Hume's Account of Personal Identity

David Pears
Oxford University

Follow this and additional works at: http://digitalcommons.brockport.edu/phil_ex

Part of the Philosophy Commons

Repository Citation

This Article is brought to you for free and open access by Digital Commons @Brockport. It has been accepted for inclusion in Philosophic Exchange by an authorized editor of Digital Commons @Brockport. For more information, please contact kmyers@brockport.edu.
Hume's account of personal identity is superior to most other accounts because it is more penetrating. It is given in two parts. First, there is the reductive theory set out in the main body of the *Treatise*, and then in the Appendix there is the recantation of that theory. Since the two parts cancel out, it is unrealistic to ask whether the whole account contains more truth, or less error, than other accounts. But we can assess it in another way. We can ask whether the trip was a good one, and whether we learned anything from the arguments which Hume uses, first against rival theories, and then against his own. When his discussion is assessed in this way, it should get high praise. It made a permanent difference to the subject.

The theory offered in the body of the *Treatise* was intended to explain the fact that a person is a single unified being, persisting through time. It is a theory based on several axioms, and Hume builds it up on this basis very carefully and very scrupulously, never allowing himself to slip extra material into the structure, or extra assumptions. But in the end he finds that the theory does not fit the fact which it was designed to explain. So he is faced with a dilemma: either he must argue that the fact is not really what he had taken it to be, or else he must abandon the theory. He chose to abandon the theory, and he confessed that he was unable to find a better one to put in its place.

That is a summary description of an investigation which is lengthy and complex. I shall now go back to the starting-point and explain the axioms on which Hume based his supposedly unsuccessful theory. These may be divided into two groups. The first group is concerned with the conditions of perfect identity; and the second is concerned with the nature of the connection between the components of a composite thing.

The axioms in the first group are the following:

1. An incomposite thing enjoys perfect identity so long as it lasts.
2. A composite thing enjoys perfect identity so long as there is no change in the identity of its incomposite components.
3. There is no third way of achieving perfect identity through time.

Axiom (3) is directed against any third way that might be suggested. But the suggestion which Hume had chiefly in mind, and on which he spends a lot of argument, is the suggestion that a composite thing might achieve perfect identity through a constant substrate. This substrate would be a substance in one sense of that word, but not in the sense in which Hume allows that an incomposite thing would be a substance. That is a different use of the word. An incomposite thing would be a detectable substance, but a substrate would be an undetectable substance, and so, according to Hume, even if its nature were intelligible, its existence would be dubious. An incomposite thing is a kind of substance which
HUME'S ACCOUNT OF PERSONAL IDENTITY

avoids these disadvantages, but has another disadvantage instead; it selfishly refuses to extend perfect identity to anything other than itself.

Let us now look at the axioms in the other group, which is concerned with the nature of the connection between the components of a composite thing. I give these in Hume's words:

(4) "All our distinct perceptions are distinct existences." ¹
(5) "The mind never perceives any real connection among distinct existences." ²

These two axioms are stated for the special case in which the composite thing is a person's mind. But Hume really subscribes to generalised versions of them, which apply to composite material objects as well as to minds.

The next thing is to ask how he proceeds to build up his supposedly unsuccessful theory on the basis provided by these axioms. But before tackling that question I want to divide his discussion of personal identity into two layers, in order to make it more manageable. On the surface there is his elaborate argument about identity, diversity and change, leading to the conclusion that persons do not enjoy perfect identity through time. Let me call this "the surface plot", meaning not that it is superficial or unimportant, but only that it is explicit. Beneath the surface plot, and largely hidden by it, there is what I shall call "the underplot".

This division of Hume's discussion may be pictured as a horizontal line, whereas the other division that I mentioned, between the theory itself and the recantation of it, would be drawn as a vertical line. Above the horizontal line the surface plot is played out. The distinctive mark of the surface plot is that it treats persons like ordinary composite material objects. I do not mean that Hume treats persons as material objects. Far from it. After a few allusions to human bodies, he turns his back on them, and addresses himself to the narrower task of explaining the unity of human minds rather than human beings. What I mean is that, having restricted himself to human minds, he treats them like ordinary material objects. E.g. when he asks whether the impressions and ideas, which go to form a mind persisting through time, are organised in a way that confers perfect identity on that mind, he construes this question like the question posed by William James in his discussion of this subject: Are the individual beasts in a herd of cattle persisting through time organised in a way that confers perfect identity on that herd? ³ Both these questions get a negative answer based on Axioms (2) and (3), according to which, whenever there is any change in the identity of the components of a composite thing, physical or mental, it loses its title to perfect identity through time.

