they are indeed conceiving of reference in this mechanistic way. These people seem struck by the fact that when we utter some proper name in a sentence, we often refer to objects that are a long distance away, both spatially and perhaps even temporally. It's almost as if, when they think of an instance of using a proper name to refer to something, they can picture the sound waves coming out of someone's mouth, and after travelling a particular path through some sort of mechanism, "landing" on the referent of the name. This doesn't seem to be much different from the way in which some people talk about the problem of, not referring to someone, but merely thinking about someone. The picture there is that, when we think about someone, somehow a beam or ray comes out of our brains and focuses on the person about whom we are thinking.

However, while it is possible to use these metaphors to give sense to the questions which I've been considering, this mechanistic interpretation is inadequate in several respects. First, it isn't completely clear that likening the relationship between the use of a proper name and the referent of that name to the mechanistic and causal sequence of events that take place in the metaphors which I have offered is appropriate. In many, if not most, of the

cases in which distinct events are causally related (i.e., where C causes something which causes something else which causes E), time passes between the occurences of the events. This is true of both the causal sequence involved in Bob's act of standing up and the causal sequence involved in the dollar machine's giving out change when a dollar is put in the slot. In contrast, the events of my uttering a proper name in some sentence and my referring to a particular object are not separated in time. No philosopher would claim that my referring to the referent of a proper name takes place at some time after my using the name in some sentence.

More importantly, even if we suppose that the two events are causally related and occur at the same time and that questions like "How is the reference of a proper name determined?" should be viewed as asking for a description of that relationship, this interpretation does not satisfy one of the requirements of this paper. I have argued that an adequate interpretation of the questions I've been examining requires that it is possible to view the causal and description theories of reference as initially plausible, but conflicting responses. Perhaps it is possible to view the causal theory, with its talk of a proper name passing from speaker to speaker along a "causal chain," as a description of some sort of causal or physical relationship between the referent of a name and various utterances of the name itself.

However, the description theory says only that a proper name is synonymous with a particular definite description or cluster of descriptions. It obviously does not even attempt to describe any causal or physical relationship between a proper name and its referent. Thus if Kripke is correct when he says that the description theory gives a natural response to the question "How is the reference of a proper name determined?" (and that Mill gives no response), we haven't yet found the intended sense of the question.

VI

Some readers might object that I've been looking for too deep and profound a sense for the questions I've been examining. Perhaps the proper way to understand them has been under our noses the whole time. Perhaps all that philosophers who ask these questions desire is a set of necessary and sufficient conditions for reference. That is, perhaps all that these philosophers want is to be able to correctly "fill in the blank" in expressions like (26).

(26) (x) ("Aristotle" refers to x if and only if _____).[24]

This interpretation of the question I've been considering has several features that recommend it. First, it satisfies the two requirements that I mentioned at the beginning of this paper. That is, it allows us to view the causal and description theories of reference as initially plausible, but conflicting answers to these questions. The fact that so many philosophers have supported the various versions of these theories would seem to indicate that they both have, at least initially, some degree of plausibility. And Kripke and others have provided many ingenious examples that show that the two theories are fundamentally incompatible.

Also, on this interpretation, it is still possible to see why someone might be inclined to ask questions like "How is reference possible?" Perhaps the best illustration of this is found in Putnam's book Reason, Truth and

History. He begins that book by telling a story about an ant which, while crawling about in the sand, accidently traces out a figure which strongly resembles Winston Churchill.[25] Putnam argues that, in spite of this, the ant's figure clearly does not "stand for," "represent," or "refer to" Winston Churchill. Moreover, particular words such as "Churchill" can and do "represent" or "refer to" Churchill even though they do not resemble him at all. Thus, says Putnam, mere similarity is neither necessary nor sufficient to make something stand for, or refer to something else.

