Kripke's Puzzle And The Practice Of Logical Appraisal
Winner of the 1997 Larry Taylor Award
Scott Kimbrough

"So what's the puzzle?" That's the reaction many have had upon being introduced to Kripke's famous puzzle about belief.[1] In this paper, I offer an interpretation of Kripke's puzzle which explains both why this reaction is the natural one for well-adjusted individuals, and why Kripke and others nevertheless find themselves mired in paradox. I argue that the puzzle cannot be blamed, as Kripke contends, on an incoherence in our ordinary practices of belief attribution and logical appraisal, but must rather be blamed upon a conflict between those practices and Kripke's widely shared theoretical assumptions. Rather than simply stating those assumptions up front, however, I will begin with a description of our ordinary practices of belief attribution and logical appraisal which refutes Kripke's charge that those practices are incoherent. I will then isolate Kripke's puzzle-engendering assumptions by identifying the point at which those assumptions distort our ordinary ways and means for assessing charges of inconsistency. Finally, I will consider the implications of solving Kripke's puzzle by rejecting those assumptions.

The first order of business is to give a brief reminder of what Kripke's puzzle is. The protagonist of the puzzle, Pierre, is a Frenchman who believes that London is pretty. The attribution of this belief is based on Pierre's sincere testimony, in particular his assertion of the French sentence 'Londres est jolie.' If any monolingual Frenchman can believe that London is pretty, Pierre certainly does.[2] Later, Pierre moves to London and learns English without realizing that the names 'London' and 'Londres' are inter-translatable. His subsequent English testimony shows him to believe that London is not pretty. But then Pierre has contradictory beliefs – the beliefs that London is pretty and that London is not pretty. Kripke concludes that our ordinary practices of belief attribution and logical appraisal are incoherent: for Pierre can't be fairly charged with inconsistency, but surely anyone who has contradictory beliefs is inconsistent.

Or so Kripke would have us believe. However, by offering a description of our ordinary practice of making and answering charges of inconsistency, I will show that the possession of contradictory beliefs is not a sufficient condition for inconsistency or *self*-contradiction. For as we shall see, charges of inconsistency depend upon what the accused should take the logical relations among her beliefs to be, rather than upon what those logical relations in fact are. Thus, a space opens up between having contradictory beliefs and being inconsistent or contradicting oneself. Pierre, I shall show, rests comfortably in this space.

In levelling charges of inconsistency, we engage in arguments that are both *ad hominem* and *reductio ad absurdum*. For in these arguments, we a) restrict ourselves to the assumptions of the accused; and b) using only those assumptions, attempt to derive a contradiction. To highlight both of these elements, I call such arguments "*reductio ad hominem*" arguments. The point of *reductio ad hominem* arguments is to show the accused that she is *inconsistent by her own lights* or *contradicts herself*; given only her *own* beliefs as raw materials, she must admit that a contradiction can be derived.

To illustrate the practice of *reductio ad hominem* argumentation, consider a conspiracy theorist who wishes to reveal inconsistencies in the Warren report. Conspiracy theorists invariably have many beliefs which are not reflected in the Warren report, including (say) the belief that the FBI was involved in Kennedy's assassination. However, the conspiracy theorist cannot appeal to these beliefs in the context of a *reductio ad hominem* argument. She may, as many conspiracy theorists have, appeal to the alleged involvement of government agencies in an attempt to show that the Warren report is *inconsistent with the evidence*, but that is not to charge that the Warren report contradicts *itself*. To sustain this latter charge, our conspiracy theorist must set aside her conspiratorial views, drawing only on assumptions explicitly contained in the report and any additional assumptions upon which there is agreement (e.g., the assumption that a bullet cannot violate the laws of physics). Of course, in her view, many of the assumptions taken from the report are hopelessly false. But the truth or falsity of those assumptions is irrelevant in the context of a *reductio ad hominem* argument; all that matters is that the authors of the report take those assumptions to be true. If our conspiracy theorist can derive a contradiction while appealing only to assumptions accepted by the Warren commissioners, her attempt to reveal an inconsistency in the Warren report will succeed, failing otherwise. And so with charges of inconsistency generally: a person is inconsistent just in case a compelling *reductio ad hominem argument* may be mounted against her.