I am not yet in a position to give a detailed characterisation of the underplot. But what can be said in general is that it takes account of factors which are peculiar to persons and certain other animals, and perhaps some machines. This, of course is only a schema for describing the underplot, and different people will fill in the details in different ways. My way of filling them in will be based on Hume's text. I am interested only in those peculiarities of embodied minds which almost break the surface of his discussion of personal identity, but which do not quite succeed in breaking it. It would, of course, be possible to broaden
the scope of the inquiry, so as to bring in factors such as intentions and actions which lie entirely outside the drama as Hume presents it in Book I of the *Treatise* and in the Appendix. But I prefer to keep the inquiry more narrowly focussed on to those two texts, because I want to exhibit the tensions between his surface plot and his underplot.

I shall not say much about the surface plot, because the general pattern of it is tolerably clear. It is important to notice that right from the start Hume turns his back on the human body, and concentrates on the unity of the human mind. He easily shows that a mind does not satisfy either of the two conditions of perfect identity laid down in Axioms (1) and (2). Then he dismisses the third suggestion, that it might achieve perfect identity through a substrate. This is dismissed by an appeal to Axiom (3). Anyone who claims to have a distinct impression of his self as a separate entity is guilty of "a manifest contradiction and absurdity". 4 I take it that the absurdity that he means is the absurdity of identifying an enduring substrate with any of the components which go to make up the composite thing whose substrate it is. At least, that is how the absurdity is presented in the surface plot. Just as the substrate of a lump of rock cannot be identified with any of its physical components, so too the substrate of a mind cannot be identified with any of its mental components.

But here I must interrupt the exposition of the surface plot in order to glance at the underplot. For the surface plot does not exhaust the richness of the absurdity which Hume is trying to expose. In the case of a material object, such as a table or chair, it is absurd to identify the suggested substrate with any detectable component. But in the case of a mind there is an extra dimension to the absurdity. That is, first, the parallel absurdity of identifying the substrate with any detectable component of the mind, i.e. with any impression or idea. Then there is the additional complication that, if we did make such an identification, the impression of the self would be an impression of another impression, and, therefore, in Hume's terminology, an impression of reflection. Now this complication is not enough, in itself, to lead to any further absurdity. For it would have been possible to argue, as William James did later, that consciousness simply consists in the fact that one component of the mind reflects another. 5 But Hume's adversaries required the self to be a single subject, and such a subject could hardly get an impression of itself. James' theory that the subject is the passing thought, which apprehends earlier thoughts but not itself, avoids this absurdity, but only by abandoning the requirement of a single subject. Thus something which is simple enough in the surface plot carries richer implications in the underplot.

Let me return to the surface plot. The next step in its development concerns the relations between the mental components which go to form a single mind enduring through time. According to Hume, these relations are resemblance and causation. As demonstrated, they do not produce perfect identity, but they do produce the inferior substitute with which we have to rest content when we leave the philosopher's study. In other words, when Hume wrote the text of the *Treatise*, he believed that his theory of personal identity, founded on resemblance and causation, was an adequate theory, in spite of the fact that it
HUME'S ACCOUNT OF PERSONAL IDENTITY

does not satisfy philosophers' dreams about perfect identity. He believed as we say nowadays, that his analysis of the concept of personal identity was correct.

Since he had ceased to believe this by the time that he wrote the Appendix, it might be a good thing to mention two relations which have a strong claim to be included on his list, but which are not included on it. He does not include contiguity, in spite of the importance of temporal contiguity in the mental life of a person, and in spite of the fact that causation, which he does include, is said by him to involve contiguity. Another equally striking exclusion is the memory relation. This is the relation which holds between a memory impression and the earlier mental component of which it is a memory impression. When I say that he excludes the memory relation from his list, I mean only that he does not treat this relation as part of the basis of personal identity. Naturally, he thinks that memory is indispensable, but only as a means of acquiring knowledge of one's own identity. For without memory, how could anyone discover that a series of mental items really were related by resemblance and causation, which, according to the theory, are the two basic relations? How could he even discover the existence of the mental items? But this does not make the memory relation into a third basic relation. It is true that he allows it a minor role in producing personal identity, as opposed to discovering it. Because he thinks that memory impressions are replicas, the memory relation multiplies resemblances and so helps to produce personal identity as a sort of side effect. But this work is done through resemblance, and resemblance is a basic relation already on the list, and so the memory relation does not acquire a title to a place of its own on the list.

