

WHAT PUZZLING PIERRE DOES NOT BELIEVE

David Lewis

Kripke's puzzle about belief refutes a certain simple analysis of belief sentences. This analysis fails for another reason as well, since it requires believers to have a knowledge of essences which they do not in fact possess.

The Lesson of Pierre

The case of Pierre is presented by Saul Kripke. It runs as follows.¹

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:

(1) Pierre believes that London is pretty.

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 neighbours, rarely ever leaves this part of the city. None of his neighbours knows any French, so he must learn English by 'direct method'; 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. Pierre learns from them everything they know about London, but there is little overlap with what he heard before. He learns, of course, to call the city he lives in 'London'. Pierre's surroundings are 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: 'London is not pretty'. Of course he does not for a moment withdraw his assent from '*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. After Pierre lived in London for some time, he did not differ from his neighbours either in his knowledge of English or in his command of the relevant facts of local geography. Now Pierre's neighbours 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

¹ Saul A. Kripke, 'A Puzzle about Belief', in Avishai Margalit, ed., *Meaning and Use* (Reidel, 1979). The quoted material comes from pp. 254-257, with deletions and renumbering of sentences.

should say the same of him. But then, on the basis of his sincere assent to 'London is not pretty', we should conclude:

(2) Pierre believes that London is not pretty.

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. But there seem to be insuperable difficulties. We may suppose that Pierre is a leading philosopher and logician. He would *never* let contradictory beliefs pass. And surely anyone is in principle in a position to notice and correct contradictory beliefs if he has them. 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.

(3) He cannot be convicted of inconsistency; to do so would be incorrect.

Kripke presents the case as an unsolved puzzle. His principal moral is that whatever is responsible for this puzzle may also be the source of difficulties commonly blamed on failures of substitutivity.

To solve the puzzle would be to show how (1), (2), and (3) are compatible. That would require an adequate analysis of (1) and (2) that would not represent them as ascribing inconsistent beliefs. I have no such solution to offer, in part because some of the makings of the sort of solution I might favour are not mine to present.²

Instead, I shall make a negative point. There is a natural and straightforward analysis of (1) and (2) that does represent them as ascribing inconsistent beliefs. The case of Pierre refutes this analysis, thus clearing the way — let us hope — for some less simple but more accurate successor. The refuted analysis consists of three parts.

(4) 'Pierre believes that $F(A)$ ', where A is an ordinary proper name and F is an easily understood predicate, ascribes to Pierre a belief whose object is the proposition (actually) expressed by ' $F(A)$ '.

(5) This proposition holds at exactly those possible worlds where the thing which is (actually) denoted by A has the property which is (actually) expressed by F .

(6) Beliefs are jointly inconsistent if there is no possible world where their propositional objects hold true together.

Given that we accept (1), (2), and (3), as I think we clearly must, this analysis stands refuted. For 'London' is an ordinary proper name, which (actually, and in this context) denotes the city London; 'is pretty' and 'is not pretty' are easily understood predicates which (actually, and in this context) express strictly contradictory properties; so there is no possible world where anything, be it London or anything else, has the properties expressed by those two predicates.

I think the refuted analysis has been held, but let me not stop to point the finger. It surely deserves to be held, but for its failures: it is simple and plausible,

² I have in mind some ideas about the analysis of belief sentences suggested by Robert Stalnaker (personal communication, 1979).

and it fits perfectly into a systematic program for compositional semantics. If it did not exist, it would be necessary to invent it — and to refute it.

Several remarks before we go on. First, the refuted analysis as I stated it is limited in scope, and thereby dodges some extra problems that would plague its generalisations. Since A is an ordinary proper name, it is not a fictitious name with no actual denotation. Likewise it is not first-person or present-tensed, so we avoid problems about self-descriptive belief. Since F is easily understood, it is not one of those semi-technical terms like 'elm', 'water', or 'arthritis', that are used at least somewhat competently by laymen who lack full mastery of their meanings; neither does it have any confusingly complex internal structure.

Second, the refuted analysis has variants which equally stand refuted. There is the counterpart-theoretic variant: we amend (5) to consider not worlds where the thing denoted by A is itself present to have the property expressed by F, but rather worlds where that thing has that property vicariously, through a unique counterpart which is united to it by some sort of resemblance and which stands in for it at that world.³ (A variant permitting multiple, rather than unique, counterparts would so far escape refutation on a technicality; but would fall victim to the considerations discussed in the next section.) Thus the London of our world is vicariously pretty at a world where it does not exist, strictly speaking, if it has at that world a unique counterpart and that counterpart is pretty. Such a world might well be counted as one where the proposition (actually) expressed by 'London is pretty' holds.

Also there is the ersatz variant: we systematically replace possible worlds and individuals, and properties thereof, by stand-ins constructed entirely out of the resources of this world.⁴ An ersatz variant will be counterpart-theoretic as well, unless the ersatz worlds are so constructed that London itself, for instance, always is used as the ersatz for any unactualised alternative London.

Third, notice that to refute the analysis given by (4)-(6) is not yet to refute the theory that beliefs have propositional objects, where a proposition is something fully characterised by the set of worlds where it holds. It is one question what the objects of belief are, another question how the 'that'-clause of a belief sentence specifies the object of the ascribed belief. In fact, I think the refuted analysis is wrong on both points,⁵ but in this paper I shall be content to show that there's something wrong with it somewhere. And there is, as witness Kripke's puzzle.

Ignorance of Essence

However, we do not really need the whole of the puzzle of Pierre to see that the refuted analysis goes wrong. False accusations of inconsistency between beliefs are but a symptom. The cause of the trouble can be seen even if we consider

³ See my 'Counterpart Theory and Quantified Modal Logic', *Journal of Philosophy* 65 (1968), pp. 113-126; and *Counterfactuals* (Blackwell, 1973), pp. 39-43.

⁴ See, for instance, the construction of ersatz worlds in W. V. Quine, 'Propositional Objects', in his *Ontological Relativity and Other Essays* (Columbia University Press, 1969).

⁵ On objects of belief, but not the analysis of belief sentences, see my 'Attitudes *De Dicto* and *De Se*', *The Philosophical Review* 88 (1979), pp. 513-543.

beliefs one at a time. Let us stick to the case of Pierre, and consider whether the refuted analysis accounts even for the truth of (1). I say it does not. (It likewise fails to account for the truth of (2).) Pierre does not have as an object of his belief the proposition (actually) expressed by 'London is pretty'. For there is a possible world which fits Pierre's beliefs perfectly — it is one of his 'belief worlds' — at which that proposition is false.

I have in mind a world where the beautiful city Pierre heard about was not London but Bristol. Imagine a world just like ours until fairly recently (except to the extent that it must differ to fit Pierre's misconceptions about earlier history, if any). Then the beautification of Bristol was undertaken, and at the same time it was renamed in honour of Sir Ogdred Londer. The French called this famous city '*Londres*'; they spoke often of its beauty, and all they said was true. In due course Pierre heard of the beauty of Bristol, lately called 'Londer' in England and '*Londres*' in France, and he came to assent sincerely to '*Londres est jolie*'. What happened at his end was just like what happened at the real world.

While Bristol was beautified, London fell into decay. The better parts were demolished — copies sometimes were built in Bristol, alias 'Londer' — and only the slums remained. London became ugly through and through. Also, nothing of consequence happened there. The French had little occasion to speak of the place under any name, and indeed it never was mentioned in Pierre's presence. It was to this place that the unfortunate Pierre was made to go. Again, what happened at Pierre's end of his encounters with London was just like what happened at the real world.

This world fits Pierre's beliefs perfectly. For all that he believes, it might very well be the world he lives in. Tell him and show him all about it, claiming that it is the real world; he will never be at all surprised, unless it surprises him to find that he has turned out to be right in all his beliefs without exception. Nothing he believes — no propositional object of his belief — is false at this world.

