

OXFORD

Philosophy of Science Today

Edited by Peter Clark and Katherine Hawley

Philosophy of Mind

John Campbell

- 1 *Mental causation*
 - 2 *Functionalism*
 - 3 *Types of content*
 - 4 *Other minds*
 - 5 *Personal identity*
-

The last fifty years has seen intense and broad development in philosophy of mind, unprecedented in the whole history of the subject. In this essay I will try to convey something of the broad sweep of developments, but I may as well say now that (a) it would have been possible, even for this author, let alone someone with a quite different approach to the subject, to write a review of the period that focused mostly on material I have either not mentioned or mentioned only in passing, and (b) it is not possible within the scope of a single essay to convey the depth with which every single topic in the area has been discussed. It would, for example, be quite possible to write a long and interesting book that simply reviewed all that has been said about the inverted spectrum during the last fifty years. I have kept references to a minimum.

1 Mental causation

Wittgenstein [1953] and Ryle [1949] thought that psychological and causal explanation were of different types, that it was a grammatical mistake to suppose that psychological explanation was causal. An explanation that shows the rationality of an action or psychological state contrasts with causal explanations, which merely aim to show that what happened was to be expected. In one of the simplest and most influential arguments advanced on the subject, Davidson [1963] pointed out that we need a distinction between (a) the case in which someone has a reason to perform an action, and performs the action, but not for that reason, and (b) the case in which someone has a reason to perform the action, and performs the action for that reason. How are we to explain the distinction between these two cases? It seems that the only way of explaining it is to say that in case (a) the cause of the action is not that particular reason; whereas in case (b) the cause of the action is the reason. It was not

easy to see how Davidson's contrast could be drawn on a non-causal model of psychological explanation.

Place [1956], Smart [1959], and Feigl [1959] had already held that psychological states could be identified with brain states; to be in a particular type of psychological state was to be in a particular type of brain state. Psychological states were thus, on this account, firmly identified as causes of what they explained. But the problem, stated by Putnam [1967], was that it seemed possible that the very same psychological states could be possessed by creatures with quite different physical constitutions. Davidson's [1970] anomalous monism shifted the locus of the identity claim. In effect, Davidson makes a sharp distinction between events and their properties. And he has a monistic account of events and a dualistic view of properties: all events are physical events, but there are two types of property that events can have: physical and psychological. Events stand in causal relations to one another, in virtue of the lawlike relations holding between their physical properties. So a mental event—an event with mental properties—can cause another event, in virtue of the fact that the event also has physical properties, which stand in lawlike relations to the properties of the event caused.

This, though, leaves the causal role of the psychological properties uncharacterized; it leaves it seeming that, if the psychological properties had been varied while the physical properties were held constant, the effect of the event would have been the very same. The key problem here is that Davidson's initial argument, which is what convinced everyone that psychological explanation must be causal, does not establish that psychological properties are causally relevant; the idea of performing an action for a reason could already be explained by an anomalous monism which regarded psychological properties as epiphenomenal. Even such a view could distinguish between the case in which you perform the action and your performance is caused by your having the reason, and the case in which you have the reason and the event of your having that reason is not causally active. Although we naturally feel that psychological properties must be causally relevant, we lack a compelling argument on this point, which means that we have no grasp of why they should be taken to be causally relevant, and when it would be right to regard them as indeed causally idle.

The obvious response at this point is to appeal to a doctrine of supervenience: that, if the physical properties are held constant, the mental properties are thereby held constant. This suggests that we might appeal to counterfactuals here: if the event had not had the mental properties, it would not have had the physical properties in virtue of which it has its effects, and if the event had still had its mental properties, it would still, one way or another, have had the physical properties in virtue of which it had its upshot (Jackson and Pettit [1990]).

The appeal to supervenience here needs further discussion, but the problem of explaining how psychological properties can be causally relevant is still only beginning. Suppose we consider the onset of delusions, such as those suffered by some schizophrenic patients, which were beginning to draw philosophical attention as the millennium ended. Jaspers [1959] famously said that delusions were 'un-understandable', meaning that the successive stages of the delusion resisted psychological explanation. In particular, you could not give a causal explanation of later stages of the delusion by appealing to earlier stages of the delusion. Rather, the progress of the delusion had to be understood as epiphenomenal upon the progression of an underlying physical dysfunction—a 'virus', yet to be discovered, that produced the psychological symptoms as epiphenomena. It is a mark of how little grasp we have of the causal relevance of psychological properties that we seem to have no way, even in principle, of saying what the difference would be between a case in which Jaspers's analysis was correct, and a case in which the later stages of the delusion really were causally explained by the psychological states of the subject earlier in the delusion. In both cases there will be counterfactuals to the effect that, if the earlier psychological properties had not been instantiated, neither would the later psychological properties. At the moment we really seem to have no way of explaining how to draw the finer distinction we want, between the case in which there is psychological causation and the case in which there is not psychological causation.