It might be argued that Hume was right to refuse to put the memory relation on the list of relations that constitute personal identity. Perhaps memory does only discover personal identity. This is a difficult matter to settle, and the difficulty can be exhibited in the following way. Suppose that he had said that the memory relation also helps to produce personal identity by multiplying causal connections between mental items. This would have been a much more important point than his suggestion that it multiplies resemblances. As far as I know, he never says that memory helps to produce personal identity in this way. But if he had made this point, he could have argued very plausibly that it locates the most important thing that memory contributes to constituting personal identity. But would this give the memory relation a title to a place of its own on the list of basic relations? Probably not. For if he had made the point about memory and causality, he could still have defended his refusal to give the memory relation a place of its own on the list. Memory would play its further role only through causation, which is already on the list. So perhaps his refusal to add the memory relation to the other two is not wrong, and the only fault in this part of his argument is that he does not offer a full justification of his refusal.

What then is the dénouement of the surface plot? If we do not include the Appendix, the story ends with Hume's acceptance of a reductive theory. A human mind is composed of impressions and ideas related by resemblance and causality. He argues that there is no real alternative to this theory. Those who
say that a human mind enjoys a tidier type of identity have simply made a mistake about the nature of the scale of better and worse types of identity. When we place various kinds of things on this scale, we are merely applying axioms (1), (2) and (3) to the empirical phenomena. If the result is that a physical atom exhibits perfect identity through time, while cabbages and kings do not, that is a final result, and there is no appeal beyond it. Of course, someone might challenge Hume’s account of the empirical phenomena, and claim that the identity of a physical atom, or of a cabbage, is not as he describes it. Or someone might make the more radical suggestion that we should not use the three axioms to define perfect identity; or even that we ought to give up talking about perfect identity altogether, because each kind of thing has its own appropriate criterion of identity, and there is no competition between them. But if we do construct the scale in the way sketched by Hume, and if we cannot find any mistake in his description of the empirical phenomena, then that result is final. It is a misunderstanding of the nature of the scale to bring in the unempirical concept “substrate”, and to try to use it as the basis of a third type of perfect identity, which would be a sort of consolation prize for those who fail in the empirical competition.

So Hume is satisfied with his reductive theory, and he has an explanation of his adversaries’ dissatisfaction with it. His adversaries are obsessed with perfect identity, and try to find it where it does not exist. But when we follow the surface plot into the Appendix, there is a dramatic change. All his satisfaction with his theory vanishes. He still refuses to accept the suggestion that we have an unempirical concept of substance. But when he reviews his account of the connections between the impressions and ideas of a single person, he finds it defective. Since philosophers’ recantations are not too common, let me quote some of this one: “If perceptions (i.e. impressions and ideas) are distinct existences, they form a whole only by being connected together. But no connections among distinct existences are ever discoverable by human understanding. We only feel a connection or determination of the thought to pass from one object to another. It follows, therefore, that the thought alone feels personal identity, when reflecting on the train of past perceptions that compose a mind; the ideas of them are felt to be connected together, and naturally introduce each other... But all my hopes vanish when I come to explain the principles that unite our successive perceptions in our thought or consciousness. I cannot discover any theory which gives me satisfaction on this head.”

This really is a recantation. It is not just a case of Hume’s common flirting with his adversaries’ feelings. His treatment of causal necessity contains several examples of this kind of insincerity—or, it may be, irony. I am thinking of the passages in which he expresses the fear that his reductive account of causal necessity may not only look too sceptical, but actually be too sceptical. But on the whole he is satisfied that that theory is adequate because it covers everything that is empirically accessible. So his settled conclusion about that matter is that his adversaries are misled by the mind’s “great propensity to spread itself on external objects, and to conjoin with them any internal impressions which they occasion.”
HUME'S ACCOUNT OF PERSONAL IDENTITY

Why then, is he not equally satisfied with his reductive theory of personal identity? We may suspect that he is influenced by something in the underplot. For example, minds are self-reflexive, and so, though it may be a good explanation of causal necessity to say that the mind spreads itself on external objects, it does not sound so good an explanation of personal identity to say that it spreads itself on internal objects. However, Hume does not mention this difