WHAT'S IN A BELIEF?

J.I. BIRO

I

An alleged puzzle about belief introduced recently by Kripke has been taken to represent a serious problem for our account of belief attribution, indeed, a challenge to the coherence of our concept of belief.([1]) The purpose of this note is to suggest that such worries are exaggerated. I shall sketch an explanation of the situation Kripke describes which shows the "puzzle" to be no puzzle at all. All that is involved is a confusion resulting from overlooking some important, but not mysterious, facts about the influence of the subject's and the reporter's respective semantic beliefs on the content of a belief attribution.

First, a brief reminder of how the putative puzzle goes. We are asked to imagine someone, rational, sincere and linguistically competent, assenting to sentences, in different languages, in which contradictory predicates are applied to different subject terms whose denotation, unbeknownst to him, is the same. Such a person, it seems, must be charged with holding contradictory beliefs, if his assent to these sentences is to be taken as a basis, as it usually is, for fixing on what his beliefs are. The reason for this is that two obvious and seemingly innocent principles on which we normally proceed, dubbed by Kripke "the principle of disquotation," (PD) and "the principle of translation," (PT) appear to force these problematic belief attributions on us. The first of these, says Kripke, "can be stated as follows, where 'p' is to be replaced, inside and outside all quotation marks, by

([1]) S. KRIPKE, "A Puzzle About Belief," in *Meaning and Use*, A. Margolit, ed. (Dordrecht: 1979). There has been much discussion of the puzzle already. Some interesting examples are :R.B. MARCUS, "A Proposed Solution to a Puzzle about Belief," *Midwest Studies in Philosophy*, VI (1981); D. LEWIS, "What Puzzling Pierre Does Not Believe," *Australian Journal of Philosophy*, 59 (1981); B. HARRISON, "Description and Identification," *Mind*, XCI (1982); R.B. MARCUS, "Rationality and Believing the Impossible," *Journal of Philosophy*, June 1983. The account sketched here is presented in greater detail in J. Biro, "Propositional attitues and the semantics of names," read to the 1982 meeting of the Society for Philosophy and Psychology.

any appropriate standard English sentence: 'If a normal English speaker, on reflection, sincerely assents to 'p', then he believes that 'p'.'' The second principles goes: ''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).''([2]) Thus, if the fact that someone assents to *''Londres est jolie''* warrants our inferring that the believes that London is pretty (to use Kripke's example), his equally sincere assent to ''London is not pretty'' must justify the attribution to him of the belief that London is pretty. And, as Kripke shows, it's easy to imagine a situation where we find someone assenting to both of these sentences. We need only think of *Pierre* as coming to assent to the former, as might any other monolingual and untravelled Frenchman, on the basis of hearsay and then moving to London, learning English by immersion, and coming to assent to the latter on the basis of what he sees. Assuming that he is rational, we cannot think of him as holding both beliefs; yet we don't know which attribution to withhold. What has gone wrong?

Kripke does not offer an answer; his concern is, rather, to persuade us of the genuineness and seriousness of the dilemma. To this end, he expends most of his effort on embellishing the story in ways that make the attribution of each of the seemingly contradictory beliefs plausible and natural. In fact, he suggests that the alleged difficulty is not confined to belief attribution on the basis of sentences involving foreign languages and proper names: analogous stories can be told involving someone's assent to sentences all in English, whether these contain co-denoting names or pairs of common nouns (e.g., 'gorse' and 'furze') with the same extension. (It is easy to see how cases involving definite descriptions can be generated.)

The general pattern of all these cases is that the subject of our belief attributions has an incomplete or erroneous understanding of the references of the expressions that occur in the sentences to which he assents. When this is the case, and when disquatation is applied to those sentences, the resulting belief attribution may be misleading to

---

([2]) Principles analogous to the principle of disquotation presumably apply to other languages, so a generalized form of the principle might be: ''For every L, there is a principle statable in L, according to which if a normal speaker of L, upon reflection, sincerely assents to 'p', then he believes that p (where 'p' is to be replaced, inside and outside all quotation marks, by any appropriate standard sentence of L).''

those whose own grasp of the semantics of those referring expressions is more complete and accurate.

In talking about an accurate and complete grasp of the semantics of a referring expression I do not mean to suggest that I have any very precise and detailed idea wherein such grasp consists. There are some thorny questions here: Does one have to know everything about the use of an expression in one's linguistic community to qualify? (If so, most of us don't have a grasp of more than a few names). Just what counts as a linguistic community to begin with? Fortunately, we do not require an answer to these and other mysteries in order to appreciate that the kind of story Kripke tells depends crucially on the belief-attributor's having a *more* complete semantical understanding than has the subject of his attributions, *whatever* such understanding consist in.(3) It is essential for the Pierre story that he should be assumed *not* to know something we do know, namely, that '*Londres*' and 'London' denote the same thing. Without making such an assumption, we cannot go on to imagine Pierre assenting in the way Kripke describes him as doing, and then we are not forced by our two principles to make the paradoxical belief-attributions Kripke thinks we are. Imagine Pierre sharing our knowledge that '*Londres*' and 'London' denote the same city: now you cannot imagine him as providing the pattern of responses which would lead, by way of the principle of disquotation, to the attribution of inconsistent beliefs. Or imagine him finding out that they do: now you have to imagine his pattern of assent as changing.

This gap between the subject's and the interpreter's semantic knowledge (4) is obviously what gives rise to the appearance of paradox in cases like Pierre's. The question is, why think that we have anything more than just an *appearance* of paradox? Why think there is a real puzzle?

Kripke thinks so, and his reason appears to be the independent plausibility of the two principles of interpretation just mentioned. I

(3) Some of these mysteries are discussed in Biro, op. cit. For a discussion of some puzzles resulting from an incomplete or erroneous understanding of terms other than names, see T. BURGE, "Individualism and the mental", *Midwest Studies in Philosophy* (1979). On these matters, see also fn. 12 below.

(4) I follow Kripke here in talking about *knowledge* of denotations. But see fn. 10 below.

want to argue that that, in spite of that plausibility, we have reason to be cautious in our use of one of them; that our use of the principle of disquotation should be subject to a constraint arising from considering a speaker's semantic beliefs. I shall claim that in our everyday use of the principle we do in fact respect that constraint, and that therefore in cases like Pierre's we would not make the paradoxical-seeming belief attributions that worry Kripke.(⁵)

II

Let me begin by informally defining a notion of the *semantic content of a term* as including its actual reference or extension, and, perhaps, the properties standardly associated with its referent. Thus, denoting London is certainly part (perhaps the whole) of the semantic content of both *'Londres'* and 'London'; London's being a great, historic and cosmopolitan capital is *perhaps* another part. With 'gorse' and 'furze', having as their extension a certain kind of shrub is certainly part (perhaps the whole) of their semantic content; the possession by that kind of the properties of being tough, wiry and prickly is perhaps another.(⁶) Using this notion of the semantic content of a term, we can define two further notions. The *semantic content of a sentence* will be that part of a sentence's meaning which depends on, and varies according to, the semantic content of the terms occurring in it. (There may be other aspects of a sentence's meaning which depend on other factors). A person's *semantic beliefs* will be his beliefs about what a term's or a sentence's semantic content is.

(⁵) Unlike some others (e.g., Ron CLARK in his unpublished "On Kripke's Puzzle") I shall not challenge the principle of translation. One *could* tinker with that, trading it off against the restriction on disquotation I propose. But that restriction is *natural* in a way that qualifications of the principle of translation I think are not.

(⁶) Perhaps instead of, or in addition to, such stereotypical properties, underlying essential properties should be included. If so, this opens up the way for thinking of a whole linguistic community as having erroneous semantic beliefs about a term, if it has mistaken views about what those underlying properties are. This is a tricky matter, but it does not really bear on the present issue, which is the principles required for the interpretation of *individuals*, whether in the same community or not.

Let us think of *the content of someone's belief* as including everything the believer has in mind in virtue of having it. Whatever precisely *having something in mind* amounts to, what a person explicitly entertains, or would more or less immediately infer from what he explicitly entertains, seems a good candidate for what *we* have in mind when we say that he has a belief with a certain content.([7])

The principle of disquotation embodies the assumption that the semantic content of the terms in a sentence assented to goes over without remainder or alteration into the content of the belief attributable to the assenter. One has only to put it this way to realize that this cannot in general be the case. Indeed, Kripke's Pierre story is more properly seen not as a deep puzzle about belief but as an illustration of the rashness of such an assumption. PD would license going from Pierre's assent to

(1) *'Londres est jolie'*

to (1') 'Pierre believes that London is pretty' (given PT)
and from his assent to

