

SAUL A. KRIPKE

## A PUZZLE ABOUT BELIEF

In this paper I will present a puzzle about names and belief. A moral or two will be drawn about some other arguments that have occasionally been advanced in this area, but my main thesis is a simple one: that the puzzle *is* a puzzle. And, as a corollary, that any account of belief must ultimately come to grips with it. Any speculation as to solutions can be deferred.

The first section of the paper gives the theoretical background in previous discussion, and in my own earlier work, that led me to consider the puzzle. The background is by no means necessary to *state* the puzzle: As a philosophical puzzle, it stands on its own, and I think its fundamental interest for the problem of belief goes beyond the background that engendered it. As I indicate in the third section, the problem really goes beyond beliefs expressed using names, to a far wider class of beliefs. Nevertheless, I think that the background illuminates the genesis of the puzzle, and it will enable me to draw one moral in the concluding section.

The second section states some general principles which underlie our general practice of reporting beliefs. These principles are stated in much more detail than is needed to comprehend the puzzle; and there are variant formulations of the principles that would do as well. Neither this section nor the first is necessary for an intuitive grasp of the central problem, discussed in the third section, though they may help with fine points of the discussion. The reader who wishes rapid access to the central problem could skim the first two sections lightly on a first reading.

In one sense the problem may strike some as no puzzle at all. For, in the situation to be envisaged, all the relevant facts can be described in *one* terminology without difficulty. But, in *another* terminology, the situation seems to be impossible to describe in a consistent way. This will become clearer later.

### I. PRELIMINARIES : SUBSTITUTIVITY

In other writings,<sup>1</sup> I developed a view of proper names closer in many ways to the old Millian paradigm of naming than to the Fregean tradition which probably was dominant until recently. According to Mill, a proper name is, so to speak, *simply* a name. It *simply* refers to its bearer, and has no other

linguistic function. In particular, unlike a definite description, a name does not describe its bearer as possessing any special identifying properties.

The opposing Fregean view holds that to each proper name, a speaker of the language associates some property (or conjunction of properties) which determines its referent as the unique thing fulfilling the associated property (or properties). This property(ies) constitutes the 'sense' of the name. Presumably, if '...' is a proper name, the associated properties are those that the speaker would supply, if asked, "Who is '...?'" If he would answer "... is the man who \_\_\_\_\_," the properties filling the second blank are those that determine the reference of the name for the given speaker and constitute its 'sense.' Of course, given the name of a famous historical figure, individuals may give different, and equally correct, answers to the "Who is ...?" question. Some may identify Aristotle as the philosopher who taught Alexander the Great, others as the Stagirite philosopher who studied with Plato. For these two speakers, the sense of "Aristotle" will differ: in particular, speakers of the second kind, but not of the first kind, will regard "Aristotle, if he existed, was born in Stagira" as analytic.<sup>2</sup> Frege (and Russell)<sup>3</sup> concluded that, strictly speaking, different speakers of English (or German!) ordinarily use a name such as 'Aristotle' in different senses (though with the same reference). Differences in properties associated with such names, strictly speaking, yield different idiolects.<sup>4</sup>

Some later theorists in the Frege-Russellian tradition have found this consequence unattractive. So they have tried to modify the view by 'clustering' the sense of the name (e.g., Aristotle is the thing having the following long list of properties, or at any rate most of them), or, better for the present purpose, socializing it (what determines the reference of 'Aristotle' is some roughly specified set of *community-wide* beliefs about Aristotle).

One way to point up the contrast between the strict Millian view and Fregean views involves — if we permit ourselves this jargon — the notion of propositional content. If a strict Millian view is correct, and the linguistic function of a proper name is completely exhausted by the fact that it names its bearer, it would appear that proper names of the same thing are everywhere interchangeable not only *salva veritate* but even *salva significatione*: the proposition expressed by a sentence should remain the same no matter what name of the object it uses. Of course this will not be true if the names are 'mentioned' rather than 'used': " 'Cicero' has six letters" differs from " 'Tully' has six letters" in truth value, let alone in content. (The example, of course, is Quine's.) Let us confine ourselves at this stage to

*simple* sentences involving no connectives or other sources of intensionality. If Mill is completely right, not only should “Cicero was lazy” have the same *truth value* as “Tully was lazy,” but the two sentences should express the same *proposition*, have the same content. Similarly “Cicero admired Tully,” “Tully admired Cicero,” “Cicero admired Cicero,” and “Tully admired Tully,” should be four ways of saying the same thing.<sup>5</sup>

If such a consequence of Mill’s view is accepted, it would seem to have further consequences regarding ‘intensional’ contexts. Whether a sentence expresses a necessary truth or a contingent one depends only on the proposition expressed and not on the words used to express it. So any simple sentence should retain its ‘modal value’ (necessary, impossible, contingently true, or contingently false) when ‘Cicero’ is replaced by ‘Tully’ in one or more places, since such a replacement leaves the content of the sentence unaltered. Of course this implies that coreferential names are substitutable in modal contexts *salva veritate*: “It is necessary (possible) that Cicero . . .” and “It is necessary (possible) that Tully . . .” must have the same truth value no matter how the dots are filled by a simple sentence.

The situation would seem to be similar with respect to contexts involving knowledge, belief, and epistemic modalities. Whether a given subject believes something is presumably true or false of such a subject no matter how that belief is expressed; so if proper name substitution does not change the content of a sentence expressing a belief, coreferential proper names should be interchangeable *salva veritate* in belief contexts. Similar reasoning would hold for epistemic contexts (“Jones knows that . . .”) and contexts of epistemic necessity (“Jones knows *a priori* that . . .”) and the like.

All this, of course, would contrast strongly with the case of definite descriptions. It is well known that substitution of coreferential descriptions in simple sentences (without operators), on any reasonable conception of ‘content,’ *can* alter the content of such a sentence. In particular, the modal value of a sentence is not invariant under changes of coreferential descriptions: “The smallest prime is even” expresses a necessary truth, but “Jones’s favorite number is even” expresses a contingent one, even if Jones’s favorite number happens to be the smallest prime. It follows that coreferential descriptions are *not* interchangeable *salva veritate* in modal contexts: “It is necessary that the smallest prime is even” is true while “It is necessary that Jones’s favorite number is even” is false.

Of course there is a ‘*de re*’ or ‘large scope’ reading under which the second sentence is true. Such a reading would be expressed more accurately by

“Jones’s favorite number is such that it is necessarily even” or, in rough Russellian transcription, as “One and only one number is admired by Jones above all others, and any such number is necessarily even (has the property of necessary evenness).” Such a *de re* reading, if it makes sense at all, by definition must be subject to a principle of substitution *salva veritate*, since necessary evenness is a property of the *number*, independently of how it is designated; in this respect there can be no contrast between names and descriptions. 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,” or even, in Carnap’s ‘formal’ mode of speech,<sup>6</sup> “‘Cicero was bald’ expresses a necessary truth.” Now the Millian asserts that all these formulations retain their truth value when ‘Cicero’ is replaced by ‘Tully,’ even though ‘Jones’s favorite Latin author’ and ‘the man who denounced Catiline’ would *not* similarly be interchangeable in these contexts even if they are codesignative.

Similarly for belief contexts. Here too *de re* beliefs — as in “Jones believes, of Cicero (or: of his favorite Latin author) that he was bald” do *not* concern us in this paper. Such contexts, if they make sense, are by definition subject to a substitutivity principle for both names and descriptions. Rather we are concerned with the *de dicto* locution expressed explicitly in such formulations as, “Jones believes that: Cicero was bald” (or: “Jones believes that: the man who denounced Catiline was bald”). The material after the colon expresses the *content* of Jones’s belief. Other, more explicit, formulations are: “Jones believes the proposition — that — Cicero — was — bald,” or even in the ‘formal’ mode, “The sentence ‘Cicero was bald’ gives the content of a belief of Jones.” In all such contexts, the strict Millian seems to be committed to saying that codesignative names, but not codesignative descriptions, are interchangeable *salva veritate*.<sup>7</sup>

Now it has been widely assumed that these apparent consequences of the Millian view are plainly false. First, it seemed that sentences can alter their *modal* values by replacing a name by a codesignative one. “Hesperus is Hesperus” (or, more cautiously: “If Hesperus exists, Hesperus is Hesperus”) expresses a necessary truth, while “Hesperus is Phosphorus” (or: “If Hesperus exists, Hesperus is Phosphorus”), expresses an empirical discovery, and hence, it has been widely assumed, a contingent truth. (It might have

turned out, and hence might have been, otherwise.)

It has seemed even more obvious that codesignative proper names are not interchangeable in belief contexts and epistemic contexts. Tom, a normal speaker of the language, may sincerely assent to "Tully denounced Catiline," but not to "Cicero denounced Catiline." He may even deny the latter. And his denial is compatible with his status as a normal English speaker who satisfies normal criteria for using both 'Cicero' and 'Tully' as names for the famed Roman (without knowing that 'Cicero' and 'Tully' name the same person). Given this, it seems obvious that Tom believes that: Tully denounced Catiline, but that he does not believe (lacks the belief) that: Cicero denounced Catiline.<sup>8</sup> So it seems clear that codesignative proper names are not interchangeable in belief contexts. It also seems clear that there must be two distinct propositions or contents expressed by 'Cicero denounced Catiline' and 'Tully denounced Catiline.' How else can Tom believe one and deny the other? And the difference in propositions thus expressed can only come from a difference in *sense* between 'Tully' and 'Cicero.' Such a conclusion agrees with a Fregean theory and seems to be incompatible with a purely Millian view.<sup>9</sup>

In the previous work mentioned above, I rejected one of these arguments against Mill, the modal argument. 'Hesperus is Phosphorus,' I maintained, expresses just as necessary a truth as 'Hesperus is Hesperus'; there are no counterfactual situations in which Hesperus and Phosphorus would have been different. Admittedly, the truth of 'Hesperus is Phosphorus' was not known *a priori*, and may even have been widely disbelieved before appropriate empirical evidence came in. But these epistemic questions should be separated, I have argued, from the metaphysical question of the necessity of 'Hesperus is Phosphorus.' And it is a consequence of my conception of names as 'rigid designators' that codesignative proper names are interchangeable *salva veritate* in all contexts of (metaphysical) necessity and possibility; further, that replacement of a proper name by a codesignative name leaves the modal value of any sentence unchanged.

But although my position confirmed the Millian account of names in modal contexts, it equally appears at first blush to imply a *nonMillian* account of epistemic and belief contexts (and other contexts of propositional attitude). For I presupposed a sharp contrast between epistemic and metaphysical possibility: Before appropriate empirical discoveries were made, men might well have failed to know that Hesperus was Phosphorus, or even to believe it, even though they of course knew and believed that Hesperus was Hesperus.

Does not this support a Fregean position that 'Hesperus' and 'Phosphorus' have different 'modes of presentation' that determine their references? What else can account for the fact that, before astronomers identified the two heavenly bodies, a sentence using 'Hesperus' could express a common belief, while the same context involving 'Phosphorus' did not? In the case of 'Hesperus' and 'Phosphorus,' it is pretty clear what the different 'modes of presentation' would be: one mode determines a heavenly body by its typical position and appearance, in the appropriate season, in the evening; the other determines the same body by its position and appearance, in the appropriate season, in the morning. So it appears that even though, according to my view, proper names would be *modally* rigid — would have the same reference when we use them to speak of counterfactual situations as they do when used to describe the actual world — they would have a kind of Fregean 'sense' according to how that rigid reference is fixed. And the divergences of 'sense' (in this sense of 'sense') would lead to failures of interchangeability of co-designative names in contexts of propositional attitude, though not in modal contexts. Such a theory would agree with Mill regarding modal contexts but with Frege regarding belief contexts. The theory would not be *purely* Millian.<sup>10</sup>

After further thought, however, the Fregean conclusion appears less obvious. Just as people are said to have been unaware at one time of the fact that Hesperus is Phosphorus, so a normal speaker of English apparently may not know that Cicero is Tully, or that Holland is the Netherlands. For he may sincerely assent to 'Cicero was lazy,' while dissenting from 'Tully was lazy,' or he may sincerely assent to 'Holland is a beautiful country,' while dissenting from 'The Netherlands is a beautiful country.' In the case of 'Hesperus' and 'Phosphorus,' it seemed plausible to account for the parallel situation by supposing that 'Hesperus' and 'Phosphorus' fixed their (rigid) references to a single object in two conventionally different ways, one as the 'evening star' and one as the 'morning star.' But what corresponding *conventional* 'senses,' even taking 'senses' to be 'modes of fixing the reference rigidly,' can plausibly be supposed to exist for 'Cicero' and 'Tully' (or 'Holland' and 'the Netherlands')? Are not these just two names (in English) for the same man? Is there any special *conventional, community-wide* 'connotation' in the one lacking in the other?<sup>11</sup> I am unaware of any.<sup>12</sup>

Such considerations might seem to push us toward the extreme Frege-Russellian view that the senses of proper names vary, strictly speaking, from speaker to speaker, and that there is no community-wide sense but only a

community-wide reference.<sup>13</sup> According to such a view, the sense a given speaker attributes to such a name as 'Cicero' depends on which assertions beginning with 'Cicero' he accepts and which of these he regards as *defining*, for him, the name (as opposed to those he regards as mere factual beliefs 'about Cicero'). Similarly, for 'Tully.' For example, someone may define 'Cicero' as 'the Roman orator whose speech was Greek to Cassius,' and 'Tully' as 'the Roman orator who denounced Catiline.' Then such a speaker may well fail to accept 'Cicero is Tully' if he is unaware that a single orator satisfied both descriptions (if Shakespeare and history are both to be believed). He may well, in his ignorance, affirm 'Cicero was bald' while rejecting 'Tully was bald,' and the like. Is this not what actually occurs whenever someone's expressed beliefs fail to be indifferent to interchange of 'Tully' and 'Cicero'? Must not the source of such a failure lie in two distinct associated descriptions, or modes of determining the reference, of the two names? If a speaker does, as luck would have it, attach the same identifying properties both to 'Cicero' and to 'Tully,' he *will*, it would seem, use 'Cicero' and 'Tully' interchangeably. All this appears at first blush to be powerful support for the view of Frege and Russell that in general names are peculiar to idiolects, with 'senses' depending on the associated 'identifying descriptions.'

