

V. *Two Arguments against the Identity Theory of Mind*

IN discussions of the identity theory of mind, there is constant recourse to two related types of argument, from ordinary language usage, to the effect that the theory in question is either false or meaningless. We can refer to the two arguments under discussion as the category argument¹ (CA for short) and the meaningfulness argument (MA for short). If either one of these arguments were well founded we could decide *a priori*, without waiting for further research in the relevant sciences, whether or not the mind could be identified with processes of the central nervous system. The answer would be in the negative. It is proposed to examine these two arguments here, and to question the persuasive simplicity with which they have been frequently put forward. If they do not bear up well under closer scrutiny we may not, of course, conclude to the truth of the identity theory; we will conclude only that it has not been *a priori* eliminated, by these arguments, as a contender for the solution of the mind-body problem.

The category argument

The CA is used in other contexts besides discussions of the mind-body problem, and the legitimacy of its application in many cases is not under discussion here. In fact we can more clearly specify the use of the CA which is being rejected by distinguishing two general types of application of the CA, and labelling them CA(S) and CA(D). CA(S) refers to the CA when it is used within a single conceptual framework, while CA(D) refers to the same type of argument when it straddles conceptual frameworks in such a way that two distinct conceptual frameworks are involved in the ways in which we conceive of the two entities whose identity, or otherwise, is under discussion. Only the latter type of CA, namely the CA(D), is considered to be invalid, and it is precisely this type of argument which is often used against the identity theory. However, even the use of the CA(S) will have to be qualified in such a way that it becomes quite ineffective in discussions of the identity theory.

1. See Keith Campbell, *Body and Mind* (New York: Anchor Books; London: Macmillan, 1970) 31.

The general structure of the CA, of both sub-types, should become clear with the following examples.

The major premise of the CA has the following form:

- (i) x is identical with y only if what can be predicated of 'x' can also be meaningfully predicated² of 'y' and vice versa.

This is not the classical principle of the identity of indiscernibles, but it is close to the principle of the indiscernibility of identicals, which might be formulated as follows:

- (ii) If x is identical with y then whatever is true of x is also true of y and vice versa.

However (i) makes a slightly weaker claim than (ii). Both (i) and (ii) propose necessary conditions for the identity of x and y but they differ in this respect: (ii) requires that x and y actually have the same properties, whereas (i) only requires that it makes sense to predicate the same properties of x and y whether or not they are truthfully predicable of them. Even more accurately, (i) only requires that it is in keeping with our use of language to be able to interchange 'x' and 'y' independently of what predicate-term is used in a sentence without resulting in a meaningless string of words.

Two instances of applying the CA will help illustrate the use of (i). In one case the argument helps corroborate a presumed identity of the referent of two names, whereas in the other the CA excludes the possibility of identifying the x and y in question. *Case I:* If one were to claim that a flash of lightning is identical with an electric discharge between clouds under specifiable conditions, and if we applied the CA to this particular case, the major premise of the argument would then be:

- (iii) The a -flash-of-lightning is identical with the b -electrical-discharge only if we can meaningfully predicate of the name of the former everything that we can predicate of the name of the latter and vice versa.

Since there is presumably no systematic way of examining all the predicates which are appropriate to the particular flash of lightning in question we can only assume the identification of the two entities is correct until someone proposes a valid counterexample. In other words, the CA is only a negative criterion of identity.

Case II: Now consider the following example:

- (iv) The pain which Jones had is identical with the pain which Murphy had.

If we apply the identical criterion specified by (i) to (iv) we see immediately that the two pains cannot be identical. For we can say

2. The verb 'to predicate' is being used in such a way that properties may be said to be 'predicated' of things, and predicate-terms 'predicated' of subject-terms.

of Jones's pain that Jones felt it, whereas it would make no sense to say of Jones's pain that Murphy felt it. Yet it is quite evident that the two pains can be identical in some other sense. For instance, they could both be pains caused by the extraction of corresponding teeth in Jones and Murphy under exactly similar circumstances. Hence, although nothing depends on the success or otherwise of this application of the CA, it does appear that the sense of 'identical with' which is operating in (i) above is not above suspicion, and this problem will emerge more clearly when the CA is applied to the mind-body problem.

The CA is usually applied to the identity theory in the following way:

- (v) A particular thought (call it t_1) is identical with a particular brain-state (call it b_1) only if what can be predicated of ' t_1 ' can also be meaningfully predicated of ' b_1 ' and vice versa.
- (vi) But there are some predicates which can be applied to ' b_1 ' which cannot, without a category mistake, be applied to ' t_1 '
- (vii) Hence t_1 cannot be identical with b_1 .

It should be clear from this example that the argument is not based on our knowledge of the properties of the entities being compared, but rather on the appropriateness or otherwise of applying predicate-terms to the names of these entities in the construction of meaningful sentences in ordinary English. It is an argument from our language about x and y to the possibility of identifying x and y . If we were relying, at the beginning of the argument, on our knowledge of the properties of x and y rather than on the compatibility of various predicate-terms with their names in meaningful sentences, then we would already know enough about x and y to make such *a priori* arguments against their possible identification redundant. This is especially the case in connection with the mind-body problem where our ignorance of the nature of either term of the proposed identity leaves room for the philosopher to attempt a short circuit of the necessary empirical investigation by means of an *a priori* argument.

The kinds of predicates which are usually mentioned in connection with (vi) are such things as a brain-state being located in a particular part of the body, or any of the properties which might be measured by scientific instruments. The argument relies on the claim that these properties (for instance, having a certain voltage) cannot normally be predicated of thoughts without producing a category mistake in the proposition which is used to make the predication. The conclusion, (vii), is then thought to be unavoidable; it is *a priori* impossible to identify thoughts with brain-states.

It is at this point that we need to distinguish the two sub-types of CA previously mentioned, the CA(S) and the CA(D), and to

show that the argument above is inconclusive because it is an example of a CA(D). To make this distinction a number of assumptions can be made which rely on relatively uncontested points in the philosophy of language and philosophy of science.

In any reasonably sophisticated way of talking about the world, or our experience of the world, there will be words or concepts which refer to entities (call these entity concepts, EC) and words which refer to qualities of entities (call these quality concepts, QC). In fact, even if such is not the case in general, it is presupposed by the structure of the CA and therefore we need only consider languages which do allow for such a distinction. Secondly, without evoking the notorious problems involved in specifying what exactly is meant by a conceptual framework, we can presume that it roughly includes the EC, QC (among others) and various rules (such as syntactic and semantical rules etc.) for the meaningful combination of these words into sentences. The details of what is involved in a conceptual framework as such are not relevant to the present discussion. Furthermore, it is presumably beyond dispute that we can have alternative conceptual frameworks to match the same observational data, and this without presupposing anything *à la* C I Lewis about the nature of what is immediately 'given' in experience. Of the many alternative frameworks which are possible, we may broadly classify them under two headings which have been proposed by Wilfrid Sellars.³ These are the framework of our ordinary language and the framework of scientific explanation, called respectively the Manifest and the Scientific frameworks (abbreviated to MF and SF). It follows that there will be EC and QC in both the MF and the SF, and the rules of each framework will independently control the use of both types of words in each framework. Even at this stage, it is highly unlikely, *a priori*, that we could combine a QC from one framework with an EC from the other, or vice versa, without the possibility of conceptual confusion. We should expect this confusion because the rules of each framework are only designed to function within the framework itself and cannot normally cope with concepts which are alien to that particular framework.

Using this very rough sketch of the role of conceptual frameworks in generating category mistakes, we can now distinguish between the CA(S) and the CA(D) by introducing a rule for the correct application of the CA to identify discussions. The rule is as follows:

- (viii) The CA is applicable to a discussion of the possible identity of *x* and *y* only if 'x' and 'y' are both in the same conceptual framework.

3. See W Sellars, *Science, Perception and Reality* (London: Routledge & Kegan Paul, 1963) 1-40.

Only those instances of the CA which follow this rule will be classified as a CA(S), or category argument within the *same* framework, whereas those which break rule (viii) will be called a CA(D), or category argument in *different* conceptual frameworks.

If we now look back to (i) we can see that, as it stands, it leaves itself open to being used in a CA(D). One could, of course, leave it unchanged and maintain it as a defining statement of the notion of identity which is appropriate to the identity theory. But this would not in any way represent what its proponents wish to defend as regards the hypothesized identity of mind and body. For such a concept of identity, as specified by (i), would effectively undercut most of our accepted scientific explanations, which rely on correspondence rules⁴ to link theoretical language to observational language about a particular *explanandum*. In almost no case of a successful scientific theory can we easily interchange the predicates which are appropriate to the phenomena of experience with those which are appropriate to the theoretical entities which are postulated to explain them. Sub-microscopical particles do not have the qualities which tables and chairs have. Hence, we need to revise (i) so that it leaves open the possibility of identifying an entity referred to in the MF in one way with the same entity referred to in the SF in another way, such that the QC or the EC of one framework cannot be replaced by those of the other. In other words, to capture the sense of the original objection to the identity theory, it is necessary to amend (i) to (i'), thus:

- (i') If 'x' and 'y' are both in the same conceptual framework, then x is identical with y only if we can meaningfully predicate of 'x' what can be predicated of 'y' and vice versa.

Even this is not precise enough yet. There are various ECs which can function in different conceptual frameworks and which can, as a result, have different types of QC predicated of them, according to whether they belong to one or other framework. This opens up the possibility that we might have an example of an x and a y where, for example, the 'y' could function in either of two frameworks. framework I and framework II⁵. If 'x' belongs to framework I, then both 'x' and 'y' belong to the same framework (namely, framework I), and perhaps even have all and only the same predicates attributable to them insofar as they belong to this same framework (call these predicates 'p₁', . . . 'p_f'). In this case, if we constructed a CA in an attempt to disprove the identity of x and y , it would

4. Cf. Ernest Nagel, *The Structure of Science* (London, Routledge & Kegan Paul, 1961) 90-105; W Sellars, *op. cit.* 106-126.

5. In that case, 'y' would be ambiguous unless the framework in which it functions were specified. For example, 'force' has a different meaning in Newtonian or Einsteinian physics.

apparently be a CA(S) since it would comply with the requirement of rule (viii). Yet, because 'y' also belongs to framework II, it might have another series of predicates which are attributable to it (call these 'q₁', . . . 'q_n') in this framework, but which cannot be meaningfully predicated of 'x'. Therefore, even (i') is not sufficient to exclude the possibility of such ambiguous predication, for it would allow the construction of a CA(S) in the example given above in which the minor premise, by analogy with (vi), would mention one of the series 'q₁' . . . 'q_n'. To avoid this ambiguity, (i') needs to be further revised to the following form:

- (i'') If 'x' and 'y' are both in the same conceptual framework, then x is identical with y only if we can meaningfully predicate of 'x' whatever can be predicated of 'y' insofar as it belongs to this shared framework, and vice versa.

With (i'') as the skeleton of the major premise of a CA(S), it is now much less likely that we can construct an effective argument against the identity theory. The example given above, in (v)—(vii), traded on the ambiguity of the MF and the SF. Talk about 'thoughts' belongs to the MF, whereas talk about brain-states and their voltages etc. belongs to the SF. Hence it was relatively easy to find a QC from the SF, to substitute in (vi) above, which would result in a category mistake when combined with the concept of a thought from the MF. But as soon as the amended (i'') is introduced into the CA, so that CA(D)s are excluded and also whatever can be said of brain-states insofar as this concept is a scientific one belonging to the SF, then it becomes much more difficult to find a QC which can be predicated of 't₁' but cannot be meaningfully predicated of 'b₁' also, or vice versa. A category mistake which results from a CA straddling alternative conceptual frameworks, i.e. a CA(D), or a CA which, appears conclusive because it uses QCs which belong to different conceptual frameworks, is not an effective argument against the possibility of identifying the referents of the ECs from the two frameworks. To construct a valid CA we would need to fulfil the following two conditions:

- (a) both x and y must belong to the same conceptual framework, and
- (b) only QCs which are predicable of 'x' and 'y' insofar as they belong to this one, shared framework are candidates for consideration in the minor premise of the argument.

Although the CA(D) group includes most of the CAs which are used against the identity theory, there are others which might be more appropriately included under the CA(S) rubric. These would be CAs which accepted rule (viii) and (i'') as the revised form of the major premise, and then found an example of a predicate which could not be meaningfully applied to the concepts of a

thought and of a brain-state, both in the MF. However, this type of argument fails for a different reason. We cannot legitimately argue from the constrictions of our conceptual framework to *a priori* factual conclusions. Our language, or our conceptual framework, reflects the way we think the world is. We could hardly argue from our conception of the world to empirical claims about the world itself in such a way that these empirical claims were no longer open to experimental confirmation or disconfirmation. As a result, J J C Smart has consistently called for a revision of refractory conceptual frameworks or languages to comply with developments in our knowledge rather than the retention of languages which foreclose the possibility of even stating new scientific or philosophical hypotheses. In the present instance, we would need to introduce a semantical rule which would permit, within the MF, the combination of what otherwise appears as a category mistake of applying the wrong type of predicate to an EC.