It is not, though, as if a return to Ryle and Wittgenstein would help us here: they do not offer an alternative way of drawing the distinction we want, between a progression of psychological states that is epiphenomenal upon an underlying physical progression, and a progression of psychological states in which the earlier states explain the later states. In particular, an appeal to the notion of rationality will not help. While the deluded subject may not be our ideal of rationality, the notion of rationality simply will not take the weight here; we do not have a sharp way of determining whether the subject is sufficiently irrational that the delusion falls outside the scope of psychological explanation.

2 Functionalism

As Kim [1998] points out, when we are looking for explanations of how psychological states might have causal significance, given that all causation is physical, what we need are not bridge laws stating that we have a type of mental state when and only when we have a particular type of physical state, but functional characterizations of the psychological properties, so that we can see how they are realized by underlying physical structures. As is also evident, though, it is not easy to see how many aspects of phenomenal consciousness

could be given functional characterizations. Block and Fodor [1972] first set this line of thought in motion by pointing out that on the face of it, sameness of functional architecture would be consistent with two subjects having spectrums, one of which is an inversion of the other.

It is also a good question whether representational content could be given a functionalist characterization. One fundamental shift through the last fifty years has been in the attempt to provide direct analyses of what it is to be in states with propositional content, such as beliefs and desires, without appealing to the idea of a shared public language. Various types of functionalist account were proposed, following Putnam's early work. The basic idea is simple: if we agree that beliefs, for example, have causes and effects, then perhaps, for each particular replacement of 'p', we can define what it is for someone to believe that p by saying that it is a matter of being in a state with a particular characteristic pattern of causes and effects. A flood of illuminating work followed, canvassing various forms such definitions might take and the status that should be accorded the definition—is the definition merely, for example, an articulation of the ordinary notion of belief, or is it grounded in scientific discovery?

One fundamental argument for functionalism, still forcible today, was its consonance with a broadly materialist view of the world. But functionalism also seemed to fit with the emphasis that Wittgenstein [1953] and Quine [1960] placed on the notion of meaning as use. If the meaning of a proposition really was its place in a pattern of use, why should we be looking at use only in linguistic contexts? Surely we could find and analyse propositional states by looking only at the beliefs and desires of a single individual, whether or not they were given linguistic expression. But functionalists did not, in general, address the puzzle that Wittgenstein and Quine had identified: if we really are looking only at complex patterns of causes and effects here, why should we describe them as states capable of truth or falsity at all?

The issue was given a jolt by Kripke's [1972] idea that reference is fixed by a causal chain, made vivid by Putnam's [1975] generalization of Strawson's [1959] thought-experiment involving massive reduplication of our sector of the universe somewhere else. On a 'Twin Earth', your and my duplicates would be functionally identical to us, but their beliefs would be about different objects and stuffs. Consequently, the content of our beliefs cannot be exhausted by their functional roles. Functionalism has to be supplemented by some appeal to causal chains to specify just which things or stuffs were being thought about. If a causal chain between the name and the object fixes the reference of the name, does that not imply that we have a way of setting standards of correctness for the use of the term? Of course, the causal chain may of itself be a causal factor underlying the use that is made of the term. But the mere fact that it is a causal factor does not of itself mean that it has a role

to play in setting standards of rightness and wrongness. Why causal chains should play such a role was not explained in Kripke's account. The fundamental issue of just how it comes about that we have here not merely a pattern of causes and effects, but uses of language or thought which can be right or wrong, is simply not addressed. A bold attempt to deal with this problem, by embedding Kripke's causal picture in a view of signs as having a set of functions for which evolution selected their use, was developed by a number of writers, in particular Millikan [1984]. But the fundamental question, of whether this provides an explanation of the basis on which propositions can be judged right or wrong remains; it is not obvious that an account of evolution does have to provide standards of right and wrong, rather than a non-normative causal story.