(2) 'London is not pretty'

to (2') 'Pierre believes that London is not pretty'
We regard (1) and (2) as inconsistent just in case we take the names in them to be co-referential. The embedding of (the translation of) (1) in (1') and of (2) in (2') suggests that they can be used to express the content of Pierre's beliefs. But this cannot be right if we take *'Londres'* and 'London' to be co-referential and Pierre does not. In making these inferences, we imply that he does. But in Kripke's story he is explicitly said not to be doing so. Therefore we should not include in our specification of his beliefs such part of the semantic content of the referring expressions occurring in (1) and (2) as he is, by hypothesis, unaware of. PD forces us to do so, therefore PD should not be used.

The fact that the two names, *'Londres'* and 'London' belong to different languages may make it appear that the problem is one of

---

([7]) Things get a little more complex with tacit beliefs (if there be such). But since the present issue concerns the interpretation of overt linguistic behaviour, we need worry only about explicit beliefs.

translation.([8]) But, as I noted, Kripke makes it clear that that is in fact an accidental feature of the story, albeit one that makes it especially dramatic and more easily grasped than single-language versions. Nevertheless, the latter are quite easy to construct once the general point is seen. The 'gorse'-'furze' case of course involves only one language; and Kripke offers another one of someone who assents to seemingly contradictory sentences containing the name 'Paderewski', believing that in each it denotes a different person. (Peter has learnt about a famous pianist called 'Paderewski' and – in a different context – about a great statesman called 'Paderewski'. Believing them to be different people, and sceptical of the artistic abilities of politicians, he assents to both 'Paderewski had musical talent' and 'Paderewski had no musical talent'. Should we now say, this time on the basis of the disquotational principal alone, that he believes that Paderewski both had and lacked musical talent?)([9]) In such cases, the crucial fact that our subject's semantic beliefs are deviant ought to show up in our statement of the contents of his beliefs – the trouble is that the principle of disquotation provides for no machinery for this. Worse, it suggests that it *need* not show up. In reports of belief based on that principle, such semantic beliefs of the subject are crowed out by the standard semantic content of the terms involved. Only if in attributing beliefs to our subject we rely on the actual semantic content of the sentences to which he assents will it seem that his beliefs are inconsistent; yet it is just of that actual semantic content that he is assumed in Kripke's story to be ignorant. Given that ignorance, we should not attribute to him beliefs whose content matches the semantic content of the sentences to which he assents; and of course in pratice we don't, if we are aware of his ignorance. It should not be surprising that a blind application of the principle of disquotation which ignores such factors creates an air of paradox.

([8]) The status of names in translation is a tricky matter; see, again, BIRO, op. cit. But note that if the principle of disquotation, together with the principle of translation, licenses going from (1) to (2), '*Londres*' must be translatable. The alternative is the unacceptable Frenglish (2''): 'Pierre believes that Londres is pretty'.

([9]) The Paderewski' case is discussed by K. BACH in "What's In a Name," *Australian Journal of Philosophy*, 59 (1981), whose treatment is criticised in BIRO, op. cit. See also M. DEVITT, "Thoughts and the attribution of thoughts," *Midwest Studies in Philosophy*, Vol. IX (forthcoming).

III

The feature common to all these cases is that in order to get the appearance of inconsistency among the subject's beliefs, we must trade on the very semantic facts that are by hypothesis not available to the subject. If they *were* available to him, we could give no rational account of his pattern of assent: in which case the assumption of nationality, on which the seriousness of the 'puzzle' depends, would have to be given up. (There is no temptation to draw radical conclusions about belief if an *irrational* person contradicts himself.) With a rational subject, on the other hand, we have only to think of him being put right in his semantic beliefs to 'see' him change his responses. Tell Pierre that *'Londres'* and 'London' name the same city and watch him take *something* back. The general principle about belief attribution which these cases serve to underline is, roughly, this: to the extent to which someone is not fully aware of the actual or conventional references of the names he uses, such references should not be used by third parties to land him with beliefs which only someone who *is* aware of them would rationally hold. The belief we must attribute to some one we are trying to interpret is *his* belief, not ours, and there is no guarantee that we do this if in specifying the content of the belief we use expressions whose semantic content may be obscured from him by idiosyncratic semantic beliefs he happens to hold concerning those expressions.