At this juncture, Putnam asks,

> If similarity is not necessary or sufficient to make something represent something else, how can anything be necessary or sufficient for this purpose? How on earth can one thing represent (or "stand for," etc.) a different thing?[26]

The second of these questions clearly could be restated as "How (on earth) is reference possible?" or "How do we manage to refer to things?" Thus we have here an example of a philosopher asking one of the questions I've been examining in the context of a search for a set of necessary and sufficient conditions for reference. Moreover, his reason for expressing his desire in such a seemingly desperate tone is clearly that he thinks it will be exceedingly difficult to provide such a set.

However, although it may well be true that philosophers have been searching for a set of necessary and sufficient conditions for reference, this is not all that they desire

when they ask the questions I've been considering. To see this, consider the following four sentences:

(L) (x) ("Aristotle" refers to x if and only if "Aristotle" refers to x).

(T) (x) ("Aristotle" refers to x if and only if x is Aristotle).

(D) (x) ("Aristotle" refers to x if and only if x is denoted by the definite description "the teacher of Alexander" which is synonymous with the proper name "Aristotle").

(C) (x) ("Aristotle" refers to x if and only if x lies at the beginning ("point of baptism") of the causal chain associated with our current usage of the proper name "Aristotle").

Since all four of these sentences offer necessary and sufficient conditions for reference (to Aristotle), they all offer, in one sense, a "definition" or "theory" of reference. Moreover, both (L) and (T) are obviously true. Thus, if all that philosophers have been after is a set of necessary and sufficient conditions for reference, there would seem to be nothing left to do.

But philosophers have not been satisfied with these "definitions" or "theories." And it is not difficult to see why many philosophers would find (L) and, to a lesser degree, (T) of little interest. After all, one wouldn't even have to know what "refers" means in order to be able to tell that (T) is true. It is a logical truth. Similarly, knowing what "refers" means would seem to be about all that

one would need to know in order to be able to tell that (T) is true. It is very hard to imagine a person who both understands the terms "refers" and also thinks that (T) is false.

Thus it is clearly not plausible to interpret questions like "How do we manage to refer to things?" as merely asking for necessary and sufficient conditions for reference. The philosophers who ask these questions are after a <u>particular kind</u> of definition or theory. But what kind of definition should we try to give in response to these questions? Why are the definitions offered by sentences like (L) and (T) inadequate? Interestingly, these very questions are the central concern of several recent papers by Hartry Field and Hilary Putnam. Since these philosophers explicitly consider these questions, I think it will be helpful to look at the general character of their debate.

In his paper "Tarski's Theory of Truth," Field argues that just because we can generate a list of true sentences like (T) by substituting different proper names for "Aristotle" we should not think that we have given an adequate definition of reference. Here is the crucial passage of Field's article.

> Now, it would have been easy for a chemist, late in the last century, to have given a 'valence definition' of the following form:

(3) (∀E) (∀n) (E has valance n ≡ E is potassium and n is +1, or . . . or E is sulphur and n is -2)

where in the blanks go a list of similar clauses, one for each element. But, though this is an extensionally correct definition of valence, it would not have been an acceptable reduction; and had it turned out that nothing else was possible -- had all efforts to explain valence in terms of the structural properties of atoms proved futile -- scientists would have eventually had to decide either (a) to give up valence theory, or else (b) to replace the hypothesis of physicalism by another hypothesis (chemicalism?). It is part of scientific methodology to resist doing (b); and I also think it is part of scientific methodology to resist doing (a) as long as the notion of valence is serving the purposes for which it was designed (i.e., as long as it is proving useful in helping us characterize chemical compounds in terms of their valences). But the methodology is not to resist (a) and (b) by giving lists like (3); the methodology is to look for a real reduction. This is a methodology that has proved extremely fruitful in science, and I think we'd be crazy to give it up in linguistics. And I think we are giving up this fruitful methodology, unless we realize that we need to add theories of primitive reference to T1 or T2 if we are to establish the notion of truth as a physicalistically acceptable notion.[27]

So Field is arguing that the notions of truth and reference, if they are to be "scientifically acceptable," must not be defined by giving a list of sentences like (T). Rather, we must also seek a "physicalistic" definition. In particular, Field mentions the need for causal theories of reference like the theory which is proposed by Kripke and offered by (C).[28]

In his John Locke lectures of 1976, Putnam (citing a point made by Stephen Leeds) responds that we do not need a "physicalistic" definition of reference because the notion of reference is different in important ways from the concepts, like valence, of natural science. He argues as follows:

> The real point of Leed's reply to Field is this: valence is an explanatory notion (i.e., a causal-explanatory notion). Since we intend the existence of various valences to figure in chemistry as a cause, we have to say what valence is, not just give the numerical values. But Leeds is denying that reference is a causal-explanatory notion. We need notions like truth and reference to express certain things (which could, in principle, be expressed in other ways -- by using infinitary languages). For this purpose, it is immaterial if primitive reference is defined in what Field calls a "crazy" way. Reference isn't (or, anyway, Field hasn't shown that it is) a causal-explanatory notion. In short, we can give a "transcendental argument" for Tarski's procedure by appealing to a purpose for having notions like truth and reference which is not at all parallel to the purpose for which we have notions like valence.[29]

So Putnam, although he admits that it might be possible to give a correct "physicalistic" definition or analysis of reference, contends that Field has not established that it is necessary to do so.[30]

Regardless of whether it is, in fact, necessary that we provide a "physicalistic" definition of reference, this debate illustrates several important points. For all we know, it might be possible to give several, perhaps many, definitions of reference which are both true and, for one purpose or another, interesting and useful. And a definition which is adequate for one purpose may not be adequate for another. For example, a Tarski style definition of reference may be useful for some purposes, but it does not, as Field points out, serve to define reference in terms of the concepts of physics.[31] Similarly, a Kripke style causal theory of reference may be useful for some purposes but not for others.[32] Consequently, if a philosopher asks us to provide a theory or definition of reference, we can respond adequately only if we have a clear idea why that philosopher is asking for a definition.

Unfortunately, while all of this may be interesting and important, it doesn't solve the problem of this paper. For what purpose have philosophers proposed the causal and description theories of reference? To what problem or question are these theories addressed? The point of this paper has been that it is inadequate to respond by claiming that the causal and description theories are intended to tell us how reference is "determined" or how reference is "possible." And I'm certain that the reader is by now

aware that I don't believe that there is one clear problem to which the various theories of reference are addressed.

In order to drive this home, I would like to conclude by pointing out what I find to be a particularly striking example of how unclear philosophers can be about their reasons for proposing a particular theory of reference. As we have seen, Putnam, in his Locke lectures, has engaged in an extensive debate concerning the reasons why we ought to try to give particular kinds of definitions of reference. Thus one would expect that he would be especially clear about his reasons for proposing his own causal theory of reference. But consider what he says about that theory in the fifth Locke lecture.

> To say this is not to repudiate the 'causal theory of reference' (I would rather call it the 'social co-operation plus contribution of the environment theory of the specification of reference') that Kripke put forward and that I developed in 'The Meaning of 'Meaning''. Kripke and I were doing two things:
>
> (1) We were attacking the idea that speakers pick our referents in the following way: each term T is 'associated' by each speaker with a property P_T (the 'intension' of T). The terms applies to whatever has the property P_T.
>
> (2) We were giving an alternative account of how speakers do pick out referents if they don't associate terms with necessary and sufficient conditions (or properties P_T) as required by, say, Russell's theory.

> Both (1) and (2) still seem right to me, and worth doing. But a theory of how reference is specified isn't a theory of what reference is; in fact, it presupposes the notion of reference.[33]

According to this passage, the Putnam-Kripke causal theory of reference "presupposes" the notion of reference and is, therefore, only intended to answer the questions "How is reference specified?" and "How do speakers pick out referents?" In contrast, Field's sought after physicalistic definition of reference would not presuppose the notion of reference and would answer the question "What is reference?" However, expressions like "to specify" and "to pick out" are just as unclear as the expressions "to determine" and "to fix." Consequently, since he nowhere explains what he means by these terms, Putnam's own attempt to explain why he proposed his causal theory is a complete failure.

VII

In this paper, I have considered many possible interpretations of questions like "How is the reference of a proper name determined?" and "How do we manage to refer to anything at all?" Obviously, not every interpretation which might be given has been or even could be considered. Thus, while I think that I have covered their most natural interpretations, it is still possible that I have simply missed the intended senses of these questions. However, the main point of this paper has been to indicate how difficult it is to give sense to these questions in a way that permits us to regard the causal and description theories as plausible, though conflicting responses. And, if this paper inspires the reader to be more careful and clear about his or her reasons and goals in proposing a theory of reference, it has fulfilled its purpose.

Notes

[1]S.A. Kripke, Naming and Necessity, (Cambridge, Mass.: Harvard University Press, 1980); originally in Semantics of Natural Language, ed. D. Davidson and G. Harman, (Dordrecht: Reidel, 1972), pp. 27-28.

[2]J.R. Searle, "Proper Names," Mind, LXVII, No. 266 (1958); reprinted in Readings in the Philosophy of Language, ed. J. Rosenberg and C. Travis, (Englewood Cliffs, N.J.: Prentice Hall, 1971), p. 213.

[3]Ibid., pp. 215-216.

[4]In what follows, I will consider only names which have a reference. Also, I will not consider imperfections of ordinary language such as names which have more than one sense. I will talk about "the" sense of a proper name. In addition, I am using "proper name" in its ordinary sense, not in Frege's expanded sense.

[5]This presentation of some of the views of Frege is derived from a presentation of Frege's views given in a course by Richard Cartwright.

[6]Curiously, Putnam (another important causal theorist) in his new book, Reason, Truth and History, also uses this expression.

[7]I think this is fairly obvious to readers of the text. Perhaps the passages where this comes out most clearly occur around footnote 33 and after. Footnote 44 is itself strong evidence for this hypothesis.

[8]Kripke, Naming and Necessity, pp. 54-56.

[9]Actually, it is unfortunate that Kripke uses the expression "meter" as an example. If "meter" is substituted for α in the intuitive tests (T_1) and (T_2) discussed in the appendix to Part One of this thesis, the result (for example,

"meter might not have been meter") is not even grammatical. Perhaps Kripke could have avoided this problem by using either "one meter" or "the length which is one meter" instead.

[10]I've been talking only about "verbal" definitions. However, it is obvious that ostensive definitions serve only to fix (or determine) reference and not to give meaning. In these definitions, no other expression is being offered which could be taken to be synonymous with the expression being defined.

Also, it might be useful to mention that there is a fairly natural way of interpreting "fixing the meaning" or "giving the meaning" in which someone might feel that definitions which were claimed above only to fix the reference of an expression actually do "fix" or "give" its meaning. Since once someone is told to what some proper name refers, that person will then be able to use the name correctly, he will understand it. So, in the sense that some person who didn't understand how to use an expression acquires this ability once he is given the definition, the definition has "given" the meaning (or use) of the expression to the person.

[11]Actually the name "Aristotle" was probably introduced with a verbal definition. At some point, someone or some linguistic community probably stipulated that "Aristotle" is to have the same reference as "β" , where "β" is the name, in some other language, of Aristotle.

[12]Kripke, Naming and Necessity, pp. 27-28.

[13]Ibid., p. 83.

[14]Of course, this is just a special case of a more general phenomenon. If I already know something, I would not seem to have any reason for asking how I could find it out.

[15]S.A. Kripke, Wittgenstein on Rules and Private Language, (Cambridge, Mass.: Harvard University Press, 1982).

[16]W.V.O. Quine, Word and Object, (Cambridge, Mass.: MIT Press, 1960), Chapter 2, pp. 26-80.