These points bear on the striking shift between 1950 and 2000 in the balance of power between philosophy of mind and philosophy of language. From 1950 until about 1980, it seemed unquestionable that the philosophy of language was the foundation of the whole subject of analytical philosophy. Since 1980, it has come to seem that the philosophy of language will have to depend to a considerable extent on a philosophy of mind. The issue is the relative priorities of thought and language in analysing content. There is a common-sense picture on which we use a word the way we do because we know what it stands for. And what the term stands for sets standards of correctness for the use that we make of the term. The later Wittgenstein and Quine thought that this picture could not be sustained: there is only the use; there is no such thing as a knowledge of reference that causes you to use the term in the way you do, and sets standards for the correctness of use. The later Wittgenstein therefore faced the problem that this gives the impression that the use that is made of a term is arbitrary: if there is no knowledge of reference to which use is responsible, how can there be standards of rightness and wrongness for the way in which we use the term? And, despite the difficulty in seeing how he aimed to resolve this point, it does seem that he means to be appealing to our grasp of a shared language in doing so. Quine, in contrast, seems to have thought that ultimately there are no standards of right and wrong to which the use is responsible: there is only the use, and standards of right and wrong are read in by the theorist. The issue of the relative priorities of thought and language is by no means resolved, and whether philosophy of mind ultimately depends on philosophy of language is an open question. It would still be open to us to go back to the common-sense picture that Wittgenstein and Quine both rejected, on which we use words the way we do because we know what they stand for. Russell [1910–11] seems to have been the only major twentieth-century philosopher to try to develop this view, the key here being his notion of 'acquaintance' as what provides us with knowledge of the reference of a term. On this view, it is experience of the object that provides you with your knowledge of the reference of

the term, and it is consequently this experience of the object that determines what it is for you to be going right or wrong in your use of the term.

3 Types of content

Through the 1950s and 1960s, experimental psychology invested increasingly heavily in explanations that appealed to the idea of human information-processing. The idea was that human performance on various tasks—whether the ability to produce and recognize grammatically well formed sentences or to keep track of one voice in a sea of conversation—could be explained in terms of the operation of information-processing systems which carried out computations on input representations. In effect, this work challenged the unity of the philosophers' notion of a representational state. Philosophers had been accustomed to think of beliefs and desires as representational, and to suppose that it would be possible to explain all other representational notions in terms of their relations to beliefs and desires. However, the representational states postulated by cognitive scientists were not themselves beliefs and desires, but nor did it seem possible to explain their contents in terms of their relations to beliefs and desires. One picture of what was happening here was set out by Chomsky [1976]. According to Chomsky, we do have a unitary notion of 'representational state' which covers both ordinary beliefs and the states involved in, for example, analysing the grammatical structure of a heard sentence. The difference is only that some of these states—in particular, ordinary beliefs and desires—are accessible to consciousness, while others, such as those postulated by linguists, are not. But according to Chomsky, the distinction between representational states accessible to consciousness and those not accessible to consciousness is of no particular theoretical interest.

An alternative reaction was developed by Searle [1995]. According to Searle, there is only one notion of representation, and it is constitutively linked to consciousness; there are only the ordinary beliefs and desires, of which we can in principle become aware, and it does not really make sense to talk in terms of other kinds of representational state. We can have intentionality only on the part of states that are in principle accessible to consciousness.

The cautious view, though, is that we have (at least) two different types of representation or intentionality. On the one hand, there are the conceptual contents of ordinary beliefs and desires, to which consciousness may constitutively attach. On the other hand, there are the non-conceptual contents of information-processing states. While in principle it may be possible to reduce one to the other, the reduction would be an extremely substantial piece of work (Evans [1982]; Davies [1995]).

Fodor [1983] maintained that 'central processing'—the home of ordinary beliefs and desires—would elude study by cognitive science, which had no

methodology for studying anything other than information-processing modules. The processing carried out by a modular system has an independently identified point: in advance of knowing the details of just how the computation is carried out, or how the computation is implemented, we know the target of the computation, what it is that has to be figured out. This means that we have a way of setting standards of right and wrong for the processing carried out by the system. However, ordinary thinking—presumably carried out in ‘central processing’—is general-purpose. It has no externally identifiable point. It is, therefore, problematic to explain how we can speak of there being standards of right and wrong for central processing: the Wittgenstein/Quine problem is recast at this point. Moreover, the very idea of a single type of content, common to both information-processing modules and central processing, seems to be undermined, though this was not obviously Fodor’s intention. Since the two types of system have such different architectures, and since the standards of right and wrong are being defined so differently in the two types of system, it is hard to sustain the idea that we have the same style of representation both times. Indeed, from this perspective it is even arguable that information-processing systems can have content only because they are providing input to some central processing system, which defines the objectives of the modular inputs.

I said earlier that phenomenal consciousness seemed to provide an obstacle to functionalist explanations of how mental states could have causal efficacy (namely, by being realized by suitable micro-physical states). It is, however, possible to argue that all aspects of phenomenal consciousness are representational; and it might then be argued that we can give an analysis of representation on which we can understand how phenomenal consciousness can fit into a broadly physicalist picture of the causal. As Tye [1995] puts it, the thesis is that ‘all phenomenal character is phenomenal content’, and then we might look for a naturalistic account of representational content (Dretske [1995]). Peacocke [1983] in effect provided the basis of a library of putative counterexamples to representationalism, which was subsequently added to by Block [1996] and others. If, though, we accept the cautious picture on which there are two types of content here, what are the implications for an analysis of consciousness in terms of content? There is, on the one hand, the richer notion, where we are talking about the kinds of content possessed by ordinary beliefs and desires; possession of states with those types of content may be sufficient for consciousness, but it does not seem likely to offer much in the way of a naturalist account, just because consciousness is arguably presupposed by this type of content. On the thinner notion, on which we are dealing with the same type of content as is possessed by sub-personal information-processing states, consciousness does not seem to be directly presupposed; but for that very reason, an explanation of consciousness in terms of this variety of content is

unlikely to carry conviction: it is likely to seem obvious to the uncommitted reader that you could have creatures with quite complicated states with this type of content which are not themselves conscious at all. Chalmers [1996] gives eloquent expression to this kind of scepticism.

There is a subsidiary problem here anyhow, which is that it remains an open question whether the two types of content are independent: as I said, it could be argued that sub-personal information-processing states count as contentful only because they are functionally connected, whether as input or output, to the conscious states of the creature. In the case of human subjects, one of the fundamental reasons why sub-personal representational states are postulated is to explain the judgements and verbal reports of subjects.

The whole debate over the analysis of phenomenal consciousness suffers from the fact that the objective of an analysis of consciousness has never been specified in a way that commands general assent. In particular, there has been little discussion of why we need the notion of conscious states with phenomenal character at all; what the point is of the notion. So it is actually quite difficult to know what it would take for us to have a satisfactory analysis. Proponents of one type of naturalist analysis or another will typically claim to have included in the analysis everything that is required for there to be 'something it is like' to be in a mental state—such as, for example, the capacity for verbal report of the state, or to act on the basis of the state—whereas opponents of the analysis find it obvious that it leaves out the critical element of experience itself. This deadlock arises, I think, just because the point of the notion of consciousness has not been adequately specified, so that it is not apparent what conditions would have to be met for the analysis to be satisfactory.

It is striking that Searle, for example, a vigorous proponent of the idea that consciousness is needed for there to be any kind of representational state, seems in effect to suppose that the role of consciousness is to make it apparent to you that you are in that very state. Once this is supposed, all the naturalist has to do is to find representational counterparts of this phenomenon, making it apparent to you that you are in a particular state, and this is not all that hard to do. But at this point it appears that there is after all more to the point of the notion of consciousness than that. For one feels that the naturalist account, which simply appeals to, for example, the ability to represent one's representational states, has missed something out. But the point of the missing element has not been made explicit. So we reach a deadlock between the exasperation of the reductionist and the frustration of the Searlean.

The only way through this deadlock is to rethink just why we need the notion of consciousness. In particular, we could rethink why the contents of beliefs and desires depend on consciousness. I earlier mentioned Russell as having held that it is conscious acquaintance with the objects of our thought

that provides us with knowledge of what we are talking about. On this view, it is consciousness of the world that provides us with knowledge of the references of our terms. This gives a significant explanatory role to the notion of consciousness: knowledge of reference is itself a notion that bears a considerable theoretical role. So an analysis of consciousness in terms of naturalized content would face the challenge of explaining what it is to know the references of the terms in which we ordinarily express our beliefs and desires. Whether this challenge can be met remains an open question.