Indeed, the point here goes beyond just beliefs about the references of names and the extensions of kind terms. We need to be concerned about a person's understanding of all the terms that occur in the sentences he utters or accepts. This includes his understanding of the references of names and definite descriptions, of the extensions of common nouns, as well as of the rules governing the uses of non-substantival expressions, *including* syncategorematic terms. Thus, disquotation may run into trouble even on account of deviant beliefs about logical connectives. If I assent to some instance of $\ulcorner \phi \ v \ \psi \urcorner$ but think that the wedge means 'either one, but not both' (as some people are tempted to think at least some of the time about 'or'), your report that I believe that p v q will mislead an orthodox logician as to what I really believe.[10] Quite generally, there is no guarantee that our subject shares our understanding of the *sentences* we use in

stating what we think he believes *even if* he uses those very sentences (or, more cautiously, the same strings or sounds). It is this possible gap between what the subject understands by a sentence he uses and what his interpreter understands by it that is responsible for the trouble Kripke runs into; or, rather, it is the fact that he relies on a principle – that of disquotation – which takes no account of the possibility.

There will be no puzzle if we restrict the use of that principle in such a way as to ensure that a belief report based on it can carry no more information about the references and meaning of the terms occurring in the sentence used to express it than we have reason to believe is available to the subject, even if such information *is* available to the reporter. Such a restriction is easy enough to state in a general way, though difficult to apply in particular cases. We need merely require that in using the principle of disquotation in interpreting a speaker we pay due regard to his semantic beliefs. Should we have reason to think that these may be unorthodox, we will know that simple disquotation runs the risk of getting the subject's beliefs wrong. What we do then will depend on the circumstances of the case and especially on how much we know about the deviant semantic beliefs that thus block disquotation. Sometimes we will be able to proceed by semi-disquotation, as it were, by marking the presence and nature of the deviation by the inclusion of some description (of *ours*) accompanying the problematical term. Thus we may cautiously report someone assenting to 'Paderewski had no musical talent' as believing that Paderwski «the politician» had no musical talent. The inclusion of the descriptive phrase signals that we do not take the speaker's assent as expressing the belief that Paderewski the pianist had no musical talent. Unless we have reason to think that the speaker holds some idiosyncratic view of Paderewski *qua* pianist, it is a better explanation of his behaviour – as long as the rationality assumption is in force – that he does not believe that Paderwski is a pianist (and believes, not irrationally and not altogether implausibly, that politicians in general have no musical talent). Sometimes, in the absence of sufficient information about just

---

([10]) More serious troubles can arise when people subscribing to a "deviant logic" are subjected to unrestricted and incautious disquotation. (On "deviant logicians," see W.V. QUINE, *Philosophy of Logic*, Ch. 6).

what the subject thinks his terms mean, we will have to refrain from using disquotation altogether, in order not to saddle him with beliefs the possession of which does not follow from his assenting to the sentences he does. Speech and belief are in such cases severed.

IV

It may be objected that none of this helps with answering Kripke's straightforward question: what does Pierre believe? It's one thing to insist in general terms that a theory of interpretation should recognize the assumptions we make about, and the allowances we should make for, our subject's semantic beliefs; it's another to translate that into concrete judgements about a particular well-specified case. And, the complaint might run, the kind of caution about disquotation I have urged does not allow us to give a determinate answer to Kripke's question and so leaves us with nothing firm to say about Pierre. If this were so, there would be a puzzle, after all.

But it is not so. In Pierre's case, we know that he does not hold all the usual semantic beliefs about *'Londres'* and 'London', since, as Kripke tells us, he does not believe them to be co-referring.([11]) (If Kripke had not told us this, we would have very likely come to the same conclusion through observing Pierre's linguistic behaviour, as long as we found no evidence in that, or in his behaviour more generally, to prompt us to abandon the rationality assumption.([12]) So

([11]) Kripke talks about Pierre's not *knowing* that the two names are coreferring, as well as about his *ignorance*, and so, at times, have I. But, strictly speaking, it is Pierre's semantic *beliefs* that matter, not his semantic *knowledge*. It is what he believes that explains behaviour, whether what he believes is true or not. (On the other hand, *our* beliefs about whether his beliefs are true or not *do* play a rôle – not in explaining his behaviour, but in justifying certain interpretations of what he says.)

