

# Philosophical Papers

---

Volume I

DAVID LEWIS

New York Oxford  
OXFORD UNIVERSITY PRESS  
1983

Copyright © 1983 by David Lewis

Library of Congress Cataloging in Publication Data

Lewis, David, 1941-  
Philosophical papers.

Includes index.

1. Philosophy—Addresses, essays, lectures.

I. Title.

B29.L49 1982 110 82-22551

ISBN 0-19-503203-9 (v. 1)

ISBN 0-19-503204-7 (pbk. : v. 1)

Printing (last digit): 9 8 7 6 5 4 3

Printed in the United States of America

## *Postscripts to* “Radical Interpretation”

### A. KARL AND OTHERS OF HIS KIND

I stated my problem in an unduly individualistic way: given the facts *about Karl* as a physical system, solve for the facts *about him* as a person—*his* beliefs, desires, and meanings. If Karl were a unique being, this would be the right question to ask. If not—if he is, for instance, human—it is not. In “Mad Pain and Martian Plain” (in this volume), I argued that a “madman” might be in pain not because his state occupied the causal role of pain in him but rather because that state occupies that role, for the most part, in members of the kind to which he belongs. The same possibility should be recognized for attitudes as well. Karl might believe himself a fool, and might desire fame, even though the best interpretation of Karl considered in isolation might not assign those attitudes to him. For the best interpretation of Karl’s kind generally might be one that interprets two states respectively as belief that one is a fool and as desire for fame, and Karl might be in those two states.

Compare a less controversial case. We have a certain kind of mass-produced calculator: the Texas Instruments 58C, let us say. A certain hardware state S of the 58C is the state of having the number 6099241494 stored in register 17. The state S is to be thus interpreted because of its causal role in the functional organization of the 58C. Now suppose that one 58C comes out defective. It still *is* a 58C, and the defect is not such as to prevent it from ever being in state S; but the defect does mean that the role of S in this calculator differs from the normal role of S in the 58C. Considering the defective calculator in isolation, there is no reason to interpret S as the state of having 6099241494 in register 17. But that is how we interpret S for the 58C generally, so that is what S is for any 58C—even the defective one. We would say, and rightly so, that when the defective 58C is in state S, then it does have 6099241494 in register 17.

An interpretation just of Karl at the present moment need only specify his attitudes and his meanings. But an interpretation of Karl’s kind generally—or even of Karl himself as he is at various times, or as he might have been under various different circumstances—must be something more complicated. It must be a scheme of interpretation specifying the attitudes and meanings as a function of the momentary total physical state. On the basis of such states, the scheme assigns interpretations to individuals at times. (Indeed it might—and should, I think—do this simply by identifying certain attitudes with certain (partial) physical states.)

The best scheme is the one that does the best job overall of conforming to the constraining principles, taking one individual and time with another. (The individuals in question being not only Karl and others of his kind as they actually are, but also some of their might-have-been counterparts.) The best interpretation for Karl is the one assigned to him by the scheme of interpretation that does best overall, even if it does not do so well in his exceptional case. And the best scheme is *ipso facto* correct: to believe that one is a fool, or to desire fame, is to be such that the best scheme for your kind assigns these attitudes to you.

(What if there is no unique best scheme? Suppose there are several that beat the rest, these are tied or incomparable, and they yield conflicting assignments of attitudes and meanings. Then to the extent that there is conflict, Karl's attitudes and meanings are indeterminate. And what if it is not clear which of all the broader and narrower kinds Karl belongs to should be taken as "his kind" in determining the best scheme of interpretation for his kind?<sup>1</sup> As I noted in "Mad Pain and Martian Pain," we cannot rely on our criteria of selection to settle hard cases unequivocally. If they do not, and if different choices favor schemes that assign conflicting attitudes and meanings to Karl, we have here a further source of indeterminacy. Given indeterminacy, there are two natural senses in which we may nevertheless speak of Karl's attitudes and meanings. (1) We may take only what is agreed upon between the conflicting eligible schemes. Or, more daringly, (2) we may take him to have all the attitudes and meanings assigned by all the conflicting eligible schemes. The problem resembles that of truth in inconsistent fiction, discussed in Postscript B to "Truth in Fiction" (in this volume), and the alternative proposed solutions to the problems are parallel.