However, the proposition (actually) expressed by 'London is pretty' according to (5), or a counterpart-theoretical or ersatz-theoretic variant thereof, is false at this world. 'London' denotes London and 'is pretty' expresses the property of being pretty, and this is a world where London, or a good counterpart or ersatz, is present and is not at all pretty.

(In a parallel way, we could find a world which fits Pierre's beliefs, but in which the proposition expressed according to (5) by 'London is not pretty' is false. We could even find a belief world for Pierre where the propositions expressed by 'London is pretty' and 'London is not pretty' both are false. Let the city he heard of in France be a beautified Bristol, renamed 'Londer' or '*Londres*'; let the city he fetches up in be a worsened Manchester, renamed 'London'; and let London itself be absent altogether.)

Therefore (4) and (5) are not both true. ((6) is not involved in this refutation.) I see no grounds for blaming the trouble on (5); I take (5) to be correct, either as it stands or when restated in terms of counterparts. The culprit is (4). The role of the 'that'-clause in a belief sentence is not, in such cases as this, to express the proposition which is the object of the ascribed belief.

The error of (4) amounts to ascribing a knowledge of essences that we may

not in fact possess. Let us define the *essence* of London as that property that belongs to London at every world where it exists, and to nothing else at any world. (Or it belongs to all and only counterparts of London, or to all and only ersatz Londons.) What Pierre's belief that London is pretty is about is that which plays the role of London for him; or rather, that which plays one of the London-roles for him. It may or may not be London. If being the X is London's essence, it may or may not be the X. Having London's essence is not a prerequisite for playing the London-role. Thus Pierre does not believe that London is the X, or even that London is the X if anything is; and this despite the fact that he believes the proposition that holds at every world, and every world is a world where London is the X if anything is. Because of this ignorance, he is unable to get from his belief that London is pretty to a belief having as object the proposition that the X is pretty — or in other words, to the proposition expressed by 'London is pretty'.

Objections and Replies

Objection. Why are you entitled to assume that the so-called Bristol (or 'Londer', or 'Londres') of the counterexample world is not really London? *I reply:* on my view of these matters, it isn't London because there is no identity across worlds; and it isn't a counterpart of London because it isn't very much like London, and because there is a rival candidate that resembles the real London very much better. This is so if we consider location; or match of origins; or resemblance of histories; or resemblance of present geography; or even, not completely but to a sufficient degree, resemblance of landmarks. (All the ugly landmarks of London and all the pretty landmarks of Bristol remain. Although Bristol does have its copies of the Houses of Parliament and Tower Bridge and one or two more, these being the ones Pierre was shown pictures of when he was in France, they're quite poor copies — he was shown quite poor pictures.) On an opposite view, according to which there is identity across worlds and it varies quite independently of qualitative character, there are worlds of the sort I imagined where the so-called Bristol is really London, and others where it isn't. I hereby stipulate that the counterexample world is one of the latter ones. On an intermediate view, essentialist but Haecceitist, there are no qualitative conditions sufficient for identity across worlds, but match of origins is a necessary condition; and the so-called Bristol fails to satisfy this qualitative prerequisite for identity with the real London. It may or may not be the real Bristol, but London it definitely isn't. These three views seem to cover the spectrum of reasonable positions fairly well.

Objection. The so-called Bristol is a counterpart of London at least *for Pierre*. His epistemic rapport with it at the counterexample world is just like one part of his epistemic rapport with London at the real world. In this respect, the otherworldly Bristol resembles the real London. It is London if, in Jaakko Hintikka's terminology, we cross-identify by acquaintance.⁶ *I reply:* yes. This is

⁶ See his 'The Logic of Perception', in his *Models for Modalities* (Reidel, 1969), and in Norman S.

true, and relevant, and perhaps something that would figure in a satisfactory analysis of (1) and (2). But it cannot save the refuted analysis. A counterpart by acquaintance for Pierre is not a counterpart *simpliciter*; the relativity to Pierre is not provided for in (4) and (5). They are stated in terms of a sentence expressing a proposition, and our notion of expressing includes no such relativity. Other relativities, yes: to language, and to context. But not the relativity involved in cross-identification by acquaintance for Pierre.