([12]) Another key assumption Kripke needs to generate his "puzzle" is that of the speaker's linguistic competence. The notion of linguistic competence is an imprecise one at the best of times, and involves a considerable degree of idealisation. Here it is clear that it cannot include competence with respect to the conventional use of names. This is fairly plausible, if not altogether unproblematic. But since, according to Kripke, analogous "puzzles" can be generated involving other kinds of terms (kind terms, definite descriptions), the question arises about the status of a speaker's beliefs about these *vis-à-vis* that speaker's linguistic competence. While ideal competence in a

one thing we know is that disquotation should not be applied to *both* of his assents. But it does not follow from this that it cannot be applied to either one, and to decide to which it can and to which it cannot we need to look at each sentence assented to separately and ask whether there is reason to think that Pierre holds unorthodox semantic beliefs about the referring expressions occurring in *it*. We obviously cannot infer from the fact that someone believes that 'A' and 'B' do not denote the same thing (when they in fact do so) that his beliefs about what 'A' denotes are mistaken *or* that his beliefs about 'B' are mistaken. His failure to recognize their co-referentiality may be a consequence of mistaken beliefs about what 'A' denotes *or* of mistaken beliefs about what 'B' denotes, either of which would be sufficient. (Though the failure may be overdetermined by ignorance of the actual denotation of both.(13)) Which of these is at the root of a person's error about a pair of co-denoting names he fails to recognize as such will vary from case to case and will be sometimes difficult to discover.

language may include standard beliefs about at least all terms other than proper names (which are, arguably, not terms *of* any language), there are looser senses of 'competence' which clearly do not. You and I don't meet the stricter criterion, and it would be unjustified to require that Pierre should. If we did require this, the "puzzle" would be less real-world looking and thus much less compelling. It *would* show that with a less than ideally linguistically competent speaker (i.e., one who knows less than *everything* about the language he speaks), speech is less than a fully reliable guide to belief. It would *not* show that this is so with ordinary speakers. The latter might be puzzling; the former is surely not.

All this goes not just for the extensions of terms, but for their meanings as well. Not only do I not have to know for *all* pairs of terms like 'gorse' and 'furze' that they are co-extensive in order to be a fully (though not *ideally*) competent speaker of English, I don't even have to have command of *all* synonymies. *Mutatis mutandis*, Pierre can be imagined to be completely bilingual, by any real-life standard, without knowing all standard French-English pairings. Part of the point I am trying to make about the principle of disquotation can be put by saying that it (or, rather, its *unrestricted* use) assumes that actually competent speakers meet a standard of ideal competence.

(13) Oddly enough, in the last kind of case we do not have even a sufficient condition for ignorance of co-refernetiality. Our subject may be making the *same* mistake about both names, in which case substitutivity will hold and disquotation will not get us into the kind of trouble Kripke finds with Pierre. Gaston may believe that both *'Londres'* and 'London' denote the same Canadian city; while mistaken, he would never manifest the kind of inconsistency Pierre does.

But, again, in Pierre's case things are made easy for us, for Kripke tells us enough to make one explanation much more likely than any other. The crucial difference between Pierre's assent in France and his assent in England is that the latter occurs *in* London – the city the referring expression in the sentence assented to actually denotes – while the former does not. This means that the possibility of ostension is present in the second case, but not in the first. Pierre-in-London would be likely to assent to 'This city is not pretty' as much as to 'London is not pretty', whereas Pierre-in-Paris would not assent to *'Cette ville est jolie'* as a substitute for *'Londres est jolie'*. (He *would*, of course, being sane, assent to *'Cette ville est jolie'*. But here our substitution test could work the other way: his willingness to substitute *'Paris est jolie'* would show that he took himself to be talking about Paris and that he held orthodox semantic beliefs about the name *'Paris'*.)

In general, the willingness of a subject to substitute an indexical term for a referring expression can anchor that referring expression in such a way as to betray his semantic beliefs about that term. So Pierre's willingness to substitute 'This city' in the subject position when in London tells us that he takes the name 'London' to denote London and thus that he is talking about London. His unwillingness to substitute *'Cette ville'* for *'Londres'* when in Paris tells us that he does not take *'Londres'* to denote Paris and thus that he is *not* talking about Paris. Why should that be interesting? Not because it helps us decide what he *is* (taking hemself to be) talking about. Obviously it doesn't do that. It is interesting only as a means of bringing out the contrast with the willingness to substitute the expression 'This city' for the name 'London'. If Pierre-in-Paris *were* willing to substitute *'Cette ville'* for *'Londres'*, that *would* justify our saying that he believed that Paris was pretty (in spite of mistakenly thinking that Paris was called *'Londres'*). Actual ostension always overrides ostension by what Ruth Marcus has called the "long arm" of a proper name.