Any broadly functionalist theory of mind is under intuitive pressure from two directions. On the one hand, it seems wrong to make it invariable or necessary that the mental states occupy their definitive causal roles. Couldn't there be occasional exceptions, comparable to the case of the defective calculator? On the other hand, the mental states of Karl seem intrinsic to him. Why should whether he now feels pain—or believes himself to be a fool, or desires fame—depend on what causes what in the case of someone else? I do not see any acceptable way to respect both intuitions in their full strength. Individualistic functionalism respects the second at the expense of the first. I have tried to strike a more credible compromise, one that grants some of the force of each. The first intuition is respected to this extent: indeed it is possible that there be exceptional cases in which a mental state fails entirely to occupy its definitive role. However, such cases must be exceptional, such failures cannot be too common. The second intuition is respected to this extent:

<sup>1</sup>It might indeed turn out that humans have such vast differences in hardware from one to the next (unlike 58C calculators) that the salient kind does after all contain only Karl and some of his counterparts. If so, the correction I have been advocating will turn out to make little difference. Still, I do not think we should adopt an individualistic approach to interpretation on principle, even if we should be prepared to have it forced on us by the facts of individual difference.

one's mental states are intrinsic states that one is in. However, what makes one's states be the mental states they are—what makes them occupy the role they do—is not entirely intrinsic. To a limited extent it concerns others of one's kind. But that extent *is* limited since, necessarily, most cases are not exceptional.

## B. THE SYSTEMS OF ATTITUDES

In view of my arguments in "Attitudes *De Dicto* and *De Se*" (in this volume), of course I must withdraw my statement that **Ao** and **Ak** are to be specifications of Karl's *propositional* attitudes. His egocentric, irreducibly *de se* attitudes should not be left out, although these are not (in my sense of the word) propositional. But even these might be expressed, in context, by sentences of our language, or of Karl's:<sup>2</sup> we must use first-person sentences, and we must regard these as expressing properties rather than propositions. Thus Karl may believe (to a certain degree, at a certain time) that he is a fool; this belief is expressed in our language by "I am a fool", but only if this is regarded as expressing the property of being a fool rather than the proposition that Karl is a fool. (What if Karl thinks that he himself is a fool and Karl isn't, from which he concludes that he must not be Karl?) There is of course a good sense in which a first-person sentence does express a proposition, but this sort of expression is beside the point for our present purposes.

The ascriptions of belief that comprise **Ao** will not much resemble the belief sentences of ordinary language. (Likewise *mutatis mutandis* for the ascriptions of desire.) For our object is to specify belief in a narrowly psychological sense: the belief that governs behavior, the belief that is "in the head", determined by brain states and their causal roles. But our ordinary belief sentences often are not narrowly psychological. Take the case of puzzling Pierre.<sup>3</sup> He believes that the city he has heard of under the name of *Londres* is pretty—this much is narrow psychology and belongs in **Ao**. The city he has heard of under that name is London—that is not a psychological matter at all. Thanks to this mixed situation, only partly psychological, an ordinary belief sentence is true: Pierre believes that London is pretty. This is *not* an ascription that belongs in **Ao**. If Pierre had been just the same psychologically, but otherwise situated, the city he had heard of as *Londres* might have been Bristol. Then he would have believed that Bristol, not London, was pretty. But the beliefs assigned to him in our interpretation should have been just the same.

<sup>2</sup>Here I am simplifying matters by assuming what may well be false: that the languages have resources adequate to describe the complex shapes, tunes, etc. that may figure in Karl's attitudes.

<sup>3</sup>Discussed in Saul Kripke, "A Puzzle about Belief," in Avishai Margalit, ed., *Meaning and Use* (Dordrecht: Reidel, 1979); and in my "What Puzzling Pierre Does Not Believe," *Australasian Journal of Philosophy* 59 (1981): 283–89.