[17]H. Putnam, Reason, Truth and History, (New York: Cambridge University Press, 1982).

H. Putnam, "Realism and Reason," in Meaning and the Moral Sciences, (London: Routledge & Kegan Paul, 1978), pp. 121-140.

[18]H. Putnam, Reason, Truth and History, pp. 32-38, pp. 217-218.

[19]I have only given a very general statement of what Putnam actually argues. I haven't explained what "operational and theoretical" constraints are and why Putnam feels that they are the only correct ways of narrowing down the set of admissible interpretations. However, it is unimportant for the purposes of this paper to get into these matters.

[20]This is a little too strong to be justified by the actual examples discussed by Kripke and Quine. At best, they show that meaning something by "plus" and "gavagai" are problematic. In defense of (22), Kripke does, in his book (pages 19-30), show how a skeptical paradox could be developed for other expressions in the language. Indeed, once one sees the kinds of hypotheses that the skeptic is willing to defend, it is hard to imagine how any expression could be exempt from his attack.

[21]Actually, while Kripke and Putnam both state that they don't think that these "skeptical" arguments show that meaning and reference are impossible, Quine is somewhat less clear. He says that meaning and reference just are "indeterminate," and there is no "fact of the matter" as to whether "gavagai" is to be translated as "rabbit" or as "undetached rabbit parts." However, it isn't important in this context to clarify Quine's remarks further.

[22]H. Putnam, "Realism and Reason," pp. 126-127.

[23]H. Putnam, Reason, Truth and History, pp. 45-46.

[24]Of course, by expressions "like" (26) , I mean that "grass is green" in (26) is to be replaced by other English sentences.

[25]H. Putnam, Reason, Truth and History, pp. 1-2.

[26]Ibid., p. 2.

[27]H. Field, "Tarski's Theory of Truth," The Journal of Philosophy, 69, (1972), pp. 362-363.

[28]Ibid., pp. 366-367.

Interestingly, Field would not accept the actual theory or "picture" of reference that Kripke proposes. As Kripke himself admits, that theory presupposes the notion of "intending to refer," and hence, does not completely eliminate the concept of reference in favor of the concepts of physics.

[29]H. Putnam, "Meaning and Knowledge (The John Locke Lectures, 1976)," in Meaning and the Moral Sciences, (London: Routledge & Kegan Paul, 1978), p. 16.

[30]Ibid., p. 17.

[31]In Section II of this paper, I pointed out that we sometimes offer analyses of particular terms in order to teach people who don't already understand those terms how to use them correctly. For this purpose, the Tarski style definition of reference offered by sentences like (T) would seem to be adequate. After all, that analysis does give us a way of eliminating the term "refers."

[32]For example, Kripke's causal analysis could not be used for the purpose of teaching someone who didn't understand the term "refers" how to use that term correctly. To see this, consider what Kripke says about his theory on pages 96-97 of Naming and Necessity.

> A rough statement of a theory might be the following: An initial "baptism" takes place. Here the object may be named by ostension, or the reference of the name may be fixed by a description. When the name is 'passed from link to link,' the receiver of the name must, I think intend when he learns it to use it with the same referent as the man from whom he heard it. If I hear the name 'Napoleon' and decide it would be a nice name for my pet aardvark, I do not satisfy this condition. (Perhaps it is some such failure to keep reference fixed which accounts for the divergence of present uses of "Santa Claus" from the alleged original use.)
>
> Notice that the preceding outline hardly eliminates the notion of reference; on the contrary, it takes the notion of intending to use the same reference as a given.

Thus Kripke admits that the notion of a "causal chain" employed in his theory depends upon the notion of intending to refer to the same thing as the person from whom one originally got the name. So, if the causal analysis of Kripke were to be written out completely, the expression "refer" would appear on both sides of the biconditional. Since this definition or analysis would be circular, it could not be used to teach someone what "refers" means.

[33]H. Putnam, "Meaning and Knowledge (The John Locke Lectures, 1976)," p. 58.