It is important to see that our refusal to disquote with respect to Pierre's assent to *'Londres est jolie'* is not based on the principle that we may disquote *only* in cases where we have *established* that the subject's semantic beliefs *are* orthodox. That would be to employ an unreasonably rigorous standard of belief attribution. On the contrary, our practice is to disquote in the absence of any reason for suspicion,

and my verdict on Pierre is in conformity with that practice. With Pierre, we *do* have reason for such suspicion, provided by his subsequent assent to 'London is not pretty'. Since we know that in *that* assent he is referring to London, and since we believe him to be rational, we can infer that in assenting to *'Londres est jolie'* he can't have been talking about London.

But, it might be protested, we don't know any such thing at the time of his French assent. So what reason do we have *then* for not disquoting? The answer is, none, and we would indeed do so.

But would we have been wrong? In spite of all I have said, this still seems a reasonable question. We must, at least if we are realists about beliefs, distinguish the state of affairs that consists in someone's having a belief from the reasons we may have for thinking that that state of affairs exists. Putting it another way, the realist will refuse to identify the truth condition of a belief attribution with its warrantability condition. He will insist that whatever may be the evidential situation with respect to a belief attribution we make, there is some objective fact about the subject that makes that attribution true or false. Thus, he argues, there must be a fact of the matter about whether or not Pierre-in-Paris believed then that London was pretty, a fact quite distinct from and independent of what *we* were then, or are now, entitled to say about him. So was it true then that he believed that London was pretty? Does the account I have offered require us to say now that he didn't (because it requires us to say that he doesn't now) and that therefore our justified claim that he did was mistaken? If it required this, the account would be less attractive than its rivals, since it would require accepting one of two equally unpalatable alternatives: allowing the *truth-value* of a belief attribution to be affected by events or states of affairs subsequent to the time to which the attribution refers, or abandoning realism with respect to beliefs. Both of these are highly counter-intuitive and either would be a high price to pay to secure whatever other advantages the account may have.

But this consequence can be avoided. What we need to say is that in such a case our warranted belief attribution *may* have been mistaken, and our new evidence may encourage us to reconsider it. But whether it *was* mistaken or not doesn't depend either on the evidence (then *or* now) *or* on the facts *now*. The former has a bearing only on questions

of warrantability, not on questions of truth, as the realist rightly insists. And, obviously, the only belief attributions on whose *truth* the latter have any bearing are those that involve beliefs now. But it doesn't follow that facts about states of affairs obtaining now cannot, if they are known, bear on the *warrantability* now of attributions referring to earlier times. And of course our discovery that Pierre now believes that London is ugly bears on the judgement we should now make about the likely truth of our earlier *warranted* attribution of a conflicting belief to Pierre, on attribution made prior to that discovery.

The whole matter is a special case of the distinction between evidence-oriented and truth-oriented considerations in theorizing in general. No matter what the subject matter of our theory, these should not be conflated. They are distinct, through there is a connexion between them, which can be stated crudely as follows: the point of a theory of evidence is to show what kinds of evidence how used are most likely to lead to mostly true beliefs. And this is so when the beliefs in question are beliefs about beliefs, that is, when the science in question is the science of interpretation. But of course to unpack in any detail what this means, and how a particular theory of evidence itself may be justified over its rivals in terms of its truth-maximizing properties, is a very complicated matter. For my present argument all that needs to be brought out is that a theory of evidence for interpretation that allows unrestricted disquotation as a way of generating true belief attributions is likely to get things wrong more of the time than one that doesn't.

This does not mean that by using the restricted principle we will *always* get things right. One reason for this is that the range of evidence to which we need to pay attention is far wider than the strictly linguistic evidence to which alone disquotation is relevant. What people say is only one bit of evidence to which we may appeal in generating hypotheses about what they believe. It is, in most cases, a particularly important bit of evidence, but it is not a privileged one. The evidence provided by other, non-linguistic, behavior may override it, just as it may lead to the rejection of a hypothesis based on the evidence of some other bit of behaviour, no matter how compelling the latter by itself may be. ([14]) Thus sometimes a person's sincere disclaimers will overcome our inclination to interpret him according to

the standard correlations between some pattern of behaviour and some set of beliefs. Sometimes they will not. And sometimes we won't know what to say.