Can Pierre defend himself from the charge of inconsistency? His case may seem hopeless, since we apparently can derive a contradiction among Pierre's beliefs by simply listing the contradictory pair of beliefs about London. However, Pierre cannot be convicted of inconsistency quite so easily. For the judgment that Pierre's beliefs are contradictory cannot be sustained without appeal to assumptions that Pierre does not accept. Imagine how a *reductio ad hominem* argument with Pierre might actually proceed. Upon being accused of contradicting himself, Pierre would naturally claim that his beliefs are not contradictory because, although *Londres* is pretty, London most certainly is not. In short, Pierre would deny that London is Londres (and of course the corresponding meta-linguistic claim that 'London' and 'Londres' are inter-translatable[3])[4] And if Pierre were right about

this, we would have to admit by our own lights that there is indeed no more contradiction in his beliefs than there is between the beliefs of someone who believes that Paris is pretty and that London is not pretty. Of course, Pierre is certainly wrong to deny that London is Londres. The assumption that London is Londres is easily defended, for example by reference to French and English maps of Britain or to a French-English dictionary. So Pierre is obviously going to have to revise his beliefs. But the point remains that it is the indefensibility of his assumptions, not a lack of consistency, that forces the revision. For the fact that Pierre is wrong to deny that London is Londres does not make that assumption admissible in the context of *reductio ad hominem* argument against him. Consequently, Pierre is exonerated from the charge of inconsistency: a compelling *reductio ad hominem* argument cannot be mounted against him.5

It bears emphasizing that, in escaping the charge of inconsistency, Pierre does not *ipso facto* escape the charge of having contradictory beliefs. For the key to exculpating Pierre from the charge of inconsistency is to observe that he has certain *false* beliefs which lead him to *mistake* the logical relations among his beliefs. As we have seen, the falsity of Pierre's beliefs is no reason to discount them in the context of *reductio ad hominem* argumentation. However, when we are concerned to say what the logical relations among Pierre's beliefs actually *are* as opposed to what he *takes* them to be (and should take them to be in light of his global body of beliefs), the falsity of Pierre's assumptions is obviously relevant. Because Pierre is wrong to deny that London is Londres, his beliefs are in fact contradictory even though they do not seem contradictory to Pierre. For since London is Londres, Pierre's assertion of the sentence 'Londres est jolie' is correctly taken as an indication that he, like any other French speaker who asserts the same sentence, believes that London is pretty. And the beliefs that London is pretty and that London is not pretty are contradictory if any pair of beliefs is.6

Some philosophers will no doubt balk at admitting the possibility that a logically competent individual could mistake the logical relations among her beliefs. But there is nothing terribly paradoxical about this. Drawing on the vocabulary of formal logic, we may say that Pierre's error is at bottom an error in schematization: although the conjunction of his beliefs instantiate the contradictory schema 'Fa &~Fa', Pierre takes them instead to instantiate the satisfiable schema 'Fa &~Fb'. As anyone who has taken introductory logic knows, such errors in schematization do not undermine the erring individual's claim to proficiency in the formal apparatus, nor her commitment to logical principles such as the principle that all instances of the schema 'Fa &~Fa' are logically false. Of course, most people don't know what a schema is, but their "judgments" as to which beliefs instantiate which schemata are manifest in their judgments as to which arguments are logically analogous to which: Pierre take his beliefs about London to have the same logical relations as his beliefs that Paris is pretty and that Liverpool is not pretty, and different logical relations from the beliefs (at least one of which he will disavow) that New York is pretty and that New York is not pretty. And, as I have been urging all along, charges of inconsistency are directed at what the accused takes the logical relations among her beliefs to be, rather than what those logical relations in fact are.

The foregoing description of the practice of *reduction ad hominem* argumentation resolves the alleged incoherence in our practices of belief attribution and logical appraisal: since Pierre may have contradictory beliefs without thereby contradicting himself, Pierre's case does not threaten our practices. That is why I suggested in the introduction that it is natural for well-adjusted individuals to fail to see any puzzle in Pierre's case. But the coherence of our ordinary practices does not show that there is nothing to be puzzled about. As I have been hinting, Kripke's theoretical assumptions preclude his acceptance of the foregoing defense of Pierre's consistency, distorting the ordinary practice of *reductio ad hominem* argumentation. As we shall now see, Pierre's case becomes puzzling when we step back from the dialectical context in which charges of inconsistency are made and answered and turn our attention to the notion of belief content.

The content of a belief is expressed by the that-clauses of the belief attributions used to attribute it. So, for example, the content of the belief that London is pretty is expressed by the sentence 'London is pretty.' If, as Kripke persuasively argues, inter-translatable names such as 'London' and 'Londres' make the same contribution to the contents of the sentences in which they occur, then the sentences 'London is London' and 'London is Londres' express the same content. On the basis of this alleged equivalence of content, Kripke is prepared to conclude that Pierre believes that London is Londres provided he believes that London is London.7 In other words, Kripke accepts the following conditional:

(1) If Pierre believes that London is London, and if 'London is London' expresses the same content as 'London is Londres,' then Pierre believes that London is Londres as well.

Since Pierre does of course believe that London is London, Kripke concludes that Pierre believes that London is Londres. But I earlier staked Pierre's defense against the charge of inconsistency upon his disavowal of the belief that London is Londres; Pierre's defense succeeded only because there was a belief of ours that he did not share, but which was essential to the derivation of a contradiction. If Pierre's disavowal of the belief that London is Londres is rejected on the strength of (1), so too must his claim to consistency.

Why accept (1)? The details of Pierre's case certainly cry out against it. (1)'s claim to truth derives from the alleged truth of a general thesis connecting the

notion of belief content to the evaluation of belief attributions for truth or falsity. In particular, (1) is an instance of the following thesis, which I call "semanticism":

> Semanticism: If 'A believes that p' is true, and if the sentences replacing 'p' with 'q' express the same content, then 'A believes that q' is true as well.

With semanticism in place, the ordinary distinction between having contradictory beliefs and contradicting oneself collapses. For when an individual accused of inconsistency mistakes the logical relations among her beliefs as Pierre does, semanticism implies that she has the beliefs which should lead her to correct that mistaken assessment.[8] Given the role semanticism plays in distorting the practice of *reductio ad hominem* argumentation, I conclude that Kripke's puzzle is best interpreted as a conflict between semanticism and our ordinary practices. As such, if my earlier defense of Pierre's consistency is to stand, semanticism must be rejected.

While I favor solving the puzzle by rejecting semanticism, I do not claim that rejecting semanticism is necessary to solve Kripke's puzzle. Indeed, that is part of the point of interpreting the puzzle as a conflict between semanticism and our ordinary practices: the conflict may be resolved, like most any conflict, at the expense of either party. Nevertheless, my interpretation of the puzzle fruitfully clarifies the structure of most of the solutions to Kripke's puzzle that have been proposed in the literature. Those solutions are attempts, in various ways, either to amend ordinary practice in ways that conform to semanticism, or to maintain allegiance to ordinary practice by adjusting the notion of content at issue in semanticism. In the former camp, Ruth Barcan Marcus attacks the connection between sincere testimony and belief; and Millian philosophers such as Nathan Salmon dispute ordinary judgments as to the truth or falsity of belief attributions, introducing the notion of "guises" or "modes of presentation" under or through which contents or propositions are believed.[9] Other more Fregean philosophers dispute Kripke's claim that inter-translatable expressions like 'London' and 'Londres' make the same contribution to content, smoothing over the conflict between semanticism and ordinary practice by distinguishing the contents of beliefs which are not cognitively equivalent.[10] Importantly, however, none of these maneuvers are necessary unless semanticism is presupposed. Semanticsm is thus the unacknowledged fulcrum supporting the see-saw battle between Millian and Fregean attempts to solve Kripke's puzzle.

As further evidence of this last claim, consider the dialectical role semanticism plays in Kripke's paper. Kripke's goal is to deflect standard Fregean objections to Millian accounts of proper names. According to Millians, coreferring proper names

make the same contribution to content. Fregeans typically object to this account of the content of proper names by pointing out that coreferring proper names are not inter-substitutable *salva veritate* in belief contexts. So, for example, it may well be that a competent speaker (Jones) believes that Cicero was bald and that Tully was not bald. The substitution of 'Tully' for 'Cicero' yields the false belief attribution 'Jones believes that Tully was bald'. Worse yet, the substitution also yields the intolerable result that Jones contradicts himself.[11] Kripke *grants* the implication from Millianism to the conclusion that Jones contradicts himself. But he urges that Millianism is not to blame for this unfortunate conclusion since Jones's case is "just like Pierre's" – a case for which Millianism is not to blame.[12] The strategy is thus to embrace the absurd consequences in *both* cases, and to take their wide occurrence to indicate that Millianism is free from blame. But we have seen that semanticism is required to derive the absurd consequences in Pierre's case, and it is equally implicated in Jones's case. For the truth of Millianism and the belief attribution 'Jones believes that Cicero was bald' do not by themselves imply the belief attribution 'Jones believes that Tully was bald'. The following instance of the semanticist schema is required to bridge the gap between preservation of the *content* of the *beliefs attributed* by these two belief attributions and preservation of the *truth* of the *attributions themselves*:

> If 'Jones believes that Cicero was bald' is true, and if 'Cicero was bald' expresses the same content as 'Tully was bald', then 'Jones believes that Tully was bald' is true as well.