The CA, then, is seen to be an ineffective argument against the possibility of the identity being proposed and seriously examined as a theory of mind, for one of the following reasons: it either confuses the conceptual framework to which the ECs or the QCs of the argument belong, or it argues invalidly from the constrictions of our present language about the world to empirical claims which are not open to empirical testing.

The Meaninglessness argument

In a reply to a series of articles by J J C Smart, Norman Malcolm has recently proposed a new line of argument against the identity theory. His objection is the following.

Malcolm admits the distinction between a contingent identity of *x* and *y* and an identity of meaning between 'x' and 'y'. The identity theory is only claiming the former, i.e. that whereas 'thought' and 'brain-state' are not identical in meaning, they do, as a matter of fact, both refer to exactly the same entity (or type of entity). If the identity of thoughts and brain-states is to be a contingent one, Malcolm argues, then there must be some empirical evidence in its favour, and there must be logically distinct methods of verifying the occurrence and properties of both thoughts and the corresponding brain-states. Otherwise, the identity in question would no longer be contingent, but analytic.

It should be noted that since the identity is presumed to be *contingent*, the method of determining that a *thought* or *thinking* was located inside the skull would have to be *logically independent* of the method of determining the occurrence of the correlated brain process. Employing the occurrence of the brain process as the *criterion* of the occurrence of the thought inside the skull would prevent the putative identity from having an empirical verification. It would no longer be

conceived of as a contingent identity. There would be a connection of *meaning* between saying that a certain brain process occurred inside someone's skull and saying that a certain thought occurred in the same place.⁶

However there is a problem in examining the occurrence and properties of a thought, since the very notion of a thought having a certain location makes no sense. To make his point, Malcolm is willing to concede what was earlier argued concerning the use of the CA(S), and to grant that a thought might be said to have a location, by means of a change in our language about thoughts. But the result of this new convention would be to change the identity of thoughts and brain-states from an empirical one to an identity of meaning.

Since the notion of a bodily location of thinking does not have any meaning at present, it *can* have meaning only if it is *given* meaning. One way this could come about would be through adopting the convention that the occurrence of a certain physiological process in a certain place (for example, the brain) would signify the occurrence of thinking in that place. This would be a definition of what the latter meant. But the identity of thinking and physiological process, which would emerge from this convention, would not be a contingent identity. It would have been 'verified', in part, by legislating sense into an expression that presently does not have sense.⁷

According to Malcolm, then, Smart is faced with this dilemma: the identity proposed by the identity theory is either analytic or meaningless. If we do not legislate a sense, by a new convention, for the notion of a thought having a certain location, it will be impossible to empirically examine thoughts, and consequently the identity theory will be nothing more than an empirical hypothesis for which it is impossible, in principle, to find any evidence. In other words, the hypothesis would be meaningless. If, on the other hand, we change our language about thoughts so that it makes sense to talk about a thought having a location, the connection between thoughts and brain-states would then be a meaning relation and the hypothesis would still not be an empirical one. In either case, the identity theory is not an empirically testable hypothesis.

The crucial step in Malcolm's line of argument is the one from his acknowledgment that the identity theory implies a contingent identity to his claim that we need to find empirical evidence in favour of the hypothesis, and consequently that we need to have logically distinct ways of examining the *x* and *y* which are to be identified. This transition seems harmless in the case where we have observa-

6. Norman Malcolm, *Problems of Mind: Descartes to Wittgenstein* (New York, Harper & Row, 1970) 67.

7. *Ibid* 69.

tional evidence for the existence of x and y in such a way that we can examine them both to determine whether we are dealing, in fact, with two distinct entities or merely with two names for the one entity. But it is a mistake to extrapolate this type of corroboration of an empirical identity into a general principle which is accepted, without question, as a necessary condition for a putative identity being empirical. For the identity theory is not claiming that we can find empirical evidence in favor of an identification of thoughts and brain-states by a distinct examination of each, but rather that we have no reason to postulate the existence of thoughts (as entities distinct from brain-states) to explain human behavior and experience. In other words, in order to make any proposed identification of an x and a y empirical, there are many ways open to the theorist, only one of which is the independent examination of x and y for empirical confirmation of the theory. Another method might be to appeal to Occam's razor, and to question the justifiability of x (or y) on empirical evidence as a postulated entity of an alternative theory. A simpler example of this approach might help clear the identity theory of the charge of meaninglessness.

Imagine a physiologist, prior to William Harvey, explaining the movement of the blood in the arteries by means of an intangible, ethereal spirit (a reconstructed Galen would do!). When Harvey discovers a new explanation, namely the muscular action of the heart, he might formulate his solution of the 'heart-spirit problem' as an identity theory as follows: the blood-spirit is identical with the pumping of the heart. Is his identity claim an analytic one (i.e. a claim that 'pumping of the heart' means the same as 'blood-spirit')? Certainly not. Is his claim meaningless? Hardly. What Harvey's identity theory amounts to is that we have no reason to postulate the existence of anything besides the pumping of the heart to explain the movement of the blood in the arteries.⁸ Likewise the identity theory of mind does not begin by admitting the existence of two types of entity, namely thoughts and brain-states, whose possible identity needs to be empirically examined. If it did, it would end up in Malcolm's dilemma. But the theory begins by challenging the reasons for the postulation of anything more than brain-states etc. for the adequate explanation of human experience and behavior. It denies that the word 'thought' has any reference which is really distinct from the reference of the word 'brain-state'. As a result, it finds no reason to look for empirical evidence in favor of its claims (or, more precisely, to empirically examine both the x and

8. The 'heart-spirit problem' is not meant to be historically accurate. W Harvey, even in his *De Motu Cordis*, continues to refer to 'spirits' in the blood, and to give a function to the innate heat of the heart (or elsewhere, of the blood) in explaining the circulation of the blood.

y of the identity theory by logically distinct methods), since the non-empirical member of the two entities to be identified is being denied a real reference which is distinct from the reference of the empirical member. Malcolm's dilemma resulted, fundamentally, from granting real, empirical reference to both 'x' and 'y' in a theory which is designed to deny exactly that. It is not surprising if a self-contradictory theory gives rise to methodological dilemmas.

The solution to Malcolm's dilemma for the identity theorist is to interpret his theory in the following way: thoughts are postulated, non-empirical entities which are invoked to explain certain features of human experience and behavior. Likewise, the processes examined by the neurophysiologist are empirical, postulated states of the body which are proposed (by the identity theorist) to explain the same features of human experience and behavior as the thoughts of the traditional philosopher. The identity theory views these two positions as alternative theories (in a very general and loose sense of the term 'theory') which purport to explain the same phenomena. Since the traditional philosopher has need of both types of postulated entity to explain the phenomena in question, the identity theorist appeals to Occam's razor, and he expresses his claim that thoughts are redundant explanatory principles by the identification of thoughts with his own neurophysiological, postulated entities and processes. While his claim is an empirical one, he obviously cannot be expected to support his claim by an empirical examination of thoughts. His theory suggests that they do not exist as entities which are distinct from brain-states. The empirical warrant for the identity theory can only come from its success as a scientific theory.

Conclusion

Nothing in the discussion of the CA and the MA would warrant the claim that the identity theory is true or justified. The rejection of the CA and the MA as convincing arguments against the identity theory does nothing more than justify the conclusion that the theory in question is not shown to be *a priori* invalid by means of these two arguments. Much work remains to be done on the theory itself. In fact, it is so little developed that it hardly merits the name of a theory, and amounts to nothing more than a project for further research in philosophy and especially in the relevant sciences. The principal difficulties, from the philosopher's point of view, are conceptual ones in the effort to bridge the gap between the language of immediate experience and the language of postulated entities of scientific theory by adequate correspondence rules. Hasty predictions are inappropriate in this area. Israel Scheffler offers more sober advice to the potential critic of the identity theory: '. . . neither identity nor non-identity, *in the sense of* definitional *reducibility* or *non-reducibility*, is something we can, or ought to try

to, decide for vague masses of future scientific content. The problem is, for the philosopher as for the fat man, to try to reduce now'.⁹

University of Notre Dame, Indiana

DESMOND M CLARKE

9. I Scheffler, 'Comments' on 'The Identity Approach to the Mind-Body Problem' by W Sellars, in *Boston Studies in the Philosophy of Science*, vol. II, ed. Robert S Cohen and Marx W Wartofsky (New York: Humanities Press, 1965) 80.