Objection. You forget Kripke's emphatic stipulation (*op. cit.*, pp. 242 and 246) that he is speaking always of belief *de dicto*, never of belief *de re*. Given this stipulation, (1) and (2) *must* be interpreted according to (4) and (5). *I reply* that Kripke may stipulate that (1) and (2) are to be interpreted according to a certain analysis, or he may invite us to rely on our intuitions about their truth. But he cannot do both without begging the question in favour of the preferred analysis. However, it is not at all clear to me what Kripke is stipulating. The notion that ordinary language belief sentences divide into *de dicto* and *de re*, without residue or overlap, may be part of an oversimplified semantic analysis. The stipulation that (1) and (2) are not *de re* seems clearer and safer than the stipulation that they are *de dicto*; for we can support it by a parallel puzzle in which no *res* is available. (Pierre has been told in France that 'Pere Noel' brings presents to all the children, and has been told in England that Father Christmas brings presents only to the good children. He reckons that good children get double shares.) But the stipulation that (1) and (2) are something other than *de re*, even if legitimate, is not enough by itself to settle that we must interpret them by (4) and (5).

Objection. 'Believes that' and 'believes the proposition that' are synonymous. *I reply:* maybe so. But in that case, 'believes the proposition that' must not be analysed as 'has a belief with the propositional object expressed by' (with quotation marks supplied), at least not if the latter bears the meaning I have given it in this paper.

Objection. The counterexample world is not a world that fits Pierre's beliefs. For Pierre believes that London is pretty, whereas the counterexample world is one where London is not pretty. *I reply* by posing a dilemma. When we characterise the content of belief by assigning propositional (or other) objects, are we characterising an inner, narrowly psychological state of the believer? Are beliefs in the head? Or are we characterising partly the believer's inner state, partly the relations of that state to the outer world?⁷ If it is the latter, the objection may succeed; however, Kripke's puzzle vanishes. For if the assignment of propositional objects characterises more than the believer's inner state, then there is no reason to suppose that a leading philosopher and logician

Care and Robert M. Grimm, eds., *Perception and Personal Identity* (Case Western Reserve University Press, 1969).

⁷ Compare characterising perceptual experience without regard to whether it is veridical versus characterising the perceiver's accomplishments in gaining information about the world around him. I owe this comparison to John Perry.

would never let contradictory beliefs pass, or that anyone is in principle in a position to notice and correct contradictory beliefs if he has them. Anyone is in principle in a position to notice and correct a state of the head which can be characterised by assigning contradictory propositional objects, but why should philosophical and logical acumen help him if the trouble lies partly outside? As soon as we accept the consistency of Pierre's beliefs as a datum — as I did, on Kripke's invitation — we are committed to the narrowly psychological conception of belief and its objects. (I would like to think that this was what Kripke intended in instructing us to consider belief *de dicto*.) But on the narrowly psychological conception, the counterexample world does fit Pierre's beliefs, as witness the fact that it would not at all surprise him to be persuaded that the world was just that way. To be sure, it is not a world where London is pretty. That only means that (1) is not a narrowly psychological characterisation. Indeed not; we can see that directly. Someone might be exactly like Pierre psychologically and yet not believe that London is pretty. The Pierre-counterpart at the counterexample world is one such person. Another is an actual Pierre-counterpart on Twin Earth, long ago in a galaxy far, far away. His neighbours might say of him 'He believes that London is pretty'; but *we* have no business saying it, given what 'London' means in our mouths and given his isolation both from our London and from our 'London'. Since (1) is not a narrowly psychological characterisation of Pierre's beliefs, it is irrelevant to the question whether the counterexample world fits Pierre's beliefs on the narrowly psychological conception.⁸

Australian National University
and
Princeton University

Received August 1980

⁸ I am grateful to the Australian National University for research support, and to several philosophers there and elsewhere for valuable discussion. Special thanks are due to Nathan Salmon.