Given the role of semanticism in both Jones's and Pierre's cases, Kripke's acceptance of semanticism is necessary to his strategy for defending Millianism, representing the common commitment linking Fregean and Millian accounts of content.[13] This observation further strengthens my interpretation of the puzzle, revealing a strong motivation for Kripke to downplay – or perhaps even not to notice – the role semanticism plays in Pierre's case.

As I have already mentioned, I favor solving Kripke's puzzle by rejecting semanticism. To defend this approach fully, I would of course be obliged to address the positive motivations for semanticism. For example, semanticism follows from the epistemological view that sameness and difference of content is detectable upon introspection, and from the popular metaphysical thesis that beliefs are relations to contents.[14] Rather than discuss motivations for semanticism, however, I will dedicate my remaining time to fleshing out one of the most important consequences of rejecting semanticism.

By helping to smoke out the distorting influence of semanticism on our ordinary practices, Kripke's puzzle does the valuable service of helping to reveal

semanticism as the common assumption of Fregean and Millian theories. Once semanticism is out in the open as a controversial thesis, it becomes clear that the choice between Millian revisionism of ordinary practice and Fregean retreats from translation is not exclusive. In my view, this realization presents an important challenge for both of these major traditions in the philosophy of language: until and unless some positive motivation for semanticism can be provided,[15] neither Fregeans nor Millians can justify their respective excesses.[16]

Notes

1. Kripke, "A Puzzle About Belief", from *Meaning and Use*, edited by A. Margalit (Dordrecht: Reidel, 1979), 239-283. All subsequent references are to this paper, unless noted otherwise.

2. Kripke defends the belief attributions he makes in the course of the puzzle by appeal to two principles which he plausibly claims are implicit in the ordinary practice of attributing beliefs:
The disquotational principle: If a normal English speaker, on reflection, sincerely assents to 'p', then he believes that p. (pp. 248-9)
> 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). (p.250)
I do not feature these principles in my exposition since, as Kripke notes himself (p. 254), the particular principles are less important than the fact that the belief attributions derived with their help are indeed sanctioned by ordinary practice, as I agree they are.

3. Many philosophers would insist upon sticking to the meta-linguistic level of description in Pierre's case. However, to do so is incorrect. For since Pierre is a competent speaker of both French and English, his uses of the names 'London' and 'Londres' are literal expressions of belief no less than the uses of other competent speakers. Furthermore, given meta-linguistic assumptions such as the assumption that 'London' and 'Londres' refer to different cities, Pierre is presumably savvy enough to infer that London is not Londres. For a thorough discussion of a related point, see Burge "Belief and Synonymy," *Journal of Philosophy Vol. 75 No. 3 (March 1978)*.

4. Kripke's distaste for "word salads" notwithstanding, I do not see any problem with using the name 'Londres' in English sentences. In the present context, there is certainly no failure of intelligibility, and the threat of such failures strikes me as the only reason for the general prohibition against the use of foreign expressions in English discourse. At any rate, the use of 'Londres' is eliminable in favor of 'London1' and 'London2' or some such. For all that is required is some means for making clear which of Pierre's beliefs we are talking about – the beliefs about London that he acquired in France, or the beliefs about London that he acquired in England.

5. To accuse Pierre of inconsistency would be like accusing the authors of the Warren report of inconsistency on the grounds that they knew of an FBI agent named 'Oswald'. For so long as the commissioners *believe* that the Oswald in question is not Lee Harvey Oswald, there is no basis for the accusation that they believe that the FBI both was and was not involved in Kennedy's assassination. And this is true even if the Oswald on the FBI payroll was none other than Lee Harvey himself! For as in Pierre's case, there is an assumption without which a contradiction cannot be derived, but which is

not shared by the accused. This articulation of the example has links to Kripke's Paderewski case, in which Peter believes both that Pederewski had musical talent and that Paderewski did not have musical talent on account of believing that Paderewski the Polish statesman is not Paderewski the Polish pianist. In both these cases, as in Pierre's, there are assumptions which the accused do not share but which are essential to showing that the beliefs of the accused are contradictory. In the Oswald case, the belief in question is the belief that Oswald the FBI agent is Lee Harvey Oswald. In the Paderewski case, the belief is that Paderewski the statesman is Paderewski the pianist. Note that the fact that the relevant identity belief is (let us suppose) false in the Oswald case and true in the Paderewski case has *no effect whatsoever* on the proper outcome of the relevant *reductio ad hominem* argumentation.