Note that, according to the view we are now entertaining, one *cannot* say, 'Some people are unaware that Cicero is Tully.' For, according to this view, there is no single proposition denoted by the 'that' clause, that the community of normal English speakers expresses by 'Cicero is Tully.' Some — for example, those who define both 'Cicero' and 'Tully' as 'the author of *De Fato*' — use it to express a trivial self-identity. Others use it to express the proposition that the man who satisfied one description (say, that he denounced Catiline) is one and the same as the man who satisfied another (say, that his speech was Greek to Cassius). There is no single fact, 'that Cicero is Tully,' known by some but not all members of the community.

If I were to assert, "Many are unaware that Cicero is Tully," I would use 'that Cicero is Tully' to denote the proposition that I understand by these words. If this, for example, is a trivial self-identity, I would assert falsely, and irrelevantly, that there is widespread ignorance in the community of a certain self-identity.<sup>14</sup> I *can*, of course, say, "Some English speakers use both 'Cicero' and 'Tully' with the usual referent (the famed Roman) yet do not assent to 'Cicero is Tully.'"

This aspect of the Frege-Russellian view can, as before, be combined with a concession that names are rigid designators and that hence the description

used to fix the reference of a name is not synonymous with it. But there are considerable difficulties. There is the obvious intuitive unpalatability of the notion that we use such proper names as 'Cicero,' 'Venice,' 'Venus' (the planet) with differing 'senses' and for this reason do not 'strictly speaking' speak a single language. There are the many well-known and weighty objections to any description or cluster-of-descriptions theory of names. And is it definitely so clear that failure of interchangeability in belief contexts implies some difference of sense? After all, there is a considerable philosophical literature arguing that even word pairs that are straightforward synonyms if any pairs are — "doctor" and "physician," to give one example — are not interchangeable *salva veritate* in belief contexts, at least if the belief operators are iterated.<sup>15</sup>

A minor problem with this presentation of the argument for Frege and Russell will emerge in the next section: if Frege and Russell are right, it is not easy to state the very argument from belief contexts that appears to support them.

But the clearest objection, which shows that the others should be given their proper weight, is this: the view under consideration does not in fact account for the phenomena it seeks to explain. As I have said elsewhere,<sup>16</sup> individuals who "define 'Cicero'" by such phrases as "the Catiline denouncer," "the author of *De Fato*," etc., are relatively rare: their prevalence in the philosophical literature is the product of the excessive classical learning of some philosophers. Common men who clearly use 'Cicero' as a name for Cicero may be able to give no better answer to "Who was Cicero?" than "a famous Roman orator," and they probably would say the same (if anything!) for 'Tully.' (Actually, most people probably have never heard the name 'Tully.') Similarly, many people who have heard of both Feynman and Gell-Mann, would identify each as 'a leading contemporary theoretical physicist.' Such people do not assign 'senses' of the usual type to the names that uniquely identify the referent (even though they use the names with a determinate reference). But to the extent that the *indefinite* descriptions attached or associated can be called 'senses,' the 'senses' assigned to 'Cicero' and 'Tully,' or to 'Feynman' and 'Gell-Mann,' are *identical*.<sup>17</sup> Yet clearly speakers of this type can ask, "Were Cicero and Tully one Roman orator, or two different ones?" or "Are Feynman and Gell-Mann two different physicists, or one?" without knowing the answer to either question by inspecting 'senses' alone. Some such speaker might even conjecture, or be under the vague false impression, that, as he would say, 'Cicero was bald but Tully was not.' The



premise of the argument we are considering for the classic position of Frege and Russell — that whenever two codesignative names fail to be interchangeable in the expression of a speaker's beliefs, failure of interchangeability arises from a difference in the 'defining' descriptions the speaker associates with these names — is, therefore, false. The case illustrated by 'Cicero' and 'Tully' is, in fact, quite usual and ordinary. So the apparent failure of codesignative names to be everywhere interchangeable in belief contexts, is not to be explained by differences in the 'senses' of these names.

Since the extreme view of Frege and Russell does not in fact explain the apparent failure of the interchangeability of names in belief contexts, there seems to be no further reason — for present purposes — not to give the other overwhelming *prima facie* considerations against the Frege-Russell view their full weight. Names of famous cities, countries, persons, and planets are the common currency of our common language, not terms used homonymously in our separate idiolects.<sup>18</sup> The apparent failure of codesignative names to be interchangeable in belief contexts remains a mystery, but the mystery no longer seems so clearly to argue for a Fregean view as against a Millian one. Neither differing public senses nor differing private senses peculiar to each speaker account for the phenomena to be explained. So the apparent existence of such phenomena no longer gives a *prima facie* argument for such differing senses.

One final remark to close this section. I have referred before to my own earlier views in "Naming and Necessity." I said above that these views, inasmuch as they make proper names rigid and transparent<sup>19</sup> in modal contexts, favor Mill, but that the concession that proper names are not transparent in belief contexts appears to favor Frege. On a closer examination, however, the extent to which these opacity phenomena really support Frege against Mill becomes much more doubtful. And there are important theoretical reasons for viewing the "Naming and Necessity" approach in a Millian light. In that work I argued that ordinarily the real determinant of the reference of names of a former historical figure is a chain of communication, in which the reference of the name is passed from link to link. Now the legitimacy of such a chain accords much more with Millian views than with alternatives. For the view supposes that a learner acquires a name from the community by determining to use it with the same reference as does the community. We regard such a learner as using "Cicero is bald" to express the same thing the community expresses, regardless of variations in the properties different learners associate with 'Cicero,' as long as he determines that he will use the

name with the referent current in the community. That a name can be transmitted in this way accords nicely with a Millian picture, according to which only the reference, not more specific properties associated with the name, is relevant to the semantics of sentences containing it. It has been suggested that the chain of communication, which on the present picture determines the reference, might thereby itself be called a 'sense.' Perhaps so — if we wish<sup>20</sup> — but we should not thereby forget that the legitimacy of such a chain suggests that it is just preservation of reference, as Mill thought, that we regard as necessary for correct language learning.<sup>21</sup> (This contrasts with such terms as 'renate' and 'cordate,' where more than learning the correct extension is needed.) Also, as suggested above, the doctrine of rigidity in modal contexts is dissonant, though not necessarily inconsistent, with a view that invokes anti-Millian considerations to explain propositional attitude contexts.

The spirit of my earlier views, then, suggests that a Millian line should be maintained as far as is feasible.

## II. PRELIMINARIES: SOME GENERAL PRINCIPLES

Where are we now? We seem to be in something of a quandary. On the one hand, we concluded that the failure of 'Cicero' and 'Tully' to be interchangeable *salva veritate* in contexts of propositional attitude was by no means explicable in terms of different 'senses' of the two names. On the other hand, let us not forget the initial argument against Mill: If reference is *all there is* to naming, what semantic difference can there be between 'Cicero' and 'Tully'? And if there is no semantic difference, do not 'Cicero was bald' and 'Tully was bald' express exactly the same proposition? How, then, can anyone believe that Cicero was bald, yet doubt or disbelieve that Tully was?

Let us take stock. Why do we think that anyone can believe that Cicero was bald, but fail to believe that Tully was? Or believe, without any logical inconsistency, that Yale is a fine university, but that Old Eli is an inferior one? Well, a normal English speaker, Jones, can sincerely assent to 'Cicero was bald' but not to 'Tully was bald.' And this even though Jones uses 'Cicero' and 'Tully' in standard ways — he uses 'Cicero' in this assertion as a name for the Roman, not, say, for his dog, or for a German spy.

Let us make explicit the *disquotational principle* presupposed here, connecting sincere assent and belief. It can be stated as follows, where '*p*' is to be replaced, inside and outside all quotation marks, by any appropriate standard English sentence: "*If a normal English speaker, on reflection,*

*sincerely assents to 'p,' then he believes that p.*" The sentence replacing '*p*' is to lack indexical or pronominal devices or ambiguities, that would ruin the intuitive sense of the principle (e.g., if he assents to "You are wonderful," he need not believe that *you* — the reader — are wonderful).<sup>22</sup> When we suppose that we are dealing with a normal speaker of English, we mean that he uses all words in the sentence in a standard way, combines them according to the appropriate syntax, etc.: in short, he uses the sentence to mean what a normal speaker should mean by it. The 'words' of the sentence may include proper names, where these are part of the common discourse of the community, so that we can speak of using them in a standard way. For example, if the sentence is "London is pretty," then the speaker should satisfy normal criteria for using 'London' as a name of London, and for using 'is pretty' to attribute an appropriate degree of pulchritude. The qualification "on reflection" guards against the possibility that a speaker may, through careless inattention to the meaning of his words or other momentary conceptual or linguistic confusion, assert something he does not really mean, or assent to a sentence in linguistic error. "Sincerely" is meant to exclude mendacity, acting, irony, and the like. I fear that even with all this it is possible that some astute reader — such, after all, is the way of philosophy — may discover a qualification I have overlooked, without which the asserted principle is subject to counterexample. I doubt, however, that any such modification will affect any of the uses of the principle to be considered below. Taken in its obvious intent, after all, the principle appears to be a self-evident truth. (A similar principle holds for sincere affirmation or assertion in place of assent.)

There is also a strengthened 'biconditional' form of the disquotational principle, where once again any appropriate English sentence may replace '*p*' throughout: *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.*<sup>23</sup> The biconditional form strengthens the simple one by adding that failure to assent indicates lack of belief, as assent indicates belief. The qualification about reticence is meant to take account of the fact that a speaker may fail to avow his beliefs because of shyness, a desire for secrecy, to avoid offense, etc. (An alternative formulation would give the speaker a sign to indicate lack of belief — not necessarily disbelief — in the assertion propounded, in addition to his sign of assent.) Maybe again the formulation needs further tightening, but the intent is clear.

Usually below the simple disquotational principle will be sufficient for our purposes, but once we will also invoke the strengthened form. The simple

form can often be used as a test for disbelief, provided the subject is a speaker with the modicum of logicality needed so that, at least after appropriate reflection, he does not hold simultaneously beliefs that are straightforward contradictions of each other — of the forms ‘ $p$ ’ and ‘ $\sim p$ .’<sup>24</sup> (Nothing in such a requirement prevents him from holding simultaneous beliefs that jointly *entail* a contradiction.) In this case (where ‘ $p$ ’ may be replaced by any appropriate English sentence), the speaker’s assent to the negation of ‘ $p$ ’ indicates not only his disbelief that  $p$  but also his failure to believe that  $p$ , using only the simple (unstrengthened) disquotational principle.

So far our principle applies only to speakers of English. It allows us to infer, from Peter’s sincere reflective assent to “God exists,” that he believes that God exists. But of course we ordinarily allow ourselves to draw conclusions, stated in English, about the beliefs of speakers of any language: we infer that Pierre believes that God exists from his sincere reflective assent to “*Dieu existe*.” There are several ways to do this, given conventional translations of French into English. We choose the following route. We have stated the disquotational principle in English, for English sentences; an analogous principle, stated in French (German, etc.) will be assumed to hold for French (German, etc.) sentences. Finally, we assume the *principle of translation*: *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).* Some of our ordinary practice of translation may violate this principle; this happens when the translator’s aim is not to preserve the content of the sentence, but to serve — in some other sense — the same purposes in the home language as the original utterance served in the foreign language.<sup>25</sup> But if the translation of a sentence is to mean the same as the sentence translated, preservation of truth value is a minimal condition that must be observed.

Granted the disquotational principle expressed in each language, reasoning starting from Pierre’s assent to ‘*Dieu existe*’ continues thus. First, on the basis of his utterance and the French disquotational principle we infer (in French):

*Pierre croit que Dieu existe.*

From this we deduce,<sup>26</sup> using the principle of translation:

Pierre believes that God exists.

In this way we can apply the disquotational technique to all languages.

Even if I apply the disquotational technique to English alone, there is a sense in which I can be regarded as tacitly invoking a principle of translation. For presumably I apply it to speakers of the language other than myself. As Quine has pointed out, to regard others as speaking the same language as I is in a sense tacitly to assume a *homophonic* translation of their language into my own. So when I infer from Peter's sincere assent to or affirmation of "God exists" that he believes that God exists, it is arguable that, strictly speaking, I combine the disquotational principle (for Peter's idiolect) with the principle of (homophonic) translation (of Peter's idiolect into mine). But for most purposes, we can formulate the disquotational principle for a single language, English, tacitly supposed to be the common language of English speakers. Only when the possibility of individual differences of dialect is relevant need we view the matter more elaborately.

Let us return from these abstractions to our main theme. Since a normal speaker — normal even in his use of 'Cicero' and 'Tully' as names — can give sincere and reflective assent to "Cicero was bald" and simultaneously to "Tully was not bald," the disquotational principle implies that he believes that Cicero was bald and believes that Tully was not bald. Since it seems that he need not have contradictory beliefs (even if he is a brilliant logician, he need not be able to deduce that at least one of his beliefs must be in error), and since a substitutivity principle for coreferential proper names in belief contexts would imply that he does have contradictory beliefs, it would seem that such a substitutivity principle must be incorrect. Indeed, the argument appears to be a *reductio ad absurdum* of the substitutivity principle in question.

The relation of this argument against substitutivity to the classical position of Russell and Frege is a curious one. As we have seen, the argument can be used to give *prima facie* support for the Frege-Russell view, and I think many philosophers have regarded it as such support. But in fact this very argument, which has been used to support Frege and Russell, cannot be stated in a straightforward fashion if Frege and Russell are right. For suppose Jones asserts, "Cicero was bald, but Tully was not." If Frege and Russell are right, I cannot deduce, using the disquotational principle:

- (1) Jones believes that Cicero was bald but Tully was not,

since, in general, Jones and I will not, strictly speaking, share a common idiolect unless we assign the same 'senses' to all names. Nor can I combine disquotation and translation to the appropriate effect, since homophonic

translation of Jones's sentence into mine will in general be incorrect for the same reason. Since in fact I make no special distinction in sense between 'Cicero' and 'Tully' — to me, and probably to you as well, these are interchangeable names for the same man — and since according to Frege and Russell, Jones's very affirmation of (1) shows that for him there *is* some distinction of sense, Jones must therefore, on Frege-Russellian views, use one of these names differently from me, and homophonic translation is illegitimate. Hence, if Frege and Russell are right, we *cannot* use this example in the usual straightforward way to conclude that proper names are not substitutable in belief contexts — even though the example, and the ensuing negative verdict on substitutivity, has often been thought to support Frege and Russell!

Even according to the Frege-Russellian view, however, *Jones* can conclude, using the disquotational principle, and expressing his conclusion in his own idiolect:

(2) I believe that Cicero was bald but Tully was not.

*I* cannot endorse this conclusion in Jones's own words, since I do not share Jones's idiolect. I *can* of course conclude, "(2) expresses a truth in Jones's idiolect." I can also, if I find out the two 'senses' Jones assigns to 'Cicero' and 'Tully,' introduce two names 'X' and 'Y' into my own language with these same two senses ('Cicero' and 'Tully' have already been preempted) and conclude:

(3) Jones believes that X was bald and Y was not.

All this is enough so that we can still conclude, on the Frege-Russellian view, that codesignative names are not interchangeable in belief contexts. Indeed this can be shown more simply on this view, since codesignative descriptions plainly are not interchangeable in these contexts and for Frege and Russell names, being essentially abbreviated descriptions, cannot differ in this respect. Nevertheless, the simple argument, apparently free of such special Frege-Russellian doctrinal premises (and often used to support these premises), in fact cannot go through if Frege and Russell are right.

However, if, *pace* Frege and Russell, widely used names are common currency of our language, then there no longer is any problem for the simple argument, using the disquotational principle, to (2). So, it appears, on pain of convicting Jones of inconsistent beliefs — surely an unjust verdict — we must

not hold a substitutivity principle for names in belief contexts. If we used the *strengthened* disquotational principle, we could invoke Jones's presumed lack of any tendency to assent to 'Tully was bald' to conclude that he does not believe (lacks the belief) that Tully was bald. Now the refutation of the substitutivity principle is even stronger, for when applied to the conclusion that Jones believes that Cicero was bald but does not believe that Tully was bald, it would lead to a straightout contradiction. The contradiction would no longer be in Jones's beliefs but in our own.

This reasoning, I think, has been widely accepted as proof that codesignative proper names are not interchangeable in belief contexts. Usually the reasoning is left tacit, and it may well be thought that I have made heavy weather of an obvious conclusion. I wish, however, to question the reasoning. I shall do so without challenging any particular step of the argument. Rather I shall present — and this will form the core of the present paper — an argument for a paradox about names in belief contexts that invokes *no* principle of substitutivity. Instead it will be based on the principles — apparently so obvious that their use in these arguments is ordinarily tacit — of disquotation and translation.

Usually the argument will involve more than one language, so that the principle of translation and our conventional manual of translation must be invoked. We will also give an example, however, to show that a form of the paradox may result within English alone, so that the only principle invoked is that of disquotation (or, perhaps, disquotation plus *homophonic* translation). It will intuitively be fairly clear, in these cases, that the situation of the subject is 'essentially the same' as that of Jones with respect to 'Cicero' and 'Tully.' Moreover, the paradoxical conclusions about the subject will parallel those drawn about Jones on the basis of the substitutivity principle, and the arguments will parallel those regarding Jones. Only in these cases, no special substitutivity principle is invoked.

The usual use of Jones's case as a counterexample to the substitutivity principle is thus, I think, somewhat analogous to the following sort of procedure. Someone wishes to give a *reductio ad absurdum* argument against a hypothesis in topology. He does succeed in refuting this hypothesis, but his derivation of an absurdity from the hypothesis makes essential use of the unrestricted comprehension schema in set theory, which he regards as self-evident. (In particular, the class of all classes not members of themselves plays a key role in his argument.) Once we know that the unrestricted comprehension schema and the Russell class lead to contradiction by themselves, it is

clear that it was an error to blame the earlier contradiction on the topological hypothesis.

The situation would have been the same if, after deducing a contradiction from the topological hypothesis plus the 'obvious' unrestricted comprehension schema, it was found that a similar contradiction followed if we replaced the topological hypothesis by an apparently 'obvious' premise. In both cases it would be clear that, even though we may still not be confident of any specific flaw in the argument against the topological hypothesis, blaming the contradiction on that hypothesis is illegitimate: rather we are in a 'paradoxical' area where it is unclear *what* has gone wrong.<sup>27</sup>

It is my suggestion, then, that the situation with respect to the interchangeability of codesignative names is similar. True, such a principle, when combined with our normal disquotational judgments of belief, leads to straightforward absurdities. But we will see that the 'same' absurdities can be derived by replacing the interchangeability principle by our normal practices of translation and disquotation, or even by disquotation alone.

The particular principle stated here gives just one particular way of 'formalizing' our normal inferences from explicit affirmation or assent to belief; other ways of doing it are possible. It is undeniable that we *do* infer, from a normal Englishman's sincere affirmation of 'God exists' or 'London is pretty,' that he believes, respectively, that God exists or that London is pretty; and that we would make the same inferences from a Frenchman's affirmation of '*Dieu existe*' or '*Londres est jolie*.' Any principles that would justify such inferences are sufficient for the next section. It will be clear that the particular principles stated in the present section are sufficient, but in the next section the problem will be presented informally in terms of our inferences from foreign or domestic assertion to belief.

### III. THE PUZZLE

Here, finally(!), is the puzzle. Suppose Pierre is a normal French speaker who lives in France and speaks not a word of English or of any other language except French. Of course he has heard of that famous distant city, London (which he of course calls '*Londres*') though he himself has never left France. On the basis of what he has heard of London, he is inclined to think that it is pretty. So he says, in French, "*Londres est jolie*."

On the basis of his sincere French utterance, we will conclude:



- (4) Pierre believes that London is pretty.

I am supposing that Pierre satisfies all criteria for being a normal French speaker, in particular, that he satisfies whatever criteria we usually use to judge that a Frenchman (correctly) uses '*est jolie*' to attribute pulchritude and uses '*Londres*' — standardly — as a name of London.

Later, Pierre, through fortunate or unfortunate vicissitudes, moves to England, in fact to London itself, though to an unattractive part of the city with fairly uneducated inhabitants. He, like most of his neighbors, rarely ever leaves this part of the city. None of his neighbors know any French, so he must learn English by 'direct method,' without using any translation of English into French: by talking and mixing with the people he eventually begins to pick up English. In particular, everyone speaks of the city, 'London,' where they all live. Let us suppose for the moment — though we will see below that this is not crucial — that the local population are so uneducated that they know few of the facts that Pierre heard about London in France. Pierre learns from them everything they know about London, but there is little overlap with what he heard before. He learns, of course — speaking English — to call the city he lives in 'London.' Pierre's surroundings are, as I said, unattractive, and he is unimpressed with most of the rest of what he happens to see. So he is inclined to assent to the English sentence:

- (5) London is not pretty.

He has *no* inclination to assent to:

- (6) London is pretty.

Of course he does not for a moment withdraw his assent from the French sentence, "*Londres est jolie*"; he merely takes it for granted that the ugly city in which he is now stuck is distinct from the enchanting city he heard about in France. But he has no inclination to change his mind for a moment about the city he stills calls '*Londres*.'

This, then, is the puzzle. If we consider Pierre's past background as a French speaker, his entire linguistic behavior, on the same basis as we would draw such a conclusion about many of his countrymen, supports the conclusion (4) above) that he believes that London is pretty. On the other hand, after Pierre lived in London for some time, he did not differ from his neighbors — his French background aside — either in his knowledge of English or in his command of the relevant facts of local geography. His English

vocabulary differs little from that of his neighbors. He, like them, rarely ventures from the dismal quarter of the city in which they all live. He, like them, knows that the city he lives in is called 'London' and knows a few other facts. Now Pierre's neighbors would surely be said to use 'London' as a name for London and to speak English. Since, as an English speaker, he does not differ at all from them, we should say the same of him. But then, on the basis of his sincere assent to (5), we should conclude:

- (7) Pierre believes that London is not pretty.

How can we describe this situation? It seems undeniable that Pierre *once* believed that London is pretty — at least before he learned English. For at that time, he differed not at all from countless numbers of his countrymen, and we would have exactly the same grounds to say of him as of any of them that he believes that London is pretty: if any Frenchman who was both ignorant of English and never visited London believed that London is pretty, Pierre did. Nor does it have any plausibility to suppose, because of his later situation *after* he learns English, that Pierre should *retroactively* be judged *never* to have believed that London is pretty. To allow such *ex post facto* legislation would, as long as the future is uncertain, endanger our attributions of belief to *all* monolingual Frenchmen. We would be forced to say that Marie, a monolingual who firmly and sincerely asserts, "*Londres est jolie,*" may or may not believe that London is pretty depending on the *later* vicissitudes of her career (if later she learns English and . . . , . . .). No: Pierre, like Marie, believed that London is pretty when he was monolingual.

Should we say that Pierre, now that he lives in London and speaks English, no longer believes that London is pretty? Well, unquestionably Pierre *once* believed that London is pretty. So we would be forced to say that Pierre has *changed his mind, has given up his previous belief*. But has he really done so? Pierre is very set in his ways. He reiterates, with vigor, every assertion he has ever made in French. He says he has not changed his mind about anything, has *not* given up any belief. Can we say he is wrong about this? If we did not have the story of his living in London and his English utterances, on the basis of his normal command of French we would be *forced* to conclude that he *still* believes that London is pretty. And it does seem that this is correct. Pierre has neither changed his mind nor given up any belief he had in France.

Similar difficulties beset any attempt to deny him his new belief. His French past aside, he is just like his friends in London. Anyone else, growing

up in London with the same knowledge and beliefs that he expresses in England, we would undoubtedly judge to believe that London is not pretty. Can Pierre's French past nullify such a judgment? Can we say that Pierre, because of his French past, does not believe that (5)? Suppose an electric shock wiped out all his memories of the French language, what he learned in France, and his French past. He would then be *exactly* like his neighbors in London. He would have the *same* knowledge, beliefs, and linguistic capacities. We then presumably would be forced to say that Pierre believes that London is ugly if we say it of his neighbors. But surely no shock that *destroys* part of Pierre's memories and knowledge can *give* him a new belief. If Pierre believes (5) *after* the shock, he believed it before, despite his French language and background.

If we would deny Pierre, in his bilingual stage, his belief that London is pretty *and* his belief that London is not pretty, we combine the difficulties of both previous options. We still would be forced to judge that Pierre once believed that London is pretty but does no longer, in spite of Pierre's own sincere denial that he has lost any belief. We also must worry whether Pierre would *gain* the belief that London is not pretty if he totally forgot his French past. The option does not seem very satisfactory.

So now it seems that we must respect both Pierre's French utterances and their English counterparts. So we must say that Pierre has contradictory beliefs, that he believes that London is pretty *and* he believes that London is not pretty. But there seem to be insuperable difficulties with this alternative as well. 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.

We can shed more light on this if we change the case. Suppose that, in France, Pierre, instead of affirming "*Londres est jolie*," had affirmed, more cautiously, "*Si New York est jolie, Londres est jolie aussi*," so that he believed that *if* New York is pretty, so is London. Later Pierre moves to London,

learns English as before, and says (in English) "London is not pretty." So he now believes, further, that London is *not* pretty. Now from the two premises, both of which appears to be among his beliefs (a) If New York is pretty, London is, and (b) London is not pretty, Pierre should be able to deduce by *modus tollens* that New York is not pretty. But no matter how great Pierre's logical acumen may be, *he cannot in fact make any such deduction, as long as he supposes that 'Londres' and 'London' may name two different cities.* If he *did* draw such a conclusion, he would be guilty of a fallacy.

Intuitively, he may well suspect that New York is pretty, and just this suspicion may lead him to suppose that 'Londres' and 'London' probably name distinct cities. Yet, if we follow our normal practice of reporting the beliefs of French and English speakers, *Pierre has available to him (among his beliefs) both the premises of a modus tollens argument that New York is not pretty.*

Again, we may emphasize Pierre's *lack* of belief instead of his belief. Pierre, as I said, has no disposition to assent to (6). Let us concentrate on this, ignoring his disposition to assent to (5). In fact, if we wish we may change the case: Suppose Pierre's neighbors think that since they rarely venture outside their own ugly section, they have no right to any opinion as to the pulchritude of the whole city. Suppose Pierre shares their attitude. Then, judging by his failure to respond affirmatively to "London is pretty," we may judge, from Pierre's behavior as an *English* speaker, that he lacks the belief that London is pretty: never mind whether he disbelieves it, as before, or whether, as in the modified story, he insists that he has no firm opinion on the matter.

Now (using the *strengthened* disquotational principle), we can derive a contradiction, not merely in Pierre's judgments, but in our own. For on the basis of his behavior as an English speaker, we concluded that he does *not* believe that London is pretty (that is, that it is not the case that he believes that London is pretty). But on the basis of his behavior as a *French* speaker, we must conclude that he *does* believe that London is pretty. This is a contradiction.<sup>28</sup>

We have examined four possibilities for characterizing Pierre while he is in London: (a) that at that time we no longer respect his French utterance ('*Londres est jolie*'), that is that we no longer ascribe to him the corresponding belief; (b) that we do not respect his English utterance (or lack of utterance); (c) that we respect neither; (d) that we respect both. Each possibility seems to lead us to say something either plainly false or even downright contradic-

tory. Yet the possibilities appear to be logically exhaustive. This, then, is the paradox.

I have no firm belief as to how to solve it. But beware of one source of confusion. It is no solution in itself to observe that some *other* terminology, which evades the question whether Pierre believes that London is pretty, may be sufficient to state all the relevant facts. I am fully aware that complete and straightforward descriptions of the situation are possible and that in this sense there is no paradox. Pierre is disposed to sincere assent to '*Londres est jolie*' but not to 'London is pretty.' He uses French normally, English normally. Both with '*Londres*' and 'London' he associates properties sufficient to determine that famous city, but he does not realize that they determine a single city. (And his uses of '*Londres*' and 'London' are historically (causally) connected with the same single city, though he is unaware of that.) We may even give a rough statement of his beliefs. He believes that the city he calls '*Londres*' is pretty, that the city he calls 'London' is not. No doubt other straightforward descriptions are possible. No doubt some of these are, in a certain sense, *complete* descriptions of the situation.

But none of this answers the original question. Does Pierre, or does he not, believe that London is pretty? I know of no answer to *this* question that seems satisfactory. It is no answer to protest that, in some *other* terminology, one can state 'all the relevant facts.'

To reiterate, this is the puzzle: Does Pierre, or does he not, believe that London is pretty? It is clear that our normal criteria for the attribution of belief lead, when applied to *this* question, to paradoxes and contradictions. One set of principles adequate to many ordinary attributions of belief, but which leads to paradox in the present case, was stated in Section 2; and other formulations are possible. As in the case of the logical paradoxes, the present puzzle presents us with a problem for customarily accepted principles and a challenge to formulate an acceptable set of principles that does not lead to paradox, is intuitively sound, and supports the inferences we usually make. Such a challenge cannot be met simply by a description of Pierre's situation that evades the question whether he believes that London is pretty.

One aspect of the presentation may misleadingly suggest the applicability of Frege-Russellian ideas that each speaker associates his own description or properties to each name. For as I just set up the case Pierre learned one set of facts about the so-called '*Londres*' when he was in France, and *another* set of facts about 'London' in England. Thus it may appear that 'what's really going on' is that Pierre believes that *the city* satisfying *one* set of properties *is*

pretty, while he believes that *the city* satisfying *another* set of properties *is not* pretty.

As we just emphasized, the phrase ‘what’s really going on’ is a danger signal in discussions of the present paradox. The conditions stated may — let us concede for the moment — describe ‘what’s really going on.’ But they do not resolve the problem with which we began, that of the behavior of names in belief contexts: Does Pierre, or does he not, believe that London (not the city satisfying such-and-such descriptions, but *London*) is pretty? No answer has yet been given.

Nevertheless, these considerations may appear to indicate that descriptions, or associated properties, are highly relevant somehow to an ultimate solution, since at this stage it appears that the entire puzzle arises from the fact that Pierre originally associated different identifying properties with ‘London’ and ‘*Londres*.’ Such a reaction may have some force even in the face of the now fairly well-known arguments against ‘identifying descriptions’ as in any way ‘defining,’ or even ‘fixing the reference’ of names. But in fact the special features of the case, as I set it out, are misleading. The puzzle can arise even if Pierre associates exactly the same identifying properties with both names.

First, the considerations mentioned above in connection with ‘Cicero’ and ‘Tully’ establish this fact. For example, Pierre may well learn, in France, ‘*Platon*’ as the name of a major Greek philosopher, and later, in England, learns ‘Plato’ with the same identification. Then the same puzzle can arise: Pierre may have believed, when he was in France and was monolingual in French, that Plato was bald (he would have said, “*Platon était chauve*”), and later conjecture, in English, “Plato was not bald,” thus indicating that he believes or suspects that Plato was *not* bald. He need only suppose that, in spite of the similarity of their names, the man he calls ‘*Platon*’ and the man he calls ‘Plato’ were two distinct major Greek philosophers. In principle, the same thing could happen with ‘London’ and ‘*Londres*.’

Of course, most of us learn a *definite* description about London, say ‘the largest city in England.’ Can the puzzle still arise? It is noteworthy that the puzzle can still arise even if Pierre associates to ‘*Londres*’ and to ‘London’ *exactly* the same *uniquely identifying* properties. How can this be? Well, suppose that Pierre believes that London is the largest city in (and capital of) England, that it contains Buckingham Palace, the residence of the Queen of England, and he believes (correctly) that these properties, conjointly, uniquely identify the city. (In this case, it is best to suppose that he has never

seen London, or even England, so that he uses *only* these properties to identify the city. Nevertheless, he has learned English by 'direct method.' These uniquely identifying properties he comes to associate with 'London' after he learned English, and he expresses the appropriate beliefs about 'London' in English. Earlier, when he spoke nothing but French, however, he associated *exactly* the same uniquely identifying properties with '*Londres*.' He believed that '*Londres*,' as he called it, could be uniquely identified as the capital of England, that it contained Buckingham Palace, that the Queen of England lived there, etc. Of course he expressed these beliefs, like most monolingual Frenchmen, in French. In particular, he used '*Angleterre*' for England, '*le Palais de Buckingham*' (pronounced '*Bookeengam*'!) for Buckingham Palace, and '*la Reine d'Angleterre*' for the Queen of England. But if any Frenchman who speaks no English can ever be said to associate *exactly* the properties of being the capital of England etc., with the name '*Londres*,' Pierre in his monolingual period did so.

When Pierre becomes a bilingual, *must* he conclude that 'London' and '*Londres*' name the same city, because he defined each by the same uniquely identifying properties?

Surprisingly, no! Suppose Pierre had affirmed, '*Londres est jolie*.' If Pierre has any reason — even just a 'feeling in his bones,' or perhaps exposure to a photograph of a miserable area which he was told (in English) was part of 'London' — to maintain 'London is not pretty,' he need not contradict himself. He need only conclude that 'England' and '*Angleterre*' name two different countries, that 'Buckingham Palace' and '*le Palais de Buckingham*' (recall the pronunciation!), name two different palaces, and so on. Then he can maintain *both* views without contradiction, and regard *both* properties as uniquely identifying.

The fact is that the paradox reproduces itself on the level of the 'uniquely identifying properties' that description theorists have regarded as 'defining' proper names (and *a fortiori*, as fixing their references). Nothing is more reasonable than to suppose that if two names, *A* and *B*, and a single set of properties, *S*, are such that a certain speaker believes that the referent of *A* uniquely satisfies all of *S* and that the referent of *B* also uniquely satisfies all of *S*, then that speaker is committed to the belief that *A* and *B* have the same reference. In fact, the identity of the referents of *A* and *B* is an easy *logical consequence* of the speaker's beliefs.

From this fact description theorists concluded that names can be regarded as synonymous, and hence interchangeable *salva veritate* even in belief con-

texts, provided that they are 'defined' by the same uniquely identifying properties.

We have already seen that there is a difficulty in that the set *S* of properties need not in fact be uniquely identifying. But in the present paradoxical situation there is a surprising difficulty even if the supposition of the description theorist (that the speaker believes that *S* is uniquely fulfilled) in fact holds. For, as we have seen above, Pierre is in no position to draw ordinary logical consequences from the conjoint set of what, when we consider him separately as a speaker of English and as a speaker of French, we would call his beliefs. He cannot infer a contradiction from his separate beliefs that London is pretty and that London is not pretty. Nor, in the modified situation above, would Pierre make a normal *modus tollens* inference from his beliefs that London is not pretty and that London is pretty if New York is. Similarly here, if we pay attention only to Pierre's behavior as a French speaker (and at least in his monolingual days he was no different from any other Frenchmen), Pierre satisfies all the normal criteria for believing that '*Londres*' has a referent uniquely satisfying the properties of being the largest city in England, containing Buckingham Palace, and the like. (If Pierre did not hold such beliefs, no Frenchman *ever* did.) Similarly, on the basis of his (later) beliefs expressed in English, Pierre also believes that the referent of 'London' uniquely satisfies these same properties. But Pierre cannot combine the two beliefs into a single set of beliefs from which he can draw the normal conclusion that 'London' and '*Londres*' must have the same referent. (Here the trouble comes not from 'London' and '*Londres*' but from 'England' and '*Angleterre*' and the rest.) Indeed, if he *did* draw what would appear to be the normal conclusion in this case and any of the other cases, Pierre would in fact be guilty of a logical fallacy.

Of course the description theorist could hope to eliminate the problem by 'defining' '*Angleterre*,' 'England,' and so on by appropriate descriptions also. Since in principle the problem may rear its head at the next 'level' and at each subsequent level, the description theorist would have to believe that an 'ultimate' level can eventually be reached where the defining properties are 'pure' properties not involving proper names (nor natural kind terms or related terms, see below!). I know of no convincing reason to suppose that such a level can be reached in any plausible way, or that the properties can continue to be uniquely identifying if one attempts to eliminate all names and related devices.<sup>29</sup> Such speculation aside, the fact remains that Pierre, judged by the *ordinary* criteria for such judgments, *did* learn both '*Londres*' and



'London' by *exactly* the same set of identifying properties; yet the puzzle remains even in this case.

Well, then, is there any way out of the puzzle? Aside from the principles of disquotation and translation, only our normal practice of translation of French into English has been used. Since the principles of disquotation and translation seem self-evident, we may be tempted to blame the trouble on the translation of '*Londres est jolie*' as 'London is pretty,' and ultimately, then, on the translation of '*Londres*' as 'London.'<sup>30</sup> Should we, perhaps, permit ourselves to conclude that '*Londres*' should not, 'strictly speaking' be translated as 'London'? Such an expedient is, of course, desperate: the translation in question is a standard one, learned by students together with other standard translations of French into English. Indeed, '*Londres*' is, in effect, introduced into French as the French version of 'London.'

Since our backs, however, are against the wall, let us consider this desperate and implausible expedient a bit further. If '*Londres*' is *not* a correct French version of the English 'London,' under what circumstances can proper names be translated from one language to another?

Classical description theories suggest the answer: Translation, strictly speaking, is between idiolects; a name in one idiolect can be translated into another when (and only when) the speakers of the two idiolects associate the same uniquely identifying properties with the two names. We have seen that any such proposed restriction, not only fails blatantly to fit our normal practices of translation and indirect discourse reportage, but does not even appear to block the paradox.<sup>31</sup>

So we still want a suitable restriction. Let us drop the references to idiolects and return to '*Londres*' and 'London' as names in French and English, respectively — the languages of two communities. If '*Londres*' is not a correct French translation of 'London,' could any other version do better? Suppose I introduced another word into French, with the stipulation that *it* should always be used to translate 'London.' Would not the same problem arise for this word as well? The only feasible solution in this direction is the most drastic: decree that no sentence containing a name can be translated except by a sentence containing the phonetically identical name. Thus when Pierre asserts '*Londres est jolie*,' we English speakers can at best conclude, if anything: Pierre believes that *Londres* is pretty. Such a conclusion is, of course, not expressed in English, but in a word salad of English and French; on the view now being entertained, we cannot state Pierre's belief in *English* at all.<sup>32</sup> Similarly, we would have to say: Pierre believes that *Angleterre* is a

monarchy, Pierre believes that *Platon* wrote dialogues, and the like.<sup>33</sup>

This 'solution' appears at first to be effective against the paradox, but it is drastic. What is it about sentences containing names that makes them — a substantial class — intrinsically untranslatable, express beliefs that cannot be reported in any other language? At best, to report them in the other language, one is forced to use a word salad in which names from the one language are imported into the other. Such a supposition is both contrary to our normal practice of translation and very implausible on its face.

Implausible though it is, there is at least this much excuse for the 'solution' at this point. Our normal practice with respect to some famous people and especially for geographical localities is to have different names for them in different languages, so that in translating sentences we translate the names. But for a large number of names, especially names of people, this is not so: the person's name is used in the sentences of all languages. At least the restriction in question merely urges us to mend our ways by doing *always* what we presently do *sometimes*.

But the really drastic character of the proposed restriction comes out when we see how far it may have to extend. In "Naming and Necessity" I suggested that there are important analogies between proper names and natural kind terms, and it seems to me that the present puzzle is one instance where the analogy will hold. Putnam, who has proposed views on natural kinds similar to my own in many respects, stressed this extension of the puzzle in his comments at the Conference. Not that the puzzle extends to all translations from English to French. At the moment, at least, it seems to me that Pierre, if he learns English and French separately, without learning any translation manual between them, *must* conclude, if he reflects enough, that 'doctor' and '*médecin*,' and '*heureux*' and 'happy,' are synonymous, or at any rate, coextensive;<sup>34</sup> any potential paradox of the present kind for these word pairs is thus blocked. But what about '*lapin*' and 'rabbit,' or 'beech' and '*hêtre*'? We may suppose that Pierre is himself neither a zoologist nor a botanist. He has learned each language in its own country and the examples he has been shown to illustrate '*les lapins*' and 'rabbits,' 'beeches' and '*les hêtres*' are distinct. It thus seems to be possible for him to suppose that '*lapin*' and 'rabbit,' or 'beech' and '*hêtre*,' denote distinct but superficially similar kinds or species, even though the differences may be indiscernible to the untrained eye. (This is especially plausible if, as Putnam supposes, an English speaker — for example, Putnam himself — who is not a botanist may use 'beech' and 'elm' with their normal (distinct) meanings, even though he cannot himself

distinguish the two trees.<sup>35</sup> Pierre may quite plausibly be supposed to wonder whether the trees which in France he called '*les hêtres*' were beeches or elms, even though as a speaker of French he satisfies all usual criteria for using '*les hêtres*' normally. If beeches and elms will not serve, better pairs of ringers exist that cannot be told apart except by an expert.) Once Pierre is in such a situation, paradoxes analogous to the one about London obviously can arise for rabbits and beeches. Pierre could affirm a French statement with '*lapin*,' but deny its English translation with 'rabbit.' As above, we are hard-pressed to say what Pierre *believes*. We were considering a 'strict and philosophical' reform of translation procedures which proposed that foreign proper names should always be appropriated rather than translated. Now it seems that we will be forced to do the same with all words for natural kinds. (For example, on price of paradox, one must not translate '*lapin*' as 'rabbit'!) No longer can the extended proposal be defended, even weakly, as 'merely' universalizing what we already do sometimes. It is surely too drastic a change to retain any credibility.<sup>36</sup>

There is yet another consideration that makes the proposed restriction more implausible: Even this restriction does not really block the paradox. Even if we confine ourselves to a single language, say English, and to phonetically identical tokens of a single name, we can still generate the puzzle. Peter (as we may as well say now) may learn the name 'Paderewski' with an identification of the person named as a famous pianist. Naturally, having learned this, Peter will assent to "Paderewski had musical talent," and *we* can infer — using 'Paderewski,' as we usually do, to name the Polish musician and statesman:

- (8) Peter believes that Paderewski had musical talent.

Only the disquotational principle is necessary for our inference; no translation is required. Later, in a different circle, Peter learns of someone called 'Paderewski' who was a Polish nationalist leader and Prime Minister. Peter is skeptical of the musical abilities of politicians. He concludes that probably two people, approximate contemporaries no doubt, were both named 'Paderewski.' Using 'Paderewski' as a name for the *statesman*, Peter assents to, "Paderewski had no musical talent." Should we infer, by the disquotational principle,

- (9) Peter believes that Paderewski had no musical talent

or should we not? If Peter had not had the past history of learning the name

'Paderewski' in another way, we certainly would judge him to be using 'Paderewski' in a normal way, with the normal reference, and we would infer (9) by the disquotational principle. The situation is parallel to the problem with Pierre and London. Here, however, no restriction that names should not be translated, but should be phonetically repeated in the translation, can help us. Only a single language and a single name are involved. If any notion of translation is involved in this example, it is homophonic translation. Only the disquotational principle is used explicitly.<sup>37</sup> (On the other hand, the original 'two languages' case had the advantage that it would apply even if we spoke languages in which all names must denote uniquely and unambiguously.) The restriction that names must not be translated is thus ineffective, as well as implausible and drastic.

I close this section with some remarks on the relation of the present puzzle to Quine's doctrine of the 'indeterminacy of translation,' with its attendant repudiation of intensional idioms of 'propositional attitude' such as belief and even indirect quotation. To a sympathizer with these doctrines the present puzzle may well seem to be just more grist for a familiar mill. The situation of the puzzle seems to lead to a breakdown of our normal practices of attributing belief and even of indirect quotation. No obvious paradox arises if we describe the same situation in terms of Pierre's sincere assent to various sentences, together with the conditions under which he has learned the name in question. Such a description, although it does not yet conform to Quine's strict behavioristic standards, fits in well with his view that in some sense direct quotation is a more 'objective' idiom than the propositional attitudes. Even those who, like the present writer, do not find Quine's negative attitude to the attitudes completely attractive must surely acknowledge this.

But although sympathizers with Quine's view can use the present examples to support it, the differences between these examples and the considerations Quine adduces for his own skepticism about belief and translation should not escape us. Here we make no use of hypothetical exotic systems of translation differing radically from the usual one, translating '*lapin*,' say, as 'rabbit stage' or 'undetached part of a rabbit.' The problem arises entirely within our usual and customary system of translation of French into English; in one case, the puzzle arose even within English alone, using at most 'homophonic' translation. Nor is the problem that many different interpretations or translations fit our usual criteria, that, in Davidson's phrase,<sup>38</sup> there is more than one 'way of getting it right.' The trouble here is not that many views as to Pierre's beliefs get it right, but that they all definitely get it *wrong*. A straightforward

application of the principles of translation and disquotation to all Pierre's utterances, French and English, yields the result that Pierre holds inconsistent beliefs, that logic alone should teach him that one of his beliefs is false. Intuitively, this is plainly incorrect. If we refuse to apply the principles to his French utterances at all, we would conclude that Pierre never believed that London is pretty, even though, before his unpredictable move, he was like any other monolingual Frenchman. This is absurd. If we refuse to ascribe the belief in London's pulchritude only after Pierre's move to England, we get the counterintuitive result that Pierre has changed his mind, and so on. But we have surveyed the possibilities above: the point was not that they are 'equally good,' but that all are *obviously wrong*. If the puzzle is to be used as an argument for a Quinean position, it is an argument of a fundamentally different kind from those given before. And even Quine, if he wishes to incorporate the notion of belief even into a 'second level' of canonical notation,<sup>39</sup> must regard the puzzle as a real problem.

The alleged indeterminacy of translation and indirect quotation causes relatively little trouble for such a scheme for belief; the embarrassment it presents to such a scheme is, after all, one of riches. But the present puzzle indicates that the usual principles we use to ascribe beliefs are apt, in certain cases, to lead to contradiction, or at least, patent falsehoods. So it presents a problem for any project, Quinean or other, that wishes to deal with the 'logic' of belief on any level.<sup>40</sup>

#### IV. CONCLUSION

What morals can be drawn? The primary moral — quite independent of any of the discussion of the first two sections — is that the puzzle *is* a puzzle. As any theory of truth must deal with the Liar Paradox, so any theory of belief and names must deal with this puzzle.

But our theoretical starting point in the first two sections concerned proper names and belief. Let us return to Jones, who assents to "Cicero was bald" and to "Tully was not bald." Philosophers, using the disquotational principle, have concluded that Jones believes that Cicero was bald but that Tully was not. Hence, they have concluded, since Jones does not have contradictory beliefs, belief contexts are not 'Shakespearean' in Geach's sense: codesignative proper names are not interchangeable in these contexts *salva veritate*.<sup>41</sup>

I think the puzzle about Pierre shows that the simple conclusion was unwarranted. Jones' situation strikingly resembles Pierre's. A proposal that

'Cicero' and 'Tully' are interchangeable amounts roughly to a homophonic 'translation' of English into itself in which 'Cicero' is mapped into 'Tully' and *vice versa*, while the rest is left fixed. Such a 'translation' can, indeed, be used to obtain a paradox. But should the problem be blamed on this step? Ordinarily we would suppose without question that sentences in French with 'Londres' should be translated into English with 'London.' Yet the same paradox results when we apply this translation too. We have seen that the problem can even arise with a single name in a single language, and that it arises with natural kind terms in two languages (or one: see below).

Intuitively, Jones' assent to both 'Cicero was bald' and 'Tully was not bald' arises from sources of just the same kind as Pierre's assent to both 'Londres est jolie' and 'London is not pretty.'

It is wrong to blame unpalatable conclusions about Jones on substitutivity. The reason does not lie in any specific fallacy in the argument but rather in the nature of the realm being entered. Jones's case is just like Pierre's: both are in an area where our normal practices of attributing belief, based on the principles of disquotation and translation or on similar principles, are questionable.

It should be noted in this connection that the principles of disquotation and translation can lead to 'proofs' as well as 'disproofs' of substitutivity in belief contexts. In Hebrew there are two names for Germany, transliteratable roughly as 'Ashkenaz' and 'Germaniah' — the first of these may be somewhat archaic. When Hebrew sentences are translated into English, both become 'Germany.' Plainly a normal Hebrew speaker analogous to Jones might assent to a Hebrew sentence involving 'Ashkenaz' while dissenting from its counterpart with 'Germaniah.' So far there is an argument *against* substitutivity. But there is also an argument *for* substitutivity, based on the principle of translation. Translate a Hebrew sentence involving 'Ashkenaz' into English, so that 'Ashkenaz' goes into 'Germany.' Then retranslate the result into Hebrew, this time translating 'Germany' as 'Germaniah.' By the principle of translation, both translations preserve truth value. So: the truth value of any sentence of Hebrew involving 'Ashkenaz' remains the same when 'Ashkenaz' is replaced by 'Germaniah' — a 'proof' of substitutivity! A similar 'proof' can be provided wherever there are two names in one language, and a normal practice of translating both indifferently into a single name of another language.<sup>42</sup> (If we combine the 'proof' and 'disproof' of substitutivity in this paragraph, we could get yet another paradox analogous to Pierre's: our Hebrew speaker both believes, and disbelieves, that Germany is pretty. Yet

no amount of pure logic or semantic introspection suffices for him to discover his error.)

Another consideration, regarding natural kinds: Previously we pointed out that a bilingual may learn ‘*lapin*’ and ‘rabbit’ normally in each respective language yet wonder whether they are one species or two, and that this fact can be used to generate a paradox analogous to Pierre’s. Similarly, a speaker of *English* alone may learn ‘furze’ and ‘gorse’ normally (separately), yet wonder whether these are the same, or resembling kinds. (What about ‘rabbit’ and ‘hare’?) It would be easy for such a speaker to assent to an assertion formulated with ‘furze’ but withhold assent from the corresponding assertion involving ‘gorse.’ The situation is quite analogous to that of Jones with respect to ‘Cicero’ and ‘Tully.’ Yet ‘furze’ and ‘gorse,’ and other pairs of terms for the same natural kind, are normally thought of as *synonyms*.

The point is *not*, of course, that codesignative proper names *are* interchangeable in belief contexts *salva veritate*, or that they *are* interchangeable in simple contexts even *salva significatione*. The point is that the absurdities that disquotation plus substitutivity would generate are exactly paralleled by absurdities generated by disquotation plus translation, or even ‘disquotation alone’ (or: disquotation plus homophonic translation). Also, though our naive practice may lead to ‘disproofs’ of substitutivity in certain cases, it can also lead to ‘proofs’ of substitutivity in some of these same cases, as we saw two paragraphs back. 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.<sup>43</sup>

Of course nothing in these considerations prevents us from observing that Jones can sincerely assert both “Cicero is bald” and “Tully is not bald,” even though he is a normal speaker of English and uses ‘Cicero’ and ‘Tully’ in normal ways, and with the normal referent. Pierre and the other paradoxical cases can be described similarly. (For those interested in one of my own doctrines, we can still say that there was a time when men were in no epistemic position to assent to ‘Hesperus is Phosphorus’ for want of empirical information, but it nevertheless expressed a necessary truth.)<sup>44</sup> But it is no surprise that quoted contexts fail to satisfy a substitutivity principle within the quotation marks. And, in our *present* state of clarity about the problem, we are in

no position to apply a disquotational principle to these cases, nor to judge when two such sentences do, or do not, express the same 'proposition.'

Nothing in the discussion impugns the conventional judgment that belief contexts are 'referentially opaque,' if 'referential opacity' is construed so that failure of coreferential *definite descriptions* to be interchangeable *salva veritate* is sufficient for referential opacity. No doubt Jones can believe that the number of planets is even, without believing that the square of three is even, if he is under a misapprehension about the astronomical, but not the arithmetical facts. The question at hand was whether belief contexts were 'Shakespearean,' not whether they were 'referentially transparent.' (Modal contexts, in my opinion, are 'Shakespearean' but 'referentially opaque.')

<sup>45</sup>

Even were we inclined to rule that belief contexts are not Shakespearean, it would be implausible at present to use the phenomenon to support a Frege-Russellian theory that names have descriptive 'senses' through 'uniquely identifying properties.' There are the well-known arguments against description theories, independent of the present discussion; there is the implausibility of the view that difference in names is difference in idiolect; and finally, there are the arguments of the present paper that differences of associated properties do not explain the problems in any case. Given these considerations, and the cloud our paradox places over the notion of 'content' in this area, the relation of substitutivity to the dispute between Millian and Fregean conclusions is not very clear.

We repeat our conclusions: Philosophers have often, basing themselves on Jones' and similar cases, supposed that it goes virtually without saying that belief contexts are not 'Shakespearean.' I think that, at present, such a definite conclusion is unwarranted. Rather Jones' case, like Pierre's, lies in an area where our normal apparatus for the ascription of belief is placed under the greatest strain and may even break down. There is even less warrant at the present time, in the absence of a better understanding of the paradoxes of this paper, for the use of alleged failures of substitutivity in belief contexts to draw any significant theoretical conclusion about proper names. Hard cases make bad law.<sup>46</sup>

*Princeton University*



## NOTES

<sup>1</sup> "Naming and Necessity," in: *The Semantics of Natural Languages*, D. Davidson and G. Harman (eds.), Dordrecht, Reidel, 1971, pp. 253–355 and 763–769. (Also forthcoming as a separate monograph, pub. Basil Blackwell.) "Identity and Necessity" in: *Identity and Individuation*, M. Munitz (ed.), New York University Press, 1971, pp. 135–164. Acquaintance with these papers is not a prerequisite for understanding the central puzzle of the present paper, but is helpful for understanding the theoretical background.

<sup>2</sup> Frege gives essentially this example as the second footnote of "On Sense and Reference." For the "Who is . . . ?" to be applicable one must be careful to elicit from one's informant properties that he regards as defining the name and determining the referent, not mere well-known facts about the referent. (Of course this distinction may well seem fictitious, but it is central to the original Frege–Russell theory.)

<sup>3</sup> For convenience Russell's terminology is assimilated to Frege's. Actually, regarding genuine or 'logically proper' names, Russell is a strict Millian: 'logically proper names' simply refer (to immediate objects of acquaintance). But, according to Russell, what are ordinarily called 'names' are not genuine, logically proper names, but disguised definite descriptions. Since Russell also regards definite descriptions as in turn disguised notation, he does not associate any 'senses' with descriptions, since they are not genuine singular terms. When all disguised notation is eliminated, the only singular terms remaining are logically proper names, for which no notion of 'sense' is required. When we speak of Russell as assigning 'senses' to names, we mean ordinary names and for convenience we ignore his view that the descriptions abbreviating them ultimately disappear on analysis.

On the other hand, the explicit doctrine that names are abbreviated definite descriptions is due to Russell. Michael Dummett, in his recent *Frege* (Duckworth and Harper and Row, 1973, pp. 110–111) denies that Frege held a description theory of senses. Although as far as I know Frege indeed makes no explicit statement to that effect, his examples of names conform to the doctrine, as Dummett acknowledges. Especially his 'Aristotle' example is revealing. He defines 'Aristotle' just as Russell would; it seems clear that in the case of a famous historical figure, the 'name' is indeed to be given by answering, in a uniquely specifying way, the 'who is' question. Dummett himself characterizes a sense as a "criterion . . . such that the referent of the name, if any, is whatever object satisfies that criterion." Since presumably the satisfaction of the criterion must be unique (so a unique referent is determined), doesn't this amount to defining names by unique satisfaction of properties, *i.e.*, by descriptions? *Perhaps* the point is that the property in question need not be expressible by a usual predicate of English, as might be plausible if the referent is one of the speaker's acquaintances rather than a historical figure. But I doubt that even Russell, father of the explicitly formulated description theory, ever meant to require that the description must always be expressible in (unsupplemented) English.

In any event, the philosophical community has generally understood Fregean senses in terms of descriptions, and we deal with it under this usual understanding. For present purposes this is more important than detailed historical issues. Dummett acknowledges (p. 111) that few substantive points are affected by his (allegedly) broader interpretation of Frege; and it would not seem to be relevant to the problems of the present paper.

<sup>4</sup> See Frege's footnote in "On Sense and Reference" mentioned in note 2 above and

especially his discussion of 'Dr. Gustav Lauben' in "*Der Gedanke*." (In the recent Geach-Stoothoff translation, "Thoughts," *Logical Investigations*, Oxford, Blackwell, 1977, pp. 11–12).

<sup>5</sup> Russell, as a Millian with respect to genuine names, accepts this argument with respect to 'logically proper names.' For example — taking for the moment 'Cicero' and 'Tully' as 'logically proper names,' Russell would hold that if I judge that Cicero admired Tully, I am related to Cicero, Tully, and the admiration relation in a certain way: Since Cicero *is* Tully, I am related in exactly the same way to Tully, Cicero, and admiration; therefore I judge that Tully admired Cicero. Again, if Cicero *did* admire Tully, then according to Russell a single fact corresponds to all of 'Cicero admired Tully,' 'Cicero admired Cicero,' etc. Its constituent (in addition to admiration) is the man Cicero, taken, so to speak, twice.

Russell thought that 'Cicero admired Tully' and 'Tully admired Cicero' are in fact obviously not interchangeable. For him, this was one argument that 'Cicero' and 'Tully' are *not* genuine names, and that the Roman orator is no constituent of propositions (or 'facts,' or 'judgments') corresponding to sentences containing the name.

<sup>6</sup> Given the arguments of Church and others, I do not believe that the formal mode of speech is synonymous with other formulations. But it can be used as a rough way to convey the idea of scope.

<sup>7</sup> It may well be argued that the Millian view implies that proper names are *scopeless* and that for them the *de dicto-de re* distinction vanishes. This view has considerable plausibility (my own views on rigidity will imply something like this for *modal* contexts), but it need not be argued here either way: *de re* uses are simply not treated in the present paper.

Christopher Peacocke ("Proper Names, Reference, and Rigid Designation," in: *Meaning, Reference, and Necessity*, S. Blackburn (ed.), Cambridge, 1975; see Section I), uses what amounts to the equivalence of the *de dicto-de re* constructions in *all* contexts (or, put alternatively, the lack of such a distinction) to characterize the notion of rigid designation. I agree that for *modal* contexts, this is (roughly) equivalent to my own notion, also that for proper names Peacocke's equivalence holds for temporal contexts. (This is roughly equivalent to the 'temporal rigidity' of names.) I also agree that it is very plausible to extend the principle to all contexts. But, as Peacocke recognizes, this appears to imply a substitutivity principle for codesignative proper names in belief contexts, which is widely assumed to be false. Peacocke proposes to use Davidson's theory of intensional contexts to block this conclusion (the material in the 'that' clause is a separate sentence). I myself cannot accept Davidson's theory; but even if it were true, Peacocke in effect acknowledges that it does not really dispose of the difficulty (p. 127, first paragraph). (Incidentally, if Davidson's theory does block any inference to the transparency of belief contexts with respect to names, why does Peacocke assume without argument that it does not do so for modal contexts, which have a similar grammatical structure?) The problems are thus those of the present paper; until they are resolved I prefer at present to keep to my earlier more cautious formulation.

Incidentally, Peacocke hints a recognition that the received platitude — that codesignative names are not interchangeable in belief contexts — may not be so clear as is generally supposed.

<sup>8</sup> The example comes from Quine, *Word and Object*, M.I.T. Press, 1960, p. 145. Quine's

conclusion that 'believes that' construed *de dicto* is opaque has widely been taken for granted. In the formulation in the text I have used the colon to emphasize that I am speaking of belief *de dicto*. Since, as I have said, belief *de dicto* will be our *only* concern in this paper, in the future the colon will usually be suppressed, and all 'believes that' contexts should be read *de dicto* unless the contrary is indicated explicitly.

<sup>9</sup> In many writings Peter Geach has advocated a view that is nonMillian (he would say 'nonLockean') in that to each name a sortal predicate is attached by definition ('Geach,' for example, by *definition* names a man). On the other hand, the theory is not completely Fregean either, since Geach denies that any definite description that would identify the referent of the name among things of the same sort is analytically tied to the name. (See, for example, his *Reference and Generality*, Cornell, 1962, pp. 43–45.) As far as the present issues are concerned, Geach's view can fairly be assimilated to *Mill's* rather than Frege's. For such ordinary names as 'Cicero' and 'Tully' will have both the same reference and the same (Geachian) sense (namely, that they are names of a man). It would thus seem that they ought to be interchangeable everywhere. (In *Reference and Generality*, Geach appears not to accept this conclusion, but the *prima facie* argument for the conclusion will be the same as on a purely Millian view.)

<sup>10</sup> In an unpublished paper, Diana Ackerman urges the problem of substitutivity failures against the Millian view and, hence, against my own views. I believe that others may have done so as well. (I have the impression that the paper has undergone considerable revision, and I have not seen recent versions.) I agree that this problem is a considerable difficulty for the Millian view, and for the Millian *spirit* of my own views in "Naming and Necessity." (See the discussion of this in the text of the present paper.) On the other hand I would emphasize that there need be no *contradiction* in maintaining that names are *modally* rigid, and satisfy a substitutivity principle for modal contexts, while denying the substitutivity principle for belief contexts. The entire apparatus elaborated in "Naming and Necessity" of the distinction between epistemic and metaphysical necessity, and of giving a meaning and fixing a reference, was meant to show, among other things, that a Millian substitutivity doctrine for modal contexts can be maintained even if such a doctrine for epistemic contexts is rejected. "Naming and Necessity" never asserted a substitutivity principle for epistemic contexts.

It is even consistent to suppose that differing modes of (rigidly) fixing the reference is responsible for the substitutivity failures, thus adopting a position intermediate between Frege and Mill, on the lines indicated in the text of the present paper. "Naming and Necessity" may even perhaps be taken as suggesting, for some contexts where a conventional description rigidly fixes the reference ('Hesperus–Phosphorus'), that the mode of reference fixing is relevant to epistemic questions. I knew when I wrote "Naming and Necessity" that substitutivity issues in epistemic contexts were really very delicate, due to the problems of the present paper, but I thought it best not to muddy the waters further. (See notes 43–44.)

After this paper was completed, I saw Alvin Plantinga's paper "The Boethian Compromise," *The American Philosophical Quarterly* 15 (April, 1978): 129–138. Plantinga adopts a view intermediate between Mill and Frege, and cites substitutivity failures as a principal argument for his position. He also refers to a forthcoming paper by Ackerman. I have not seen this paper, but it probably is a descendant of the paper referred to above.

<sup>11</sup> Here I use 'connotation' so as to imply that the associated properties have an *a priori* tie to the name, at least as rigid reference fixers, and therefore must be true of the referent (if it exists). There is another sense of 'connotation,' as in 'The Holy Roman Empire,' where the connotation need not be assumed or even believed to be true of the referent. In some sense akin to this, classicists and others with some classical learning may attach certain distinct 'connotations' to 'Cicero' and 'Tully.' Similarly, 'The Netherlands' may suggest low altitude to a thoughtful ear. Such 'connotations' can hardly be thought of as community-wide; many use the names unaware of such suggestions. Even a speaker aware of the suggestion of the name may not regard the suggested properties as true of the object; cf. 'The Holy Roman Empire.' A 'connotation' of this type neither gives a meaning nor fixes a reference.

<sup>12</sup> Some might attempt to find a difference in 'sense' between 'Cicero' and 'Tully' on the grounds that "Cicero is called 'Cicero'" is trivial, but "Tully is called 'Cicero'" may not be. Kneale, and in one place (probably at least implicitly) Church, have argued in this vein. (For Kneale, see "Naming and Necessity," p. 283.) So, it may be argued, being called 'Cicero,' is part of the sense of the name 'Cicero,' but not part of that of 'Tully.'

I have discussed some issues related to this in "Naming and Necessity," pp. 283–286. (See also the discussions of circularity conditions elsewhere in "Naming and Necessity.") Much more could be said about and against this kind of argument; perhaps I will sometime do so elsewhere. Let me mention very briefly the following parallel situation (which may be best understood by reference to the discussion in "Naming and Necessity"). Anyone who understands the meaning of 'is called' and of quotation in English (and that 'alienists' is meaningful and grammatically appropriate), knows that "alienists are called 'alienists'" expresses a truth in English, even if he has no idea what 'alienists' means. He need *not* know that "psychiatrists are called 'alienists'" expresses a truth. None of this goes to show that 'alienists' and 'psychiatrists' are not synonymous, or that 'alienists' has *being called 'alienists'* as part of its meaning when 'psychiatrists' does not. Similarly for 'Cicero' and 'Tully.' There is no more reason to suppose that being so-called is part of the meaning of a name than of any other word.

<sup>13</sup> A view follows Frege and Russell on this issue even if it allows each speaker to associate a cluster of descriptions with each name, provided that it holds that the cluster varies from speaker to speaker and that variations in the cluster are variations in idiolect. Searle's view thus is Frege-Russellian when he writes in the concluding paragraph of "Proper Names" (*Mind* 67 (1958): 166–173), " 'Tully = Cicero' would, I suggest, be analytic for most people; the same descriptive presuppositions are associated with each name. But of course if the descriptive presuppositions were different it might be used to make a synthetic statement."

<sup>14</sup> Though here I use the jargon of propositions, the point is fairly insensitive to differences in theoretical standpoints. For example, on Davidson's analysis, I would be asserting (roughly) that many are unaware-of-the-content-of the following *utterance* of mine: Cicero is Tully. This would be subject to the same problem.

<sup>15</sup> Benson Mates, "Synonymity," *University of California Publications in Philosophy* 25 (1950): 201–226; reprinted in: *Semantics and the Philosophy of Language*, L. Linsky (ed.), University of Illinois Press, 1952. (There was a good deal of subsequent discussion. In Mates's original paper the point is made almost parenthetically.) Actually, I think that

Mates's problem has relatively little force against the argument we are considering for the Fregean position. Mates's puzzle in no way militates against some such principle as: If one word is synonymous with another, then a sufficiently reflective speaker subject to no linguistic inadequacies or conceptual confusions who sincerely assents to a simple sentence containing the one will also (sincerely) assent to the corresponding sentence with the other in its place.

It is surely a crucial part of the present 'Fregean' argument that codesignative names may have distinct 'senses,' that a speaker may assent to a simple sentence containing one and deny the corresponding sentence containing the other, even though he is *guilty of no conceptual or linguistic confusion, and of no lapse in logical consistency*. In the case of two straightforward synonyms, this is not so.

I myself think that Mates's argument is of considerable interest, but that the issues are confusing and delicate and that, if the argument works, it probably leads to a paradox or puzzle rather than to a definite conclusion. (See also notes 23, 28, and 46.)

<sup>16</sup> "Naming and Necessity," pp. 291 (bottom)–293.

<sup>17</sup> Recall also note 12.

<sup>18</sup> Some philosophers stress that names are not *words* of a language, or that names are not *translated* from one language to another. (The phrase 'common currency of our common language' was meant to be neutral with respect to any such alleged issue.) Someone may use 'Mao Tse-Tung,' for example, in English, though he knows not one word of Chinese. It seems hard to deny, however, that "*Deutschland*," "*Allemagne*," and "Germany," are the German, French, and English names of a single country, and that one translates a French sentence using "*Londres*" by an English sentence using "London." Learning these facts *is* part of learning German, French, and English.

It would appear that *some* names, especially names of countries, other famous localities, and some famous people *are* thought of as part of a language (whether they are called 'words' or not is of little importance). Many other names are not thought of as part of a language, especially if the referent is not famous (so the notation used is confined to a limited circle), or if the same name is used by speakers of all languages. As far as I can see, it makes little or no *semantic* difference whether a particular name is thought of as part of a language or not. Mathematical notation such as '<' is also ordinarily not thought of as part of English, or any other language, though it is used in combination with English words in sentences of mathematical treatises written in English. (A French mathematician can use the notation though he knows not one word of English.) 'Is less than,' on the other hand, *is* English. Does this difference have any semantic significance?

I will speak in most of the text as if the names I deal with are part of English, French, etc. But it matters little for what I say whether they are thought of as parts of the language or as adjuncts to it. And one need not say that a name such as '*Londres*' is 'translated' (if such a terminology suggested that names have 'senses,' I too would find it objectionable), as long as one acknowledges that *sentences* containing it are properly translated into English using 'London.'

<sup>19</sup> By saying that names are transparent in a context, I mean that codesignative names are interchangeable there. This is a deviation for brevity from the usual terminology, according to which the *context* is transparent. (I use the usual terminology in the paper also.)

<sup>20</sup> But we must use the term 'sense' here in the sense of 'that which fixes the reference,' not 'that which gives the meaning,' otherwise we shall run afoul of the rigidity of proper names. If the source of a chain for a certain name is in fact a given object, we use the name to designate that object even when speaking of counterfactual situations in which some *other* object originated the chain.

<sup>21</sup> The point is that, according to the doctrine of "Naming and Necessity," when proper names are transmitted from link to link, even though the beliefs about the referent associated with the name change radically, the change is not to be considered a linguistic change, in the way it *was* a linguistic change when 'villain' changed its meaning from 'rustic' to 'wicked man.' As long as the reference of a name remains the same, the associated beliefs about the object may undergo a large number of changes without these changes constituting a change in the language.

If Geach is right, an appropriate sortal must be passed on also. But see footnote 58 of "Naming and Necessity."

<sup>22</sup> Similar appropriate restrictions are assumed below for the strengthened disquotational principle and for the principle of translation. Ambiguities need not be excluded if it is tacitly assumed that the sentence is to be understood in one way in all its occurrences. (For the principle of translation it is similarly assumed that the translator matches the *intended* interpretation of the sentence.) I do not work out the restrictions on indexicals in detail, since the intent is clear.

Clearly, the disquotational principle applies only to *de dicto*, not *de re*, attributions of belief. If someone sincerely assents to the near triviality "The tallest foreign spy is a spy," it follows that he believes that: the tallest foreign spy is a spy. It is well known that it does *not* follow that he believes, *of* the tallest foreign spy, that he is a spy. In the latter case, but not in the former, it would be his patriotic duty to make contact with the authorities.

<sup>23</sup> What if a speaker assents to a sentence, but fails to assent to a synonymous assertion? Say, he assents to "Jones is a doctor," but not to "Jones is a physician." Such a speaker either does not understand one of the sentences normally, or he should be able to correct himself "on reflection." As long as he confusedly assents to 'Jones is a doctor' but not to 'Jones is a physician,' we *cannot* straightforwardly apply disquotational principles to conclude that he does or does not believe that Jones is a doctor, because his assent is not "reflective."

Similarly, if someone asserts, "Jones is a doctor but not a physician," he should be able to recognize his inconsistency without further information. We have formulated the disquotational principles so they need not lead us to attribute belief as long as we have grounds to suspect conceptual or linguistic confusion, as in the cases just mentioned.

Note that if someone says, "Cicero was bald but Tully was not," there need be *no* grounds to suppose that he is under *any* linguistic or conceptual confusion.

<sup>24</sup> This should not be confused with the question whether the speaker simultaneously believes *of* a given object, both that it has a certain property and that it does not have it. Our discussion concerns *de dicto* (notional) belief, not *de re* belief.

I have been shown a passage in Aristotle that appears to suggest that *no one* can really believe both of two explicit contradictories. If we wish to use the *simple* disquotational principle as a test for disbelief, it suffices that this be true of *some* individuals, after reflection, who are simultaneously aware of both beliefs, and have sufficient logical acumen and

respect for logic. Such individuals, if they have contradictory beliefs, will be shaken in one or both beliefs after they note the contradiction. For such individuals, sincere reflective assent to the negation of a sentence implies disbelief in the proposition it expresses, so the test in the text applies.

<sup>25</sup> For example, in translating a historical report into another language, such as, "Patrick Henry said, 'Give me liberty or give me death!'" the translator may well translate the quoted material attributed to Henry. He translates a presumed truth into a falsehood, since Henry spoke English; but probably his reader is aware of this and is more interested in the content of Henry's utterance than in its exact words. Especially in translating fiction, where truth is irrelevant, this procedure is appropriate. But some objectors to Church's 'translation argument' have allowed themselves to be misled by the practice.

<sup>26</sup> To state the argument precisely, we need in addition a form of the Tarskian disquotation principle for truth: For each (French or English) replacement for '*p*,' infer "'*p*' is true" from "*p*," and conversely. (Note that "'*p*' is true" becomes an English sentence even if '*p*' is replaced by a French sentence.) In the text we leave the application of the Tarskian disquotational principle tacit.

<sup>27</sup> I gather that Burali-Forti originally thought he had 'proved' that the ordinals are not linearly ordered, reasoning in a manner similar to our topologist. Someone who heard the present paper delivered told me that König made a similar error.

<sup>28</sup> It is not possible, in this case, as it is in the case of the man who assents to "Jones is a doctor" but not to "Jones is a physician," to refuse to apply the disquotational principle on the grounds that the subject must lack proper command of the language or be subject to some linguistic or conceptual confusion. As long as Pierre is unaware that 'London' and '*Londres*' are codesignative, he need not lack appropriate linguistic knowledge, nor need he be subject to any linguistic or conceptual confusion, when he affirms '*Londres est jolie*' but denies 'London is pretty.'

<sup>29</sup> The 'elimination' would be most plausible if we believed, according to a Russellian epistemology, that all my language, when written in unabbreviated notation, refers to constituents with which I am 'acquainted' in Russell's sense. Then no one speaks a language intelligible to anyone else; indeed, no one speaks the same language twice. Few today will accept this.

A basic consideration should be stressed here. Moderate Fregeans attempt to combine a roughly Fregean view with the view that names are part of our common language, and that our conventional practices of interlinguistic translation and interpretation are correct. The problems of the present paper indicate that it is very difficult to obtain a requisite socialized notion of sense that will enable such a program to succeed. Extreme Fregeans (such as Frege and Russell) believe that in general names are peculiar to idiolects. They therefore would accept no general rule translating '*Londres*' as 'London,' nor even translating one person's use of 'London' into another's. However, if they follow Frege in regarding senses as 'objective,' they must believe that in principle it makes sense to speak of two people using two names in their respective idiolects with the same sense, and that there must be (necessary and) sufficient conditions for this to be the case. If these conditions for sameness of sense are satisfied, translation of one name into the other is legitimate, otherwise not. The present considerations (and the extension of these below to natural kind and related terms), however, indicate that the notion of sameness of sense, if it is to be explicated in

terms of sameness of identifying properties and if these properties are themselves expressed in the languages of the two respective idiolects, presents interpretation problems of the same type presented by the names themselves. Unless the Fregean can give a method for identifying sameness of sense that is free of such problems, he *has no sufficient conditions for sameness of sense, nor for translation to be legitimate*. He would therefore be forced to maintain, contrary to Frege's intent, that not only in practice do few people use proper names with the same sense but that *it is in principle meaningless to compare senses*. A view that the identifying properties used to define senses should always be expressible in a Russellian language of 'logically proper names' would be one solution to this difficulty but involves a doubtful philosophy of language and epistemology.

<sup>30</sup> If any reader finds the term 'translation' objectionable with respect to names, let him be reminded that all I mean is that French sentences containing '*Londres*' are uniformly translated into English with 'London.'

<sup>31</sup> The paradox would be blocked if we required that they define the names by the same properties expressed in the same words. There is nothing in the motivation of the classical description theories that would justify this extra clause. In the present case of French and English, such a restriction would amount to a decree that neither '*Londres*,' nor any other conceivable French name, could be translated as 'London.' I deal with this view immediately below.

<sup>32</sup> Word salads of two languages (like ungrammatical 'semisentences' of a single language) need not be unintelligible, though they are makeshifts with no fixed syntax. "If God did not exist, Voltaire said, *il faudrait l'inventer*." The meaning is clear.

<sup>33</sup> Had we said, "Pierre believes that the country he calls '*Angleterre*' is a monarchy," the sentence would be English, since the French word would be mentioned but not used. But for this very reason we would not have captured the sense of the French original.

<sup>34</sup> Under the influence of Quine's *Word and Object*, some may argue that such conclusions are not inevitable: perhaps he will translate '*médecin*' as 'doctor stage,' or 'undetached part of a doctor'! If a Quinean skeptic makes an empirical prediction that such reactions from bilinguals as a matter of fact can occur, I doubt that he will be proved correct. (I don't know what Quine would think. But see *Word and Object*, p. 74, first paragraph.) On the other hand, if the translation of '*médecin*' as 'doctor' rather than 'doctor part' in this situation *is*, empirically speaking, inevitable, then even the advocate of Quine's thesis will have to admit that there is something special about one particular translation. The issue is not crucial to our present concerns, so I leave it with these sketchy remarks. But see also note 36.

<sup>35</sup> Putnam gives the example of elms and beeches in "The Meaning of 'Meaning'" (in: *Language, Mind, and Knowledge*, Minnesota Studies in the Philosophy of Science 7; also reprinted in Putnam's *Collected Papers*). See also Putnam's discussion of other examples on pp. 139–143; also my own remarks on 'fool's gold,' tigers, etc., in "Naming and Necessity," pp. 316–323.