4 Other minds

The problem of other minds is foreshadowed in Augustine, but it is only in the twentieth century that we find it treated as a major philosophical question. (For an excellent review of the history of the problem, see Avramides [2001].) What ignited the debate last century was the intense development of logical positivism in the 1920s and 1930s. But by 1950 there was a quite different approach. The positivists had at various points been inclined to behaviourist analyses of ascriptions of mental states to other people, so that saying of someone else that he was in pain was a matter of saying that he was disposed to behave in various ways. This meant that pain was not to be thought of as the hypothesized cause of his dispositions to behaviour, so there was not a hard problem about how we know when someone else is in pain. In Ryle [1949], and Wittgenstein [1953], there is on the one hand no sympathy for the kind of austere behaviourist reduction of mental terms provided by the positivists, but the anti-causalism of the positivists is retained: in ascribing mental states we are not talking about postulated causes of behaviour, we are talking in a way that in some sense recapitulates or redescribes the observed behaviour itself, so that there is no hard problem of how we know on the basis of behaviour what the mental states are. In Wittgenstein in particular, the view of mental states was embedded in a challenge to the idea that you could know what it means to ascribe mental states to someone else, but be at a loss when it comes to telling whether any of those ascriptions are correct. On Wittgenstein's picture, knowing the meanings of mental-state terms is a matter of knowing how to use them correctly, so there can be no possibility of understanding the term but not knowing how to use it. These approaches aimed to respect the univocity of psychological predicates as they appear in first- and third-person contexts; the idea was that the self-ascription of a psychological state may itself be part of the pattern of behaviour on the basis of which others ascribe the state. Strawson [1959] gave one of the defining statements of the new approach, in which he maintained both that, to understand psychological predicates at all, one had to grasp both first- and third-person uses, and that understanding the predicate involved knowing how to ascribe it. Consequently, there was no

possibility of, for example, knowing how to ascribe a type of psychological state to oneself while not yet understanding how to ascribe the state to other people.

In the background here was the idea, a legacy of the positivists, that understanding the meaning of a term was a matter of knowing how to use it. The opposed idea—that you use a term as you do because you know what it stands for—fell by the wayside; the challenge, which seemed unanswerable, was to explain what it would be to know what a psychological predicate stands for, if this is not simply a matter of being able to use the predicate appropriately, but rather the reason why the predicate is used as it is. But, while the doctrine of meaning as use became the orthodoxy through the 1960s, the non-causal view of the relation between psychological states and behaviour came under heavy attack. In an influential series of papers Putnam pointed out, for example, that the relation between a mental state—such as the belief that it is raining—and behaviour—such as whether one picks up an umbrella—depends on which other mental states one has—such as whether one wants to get wet. We seem to be dealing with a complex of interacting causal factors underlying behaviour, rather than anything that could be constituted by a tendency to behaviour. By the end of the 1960s, the non-causal view of psychological states had all but disappeared. This emphasis on the causal role of mental states led to the conception of knowledge of the mind of another as a kind of theory-construction, in which the observable data—various pieces of behaviour by the individual—are explained by a network of hypotheses about the underlying interacting mental causes of the behaviour.

There were, however, a number of problems with this idea. First, it was not obvious what the content was of these explanatory hypotheses. Were they exhausted by their functional content, so that they were merely hypotheses about the structure of the network of causes underpinning someone's behaviour? Or was there some knowledge of what it was to be in a particular mental state, available from one's own first-person perspective, that was being used in stating these explanatory hypotheses? Secondly, it was not obvious how the view could explain knowledge of one's own mental states. Since knowledge of one's own headache, or pleasure, does not seem in general to be knowledge of a hypothesized cause of one's own behaviour, it was not obvious how it could be the very same state that one had knowledge of in this way that one was postulating as a cause of someone else's behaviour. Finally, there was a reaction in which it was held that the mental states of another can be observed as directly as any bare behaviour. For example, if you give up trying to do a sum in your head and perform the calculation on paper instead, the making of marks by your pen constitutes your thinking, and is observable by me. It is not as if the real thinking is going on initially back-stage, so you are now performing a more onerous task, the initial mental arithmetic plus an external writing

behaviour that is caused by your purely mental calculation. So I do not have to infer from your writing on the paper to a hypothesized cause to find out what you are thinking.

The conception of knowledge of other minds as theory-construction is none the less still influential. As well as the problem of defining the data that the theory is intended to explain, there is the question how we have knowledge of our own minds. One radical line is that this too is a form of theory-construction, a line followed recently by Gopnik [1993]. On this view, we are as theorists specialists in explaining the behaviour of individuals we know well. So, for example, even the inflection of a cough on the part of someone we know well can speak volumes about the underlying mental state, just as a change in the timbre of the engine of a fishing-boat can speak volumes to the ship's engineer. But in the case of yourself, you are yourself the leading authority; so the theory construction in this case is particularly quick and easy, and it is a quickness and easiness that we mistake for having some kind of privileged access to the contents of our own minds.

In the 1990s, the view of knowledge of other minds as a matter of theory-construction was put to work as an empirical hypothesis about the way in which young children gain knowledge of other minds. There was a set of striking findings which seemed to show that children at 3 years old lack the concept of false belief. Though they can talk in terms of the beliefs of others and themselves, they will not ascribe to their own past selves, or to other children, beliefs that they themselves currently think to be mistaken. The analysis proposed by 'theory-theorists' was that at 3 years of age children are operating with a theory in which 'belief' is a primitive relation between an individual and an external state of affairs; so it makes sense to suppose that the relation fails to hold between a particular individual and a particular external state of affairs, but it makes no sense to suppose the relation could hold between an individual and a non-existent state of affairs. The mature conception of belief as a representational state, on which it does make sense to speak of false belief, does not emerge until the child is about 4 years old, and it emerges as the result of repeated testing of the old theory and the formation of new theoretical constructs, such as the conception of belief as representational, to deal with the problems encountered. The fundamental difficulty for this 'theory theory' as a hypothesis in developmental psychology is the lack of any convincing 'logic of discovery' to explain how the child gets from one theory to the next, and why it is that children are so strikingly similar to one another in the developmental progressions they make.

A great deal of excitement was generated in philosophy by the initial bold statements of the theory theory, and of a rival empirical hypothesis: simulation theory (Davies and Stone [1995]). According to simulation theory, the stages in the development of the child's grasp of other minds are stages in the

capacity to simulate other people's mental states: to take on, for example, the beliefs and desires of another person in way that is 'off-line', or in the scope of an imaginative exercise that does not immediately issue in action or the formation of enduring beliefs. Some of the excitement has now gone out of this dispute, partly because the empirical findings since the initial data have become increasingly complicated and difficult to interpret uniformly, partly because it has become difficult to see that the empirical results actually will yield a decision in favour of one or the other theory. Moreover, it is questionable to what extent this dispute really does illuminate the classical problem of other minds. The theory theory and the simulation theory are first and foremost theories about our knowledge of the causal roles of another person's mental states. In so far as there is more to our knowledge of another person's mental states than our knowledge of their causal roles, they shed no light. But, arguably, knowledge of what someone else's mental states are like, in a phrase made famous by Nagel, is more fundamental than knowledge of their causal roles.

5 Personal identity

Bernard Williams [1956–7] presented an objection to psychological continuity theories of personal identity. In principle, there could turn up in this century not one but two people both of whom are psychologically continuous with Guy Fawkes; so psychological continuity cannot be the same thing as personal identity, because psychological continuity can be a one–many relation. The subsequent massive literature on personal identity derives mostly from this remark. It was generally felt, despite Williams's protests, that this objection to psychological continuity theories applies with equal force to views of identity as consisting in, for example, bodily continuity, since 'fission' cases, in which two qualitatively identical bodies step out from the place hitherto occupied by one body, seem easily imaginable. In fact, any relation to which we might appeal in analysing personal identity seems capable in principle of being one–many.

Derek Parfit [1971] redrew the moral of Reduplication. He said that what is important to us is the potentially one–many relation in terms of which identity is analysed: 'identity is not what matters in survival'. There is a problem about how to understand this idea; let me give an analogy. Suppose that in the village we are having a discussion about whether baptism is important. A Distinction-Drawing Conservative might argue: 'What we have always valued about baptism is not the ceremony itself, but the associated celebration of the birth of the child and the inclusion of it in our community. And we could perfectly well have the celebration and inclusion without the ceremony of baptism.' A Traditionalist might reply that this is not possible: you can't, in fact, have just

the kind of celebration and inclusion that we value without the formal ceremony. And a Revisionary might enter the discussion and argue that the whole way we treat children has to be rethought from scratch—the institution of parenting, for instance, is questionable—and baptism abandoned in the process.

Parfit's argument that 'identity is not what matters in survival' is, on the face of it, the argument of a Distinction-Drawing Conservative. The argument is that when you reflect on a fission case you find that all we had ever valued about survival is still there, even though after fission there is no one around who is identical to the original person. The Traditionalist reply is that you could not in fact have what you traditionally valued in survival without identity; the things we value in survival actually depend on identity (Cassam [1992], McDowell [1997]). At this point, Parfit may reach for the Revisionary position: that we have to rethink from scratch the whole way we think about persons and survival. This is a much more ambitious line than Parfit actually argues for. It is not obvious what the standpoint is from which our current practices are to be criticized, nor what the problem is with our current evaluations; and it is hard to see what the relevance of fission cases would be to a Revisionary argument. Inspired by Parfit, though, some philosophers have none the less begun on the Revisionary project.

Whether psychological continuity was the right notion to appeal to in discussing personal identity drew a rich discussion. Shoemaker [1970], by defining the notion of 'quasi-memory', showed how psychological continuity would have to be characterized were it to be described in terms that did not simply presuppose, and so could not explain, the identity of the rememberer with the person whose states were remembered.

A number of philosophers, led by Wiggins [1980], argued that psychological continuity should not be viewed as giving a criterion of identity; its role is rather in determining whether the members of a species are persons. There is no such thing as a criterion of identity specifically for persons: persons are, usually, animals, and the criterion of identity for them is the criterion of identity appropriate for the relevant species of animal.

Others went back to Williams's original Reduplication argument and held that, in a case where there is fission, there are not one but two people, in the same place at the same time, admittedly, before the fission. This Multiple Occupancy thesis was developed in a number of ways (Lewis [1976]; Noonan [1989]).

By the 1990s this discussion had begun to peter out somewhat: the main rival views and their difficulties seemed relatively well understood, and it was not exactly clear how to make further progress. One line of thought was to ask how it could have happened that there was so much uncertainty as to what, if anything, we were talking about in using the first person. Anscombe [1975] had argued that there is no such thing as a Russellian acquaintance

with, or experience of, the 'self', to provide knowledge of the reference of the first person. Hence the first person does not refer to anything. A consequence of this is that there is no 'self' required to exist as the reference of 'I', into whose identity conditions we have to enquire. This idea complements the animalism advocated by Wiggins and others. But if the first person does refer, how could it have happened that we were using it so confidently while there was such radical uncertainty as to what we were talking about? This kind of thinking drew a mass of work on the first person, discussing questions such as the role of knowledge of one's body, or knowledge of one's spatial location, in an understanding of one's own uses of 'I' (Bermudez, Marcel, and Eilan [1995]; Cassam [1997], [Eilan, McCarthy, and Brewer [1993/1999]; Evans [1982]).

References

- Anscombe, G. E. M. [1975]: 'The First Person', in Q. Cassam (ed.), 1994, *Self-Knowledge*, Oxford: Oxford University Press, pp. 40–59.
- Avramides, A. [2001]: *Other Minds*, London and New York: Routledge.
- Bermudez, J. L., Marcel, A., and Eilan, N. (eds.) [1995]: *The Body and the Self*, Cambridge, MA: MIT Press.
- Block, N. [1996]: 'Mental Paint and Mental Latex', in E. Villanueva (ed.), *Philosophical Issues 7*, Northridge, Calif.: Ridgeview.
- Block, N., and Fodor, J. [1972]: 'What Psychological States are Not', *Philosophical Review*, **81**, pp. 159–81.
- Cassam, Q. [1992]: 'Reductionism and First-Person Thinking', in D. Charles and K. Lennon (eds.), *Reduction, Explanation and Realism*, Oxford: Oxford University Press, pp. 361–80.
- Cassam, Q. [1997]: *Self and World*, Oxford: Oxford University Press.
- Chalmers, D. [1996]: *The Conscious Mind*, Oxford: Oxford University Press.
- Chomsky, N. [1976]: *Reflections on Language*, London: Fontana/Collins.
- Davidson, D. [1963]: 'Actions, Reasons and Causes', *Journal of Philosophy*, **60**, pp. 685–700.
- Davidson, D. [1970]: 'Mental Events', in L. Foster and J. Swanson (eds.), *Experience and Theory*, Amherst, MA: University of Massachusetts Press.
- Davies, M. [1995]: 'Reply: Consciousness and Varieties of Aboutness', in C. Macdonald and G. Macdonald (eds.), *Philosophy of Psychology*, Oxford: Blackwell, pp. 356–92.
- Davies, M., and Stone, T. (eds.) [1995]: *Folk Psychology*, Oxford: Blackwell.
- Dretske, F. [1995]: *Naturalizing the Mind*, Cambridge, MA: MIT Press.
- Eilan, N., McCarthy, R., and Brewer, B. (eds.) [1993/1999]: *Spatial Representation*, Oxford: Oxford University Press.