In Pierre's case, I have argued, his current behaviour is good evidence that he now believes that London is ugly. In the light of this, we must re-examine the evidence, which was good evidence earlier for inferring that he believed that London is pretty. We must conclude, I think, that the new evidence is more likely to lead to a true belief attribution than the old. But, first, that does not mean that it is *guaranteed* to do so, any more than the old. Second, the new evidence is quite compatible with continuing to say that the old evidence was, *at the time*, good enough for the hypothesis we based on it (and that it continues to be good enough for similar hypotheses about other Parisians). The relation between *that* evidence and *that* hypothesis is not affected by the availability now of different evidence. So, to the question Kripke insists on, namely, what does Pierre believe, the answer must be that he does not believe that London is pretty. Had Kripke asked about Pierre's belief at the time of Pierre's French assent and before the subsequent events, we would have said the opposite. But Kripke is asking us now, in the light of those subsequent events. And *now* the answer must be that Pierre *does* believe that London is not pretty and does not believe that London is pretty.([15])

So the natural objection to this way with Pierre's beliefs, that it is in conflict with what we would – and should – say about his untravelled fellow-Parisians, fails. True, if *they* assent to *'Londres est jolie'*, we must say about them that they believe that London is pretty. But that is just what we would say about Pierre, had he remained like them. As long as his Parisian neighbours remain untravelled, this will continue to be the right thing to say about them. We have every reason to think that they are talking about London and none that they were not. With *Pierre-in-London* the situation is reversed. But that is fine: *all* reasons we have for belief attribution are *defeasible (especially* in the absence

---

([14]) While I do not endorse the analysis of Kripke's puzzle of which it forms a part, I do agree with what I take to be Ruth Barcan Marcus' insistence on this point in "Rationality and Believing the Impossible," *The Journal of Philosophy*, LXXX, 6, esp. pp. 334 and 336.

([15]) This answer is defended in excruciating detail in Biro, op. cit.

of the almost-decisive evidence of ostention). Our belief attributions need not be infallible in order to be justified.

<center>V</center>

There is no question but that we observe caution in our use of disquotation in everyday interpretation. We take a person at his word only to the extent to which we can assume that he is using his words in the way we would use them. We normally operate on the assumption that the requirement of orthodoxy is satisfied and get cautious only when there is reason to doubt that it is. In trying to get at what people believe, we have to rely to a great extent on what they say. Most of the time, doing so gets us the right results, but there is always the possibility that it may lead us astray. We know this, but we also know that we cannot alow that possibility to lead us to give up on disquotation altogether. The art of interpreation consists in balancing such considerations in particular cases.([16])

Theories of interpretation and of meaning often fall into the trap of leaning too far in one or another equally tempting directions. They either pay too much attention to the subject's own understanding of his words or ignore this altogether in favour of the conventional, standard, intersubjective meanings of those words. Gricean theories are an example of the first error;([17]) the adoption of unchecked

([16]) For a more detailed discussion of these matters, see J.I. BIRO, "Meaning, Translation and Interpretation," *Australasian Journal of Philosophy* 59 (1981). It should by now be abundantly clear that I take the "puzzle" to be one about the *attribution* of belief. So I agree with DEVITT (op. cit.) that Kripke's title is misleading. To the extent that the puzzlement is generated by the principle of disquotation, it *can* only be about attributions, since that is the enterprise in which that principle is used. There may be other ways of getting at what someone believes (e.g., by discovering what mental representations he has and what rôle those play for him), but Kripke's story does not involve any of these. There *may* be a fact of the matter about what someone believes distinct from what we are justified in attributing to him on the basis of the evidence we have. There is *certainly* a fact of the matter about the second of these things. The question about disquotation is whether it is a reliable way of getting *that* fact right; I have argued that it isn't.

([17]) See J.I. BIRO, "Intentionalism in the theory of meaning," *The Monist* 62 (1979 and "Meaning and What is Said", in *Mind, language and Society* (ed. Otto Neumaier) Wien (Conceptus-Studien, 2) 1984.

disquotation is an instance of the second. What puzzlement is generated by Kripke's stories is the product of inadequate theory, and not of any incoherence in our ordinary concepts or practices.([18])

J.I. BIRO

University of Oklahoma
Dept. of Philosophy
455 West Lindsey Street, Room 605
Norman, Oklahoma 73019