6. Some may suspect that the key to exculpating Pierre is to notice that he is unwilling to accept the sentence 'London is pretty' as an accurate expression of his belief, so that there are no contradictory sentences to which he is disposed to assent. However, while this may be a hopeful strategy for understanding Pierre's case, it cannot eliminate puzzle cases generally. In the Paderewski case, Peter is disposed to assent to each of the contradictory sentences 'Paderewski has musical talent' and 'Paderewski does not have musical talent'. I discuss the Paderewski case in more detail in the previous note.

7. Kripke does not discuss this particular inference. As evidence of his willingness to draw content-preserving inferences, however, see p.241.

8. For this reason, I reject the interpretation of the puzzle offered by Joseph Owens. See "Cognitive Access and Semantic Puzzle" (in Anderson and Owens (eds) *Propositional Attitudes*, CSLI Lecture Notes (1990)) and "Contradictory Belief and Cognitive Access" (in French et. al. (eds.) *Midwest Studies in Philosophy Volume CIV*, University of Notre Dame Press (1989)). Owens claims that the puzzle follows from the imposition of a Cartesian epistemology which guarantees that individuals can always tell whether pairs of beliefs have the same or different contents. However, Pierre's defense against the charge of inconsistency is staked upon his disavowal of the belief that London is Londres. If he believes that London is Londres, then he should be able to derive a contradiction whether or not he knows that this belief has the same content as the belief that London is London. This conclusion may be avoided if we deny that Pierre knows that he believes that London is Londres. But that is to deny Pierre self-knowledge – the knowledge that he has the beliefs he has – rather than the knowledge that some pair of beliefs have the same contents. And Owens agrees that our claim to self-knowledge is not based in an objectionable Cartesian epistemology. In fairness to Owens, the issues here are very complex, and he is not without avenues of defense. I hope to discuss this issue further in the future.

9. See Marcus's "A Proposed Solution to Kripke's Puzzle About Belief," in *Midwest Studies in Philosophy, VI, ed. French et. al. Minneapolis, MN: University of Minnesota Press, 1981; and Salmon's* book Frege's Puzzle, *Cambridge, MA: A Bradford Book, MIT Press, 1986.*

10. See, for example, Brian Loar's discussion in "Social Content and Psychological Content", from R. Grimm & D. Merrill (eds.) *The Contents of Thought*, The University of Arizona Press: Tucson (1988).

11. Note that, in this case, the beliefs that Tully was bald and that Tully was not bald are contradictory by Jones's lights as well as ours. Consequently, the only way for Jones to defend himself from a *reductio ad hominem* argument is to disavow the belief that Tully was bald. But if both Millianism and semanticism are in place, his disavowal cannot be taken seriously, and his defense against the charge of inconsistency will accordingly fail.

12. p.268.

13. It is worth pointing out that Kripke's defensive strategy for Millianism may nevertheless be successful as an *ad hominem* strategy against any critic of Millianism who accepts semanticism. That is, Kripke can deploy his puzzle to show that Millians are in no more puzzling a position than their Fregean opponents. At least, he can do this against a Fregean opponent who also grants that inter-translatable terms make the same contribution to the contents of the sentences in which they occur.

14. I discuss this latter motivation of semanticism in "Belief Content and Compositionality", forthcoming in *Southwest Philosophy Review*. The central focus of that paper is to show that semanticism cannot be justified by appeal to the compositionality constraint in semantic theory.

15. The first defense that comes to most philosophers' minds when faced with this challenge is the need to accommodate the rationality of individuals like Pierre and Jones or to preserve the integrity of common sense psychological explanation. However, such appeals are question-begging in the present context, since the only way to question the rationality of Pierre and Jones, or to throw a wrench into the workings of commonsense psychological explanation, is by way of drawing content-preserving substitutions in the that-clauses of belief attributions. But such substitutions are sanctioned only if semanticism is presupposed.

16. Thanks to Gary Ebbs and Thomas Ricketts for helpful comments and criticisms.