- Evans, G. [1982]: *The Varieties of Reference*, Oxford: Oxford University Press.
- Feigl, H. [1959]: 'The "Mental" and the "Physical"', in H. Feigl, G. Maxwell, and M. Scriven (eds.), *Minnesota Studies in the Philosophy of Science*, vol. II, Minneapolis: University of Minnesota Press.
- Fodor, J. [1983]: *The Modularity of Mind*, Cambridge, MA: MIT Press.
- Gopnik, A. [1993]: 'How We Know Our Minds: The Illusion of First-Person Knowledge of Intentionality', *Behavioural and Brain Sciences*, **16**, pp. 1–14.
- Jackson, F., and Pettit, P. [1990]: 'Program Explanation: A General Perspective', *Analysis*, **50**, pp. 107–15.
- Jaspers, K. [1959]: *General Psychopathology*, Manchester: Manchester University Press.
- Kim, J. [1998]: *Mind in a Physical World*, Cambridge, MA: MIT Press.
- Kripke, S. [1972]: 'Naming and Necessity', in D. Davidson and G. Harman (eds.), *Semantics of Natural Language*, Dordrecht: Reidel.
- Lewis, D. [1976]: 'Survival and Identity', in A. Rorty (ed.), *The Identities of Persons*, Berkeley: University of California Press.
- McDowell, J. [1997]: 'Reductionism and the First Person', in J. Dancy (ed.), *Reading Parfit*, Oxford: Blackwell.
- Millikan, R. G. [1984]: *Language, Thought and Other Biological Categories*, Cambridge, MA: MIT Press.
- Noonan, H. [1989]: *Personal Identity*, London: Routledge.
- Parfit, D. [1971]: 'Personal Identity', *Philosophical Review*, **64**, pp. 3–27.
- Peacocke, C. [1983]: *Sense and Content*, Oxford: Oxford University Press.
- Place, U. T. [1956]: 'Is Consciousness a Brain Process?' *British Journal of Psychology*, **47**, pp. 44–50.
- Putnam, H. [1967]: 'Psychological Predicates', in W. Capitan and W. Merrill (eds.), *Art, Mind and Religion*, Pittsburgh: University of Pittsburgh Press, pp. 37–48.
- Putnam, H. [1975]: 'The Meaning of "Meaning"', in K. Gunderson (ed.), *Minnesota Studies in the Philosophy of Science*, vol. VII, Minneapolis: University of Minnesota Press.
- Quine, W. V. [1960]: *Word and Object*, Cambridge, MA: MIT Press.
- Russell, B. [1910–11]: 'Knowledge by Acquaintance and Knowledge by Description', in B. Russell, *Mysticism and Logic*, London: Allen & Unwin, 1917, pp. 152–67.
- Ryle, G. [1949]: *The Concept of Mind*, Harmondsworth: Penguin.
- Searle, J. [1995]: 'Consciousness, Explanatory Inversion and Cognitive Science', in C. Macdonald and G. Macdonald (eds.), *Philosophy of Psychology*, Oxford: Blackwell, pp. 331–55.
- Shoemaker, S. [1970]: 'Persons and their Pasts', *American Philosophical Quarterly*, **7**, pp. 269–85.
- Smart, J. J. C. [1959]: 'Sensations and Brain Processes', *Philosophical Review*, **68**, pp. 141–56.
- Strawson, P. F. [1959]: *Individuals*, London: Methuen.
- Tye, M. [1995]: *Ten Problems of Consciousness*, Cambridge, MA: MIT Press.

Wiggins, D. [1980]: *Sameness and Substance*, Oxford: Blackwell.

Williams, B. [1956–7]: 'Personal Identity and Individuation', *Proceedings of the Aristotelian Society*, **57**, pp. 229–52.

Wittgenstein, L. [1953]: *Philosophical Investigations*, Oxford: Blackwell.