<sup>36</sup> It is unclear to me how far this can go. Suppose Pierre hears English spoken only in England, French in France, and learns both by direct method. (Suppose also that no one else in each country speaks the language of the other.) Must he be sure that 'hot' and '*chaud*' are coextensive? In practice he certainly would. But suppose somehow his experience is consistent with the following bizarre — and of course, false! — hypothesis: England and



France differ atmospherically so that human bodies are affected very differently by their interaction with the surrounding atmosphere. (This would be more plausible if France were on another planet.) In particular, within reasonable limits, things that feel cold in one of the countries feel hot in the other, and *vice versa*. Things don't change their *temperature* when moved from England to France, they just *feel* different because of their effects on human physiology. Then '*chaud*,' in French, would be true of the things that are called 'cold' in English! (Of course the present discussion is, for space, terribly compressed. See also the discussion of 'heat' in "Naming and Necessity." We are simply creating, for the physical property 'heat,' a situation analogous to the situation for natural kinds in the text.)

If Pierre's experiences were arranged somehow so as to be consistent with the bizarre hypothesis, and he somehow came to believe it, he might simultaneously assent to '*C'est chaud*' and 'This is cold' without contradiction, even though he speaks French and English normally in each country separately.

This case needs much more development to see if it can be set up in detail, but I cannot consider it further here. Was I right in assuming in the text that the difficulty could not arise for '*médecin*' and 'doctor'?

<sup>37</sup> One might argue that Peter and we do speak different dialects, since in Peter's idiolect 'Paderewski' is used ambiguously as a name for a musician and a statesman (even though these are in fact the same), while in our language it is used unambiguously for a musician-statesman. The problem then would be whether Peter's dialect can be translated homophonically into our own. Before he hears of 'Paderewski-the-statesman,' it would appear that the answer is affirmative for his (then unambiguous) use of 'Paderewski,' since he did not differ from anyone who happens to have heard of Paderewski's musical achievements but not of his statesmanship. Similarly for his later use of 'Paderewski,' if we ignore his earlier use. The problem is like Pierre's, and is essentially the same whether we describe it in terms of whether Peter satisfies the condition for the disquotational principle to be applicable, or whether homophonic translation of his dialect into our own is legitimate.

<sup>38</sup> D. Davidson, "On Saying That," in: *Words and Objections*, D. Davidson and J. Hintikka (eds.), Dordrecht, Reidel, 1969, p. 166.

<sup>39</sup> In *Word and Object*, p. 221, Quine advocates a second level of canonical notation, "to dissolve verbal perplexities or facilitate logical deductions," admitting the propositional attitudes, even though he thinks them "baseless" idioms that should be excluded from a notation "limning the true and ultimate structure of reality."

<sup>40</sup> In one respect the considerations mentioned above on natural kinds show that Quine's translation apparatus is insufficiently skeptical. Quine is sure that the native's *sentence* "Gavagai!" should be translated "Lo, a rabbit!," provided that its affirmative and negative stimulus meanings for the native match those of the English sentence for the Englishman; skepticism sets in only when the linguist proposes to translate the *general term* 'gavagai' as 'rabbit' rather than 'rabbit stage,' 'rabbit part,' and the like. But there is another possibility that is independent of (and less bizarre than) such skeptical alternatives. In the geographical area inhabited by the natives, there may be a species indistinguishable to the nonzoologist from rabbits but forming a distinct species. Then the 'stimulus meanings,' in Quine's sense, of 'Lo, a rabbit!' and 'Gavagai!' may well be identical (to nonzoologists), especially if the ocular irradiations in question do not include a specification of the geographical locality.

(‘Gavagai’ produce the same ocular irradiation patterns as rabbits.) Yet ‘Gavagai!’ and ‘Lo, a rabbit!’ are hardly synonymous; on typical occasions they will have opposite truth values.

I believe that the considerations about names, let alone natural kinds, emphasized in “Naming and Necessity” go against any simple attempt to base interpretation solely on maximizing agreement with the affirmations attributed to the native, matching of stimulus meanings, etc. The ‘Principle of Charity’ on which such methodologies are based was first enunciated by Neil Wilson in the special case of proper names as a formulation of the cluster-of-descriptions theory. The argument of “Naming and Necessity” is thus directed against the simple ‘Principle of Charity’ for that case.

<sup>41</sup> Geach introduced the term ‘Shakespearean’ after the line, “a rose / By any other name, would smell as sweet.”

Quine seems to define ‘referentially transparent’ contexts so as to imply that coreferential names and definite descriptions must be interchangeable *salva veritate*. Geach stresses that a context may be ‘Shakespearean’ but not ‘referentially transparent’ in this sense.

<sup>42</sup> Generally such cases may be slightly less watertight than the ‘London’–‘Londres’ case. ‘Londres’ just is the French version of ‘London,’ while one cannot quite say that the same relation holds between ‘Ashkenaz’ and ‘Germaniah.’ Nevertheless:

(a) Our standard practice in such cases is to translate both names of the first language into the single name of the second.

(b) Often no nuances of ‘meaning’ are discernible differentiating such names as ‘Ashkenaz’ and ‘Germaniah,’ such that we would not say either that Hebrew would have been impoverished had it lacked one of them (or that English is impoverished because it has only one name for Germany), any more than a language is impoverished if it has only one word corresponding to ‘doctor’ and ‘physician.’ Given this, it seems hard to condemn our practice of translating both names as ‘Germany’ as ‘loose’; in fact, it would seem that Hebrew just has two names for the same country where English gets by with one.

(c) Any inclinations to avoid problems by declaring, say, the translation of ‘Ashkenaz’ as ‘Germany’ to be loose should be considerably tempered by the discussion of analogous problems in the text.

<sup>43</sup> In spite of this official view, perhaps I will be more assertive elsewhere.

In the case of ‘Hesperus’ and ‘Phosphorus’ (in contrast to ‘Cicero’ and ‘Tully’), where there is a case for the existence of conventional community-wide ‘senses’ differentiating the two — at least, two distinct modes of ‘fixing the reference of two rigid designators’ — it is more plausible to suppose that the two names are definitely not interchangeable in belief contexts. According to such a supposition, a belief that Hesperus is a planet is a belief that a certain heavenly body, rigidly picked out as seen in the evening in the appropriate season, is a planet; and similarly for Phosphorus. One may argue that translation problems like Pierre’s will be blocked in this case, that ‘Vesper’ must be translated as ‘Hesperus,’ not as ‘Phosphorus.’ As against this, however, two things:

(a) We should remember that sameness of properties used to fix the reference does *not* appear to guarantee in general that paradoxes will not arise. So one may be reluctant to adopt a solution in terms of reference-fixing properties for this case if it does not get to the heart of the general problem.

(b) The main issue seems to me here to be — how essential is a particular mode of fixing the reference to a correct learning of the name? If a parent, aware of the familiar identity, takes a child into the fields in the morning and says (pointing to the morning star) “That is called ‘Hesperus,’ ” has the parent mistaught the language? (A parent who says, “Creatures with kidneys are called ‘cordates,’ definitely has mistaught the language, even though the statement is extensionally correct.) To the extent that it is *not* crucial for correct language learning that a particular mode of fixing the reference be used, to that extent there is no ‘mode of presentation’ differentiating the ‘content’ of a belief about ‘Hesperus’ from one about ‘Phosphorus.’ I am doubtful that the original method of fixing the reference *must* be preserved in transmission of the name.

If the mode of reference fixing *is* crucial, it can be maintained that otherwise identical beliefs expressed with ‘Hesperus’ and with ‘Phosphorus’ have definite differences of ‘content,’ at least in an epistemic sense. The conventional ruling against substitutivity could thus be maintained without qualms for some cases, though not as obviously for others, such as ‘Cicero’ and ‘Tully.’ But it is unclear to me whether even ‘Hesperus’ and ‘Phosphorus’ do have such conventional ‘modes of presentation.’ I need not take a definite stand, and the verdict may be different for different particular pairs of names. For a brief related discussion, see “Naming and Necessity,” p. 331, first paragraph.

<sup>44</sup> However, some earlier formulations expressed disquotationally such as “It was once unknown that Hesperus is Phosphorus” are questionable in the light of the present paper (but see the previous note for this case). I was aware of this question by the time “Naming and Necessity” was written, but I did not wish to muddy the waters further than necessary at that time. I regarded the distinction between epistemic and metaphysical necessity as valid in any case and adequate for the distinctions I wished to make. The considerations in this paper are relevant to the earlier discussion of the ‘contingent *a priori*’ as well; perhaps I will discuss this elsewhere.

<sup>45</sup> According to Russell, definite descriptions are not genuine singular terms. He thus would have regarded any concept of ‘referential opacity’ that includes definite descriptions as profoundly misleading. He also maintained a substitutivity principle for ‘logically proper names’ in belief and other attitudinal contexts, so that for him belief contexts were as ‘transparent,’ in any philosophically decent sense, as truth-functional contexts.

Independently of Russell’s views, there is much to be said for the opinion that the question whether a context is ‘Shakespearean’ is more important philosophically — even for many purposes for which Quine invokes his own concept — than whether it is ‘referentially opaque.’

<sup>46</sup> I will make some brief remarks about the relation of Benson Mates’s problem (see note 15) to the present one. Mates argued that such a sentence as (\*) ‘Some doubt that all who believe that doctors are happy believe that physicians are happy,’ may be true, even though ‘doctors’ and ‘physicians’ are synonymous, and even though it would have been false had ‘physicians’ been replaced in it by a second occurrence of ‘doctors.’ Church countered that (\*) could not be true, since its translation into a language with only one word for doctors (which would translate both ‘doctors’ and ‘physicians’) would be false. If *both* Mates’s and Church’s intuitions were correct, we might get a paradox analogous to Pierre’s.

Applying the principles of translation and disquotation to Mates’s puzzle, however,

involves many more complications than our present problem. First, if someone assents to 'Doctors are happy,' but refuses assent to 'Physicians are happy,' *prima facie* disquotation does not apply to him since he is under a linguistic or conceptual confusion. (See note 23.) So there are as yet no grounds, merely because this happened, to doubt that all who believe that doctors are happy believe that physicians are happy.

Now suppose someone assents to 'Not all who believe that doctors are happy believe that physicians are happy.' What is the source of his assent? If it is failure to realize that 'doctors' and 'physicians' are synonymous (this was the situation Mates originally envisaged), then he is under a linguistic or conceptual confusion, so disquotation does not clearly apply. Hence we have no reason to conclude from this case that (\*) is true. Alternatively, he may realize that 'doctors' and 'physicians' are synonymous; but he applies disquotation to a man who assents to 'Doctors are happy' but not to 'Physicians are happy,' ignoring the caution of the previous paragraph. Here he is not under a simple linguistic confusion (such as failure to realize that 'doctors' and 'physicians' are synonymous), but he appears to be under a deep conceptual confusion (misapplication of the disquotational principle). Perhaps, it may be argued, he misunderstands the 'logic of belief.' Does his conceptual confusion mean that we cannot straightforwardly apply disquotation to his utterance, and that therefore we cannot conclude from his behavior that (\*) is true? I think that, although the issues are delicate, and I am not at present completely sure what answers to give, there is a case for an affirmative answer. (Compare the more extreme case of someone who is so confused that he thinks that someone's *dissent* from 'Doctors are happy' implies that he believes that doctors are happy. If someone's utterance, 'Many believe that doctors are happy,' is based on such a misapplication of disquotation, surely we in turn should not apply disquotation to it. The utterer, at least in this context, does not really know what 'belief' means.)

I do *not* believe the discussion above ends the matter. Perhaps I can discuss Mates's problem at greater length elsewhere. Mates's problem is perplexing, and its relation to the present puzzle is interesting. But it should be clear from the preceding that Mates's argument involves issues even more delicate than those that arise with respect to Pierre. First, Mates's problem involves delicate issues regarding iteration of belief contexts, whereas the puzzle about Pierre involves the application of disquotation only to affirmations of (or assents to) *simple* sentences. More important, Mates's problem would not arise in a world where no one ever was under a linguistic or a conceptual confusion, no one ever thought anyone else was under such a confusion, no one ever thought anyone ever thought anyone was under such a confusion, and so on. It is important, both for the puzzle about Pierre and for the Fregean argument that 'Cicero' and 'Tully' differ in 'sense,' that they would still arise in such a world. They are entirely free of the delicate problem of applying disquotation to utterances directly or indirectly based on the existence of linguistic confusion. See notes 15 and 28, and the discussion in the text of Pierre's logical consistency.

Another problem discussed in the literature to which the present considerations may be relevant is that of 'self-consciousness,' or the peculiarity of 'I.' Discussions of this problem have emphasized that 'I,' even when Mary Smith uses it, is not interchangeable with 'Mary Smith,' nor with any other conventional singular term designating Mary Smith. If she is 'not aware that she is Mary Smith,' she may assent to a sentence with 'I,' but dissent from the corresponding sentence with 'Mary Smith.' It is quite possible that any attempt

to clear up the logic of all this will involve itself in the problem of the present paper. (For this purpose, the present discussion might be extended to demonstratives and indexicals.)

The writing of this paper had partial support from a grant from the National Science Foundation, a John Simon Guggenheim Foundation Fellowship, a Visiting Fellowship at All Souls College, Oxford, and a sabbatical leave from Princeton University. Various people at the Jerusalem Encounter and elsewhere, who will not be enumerated, influenced the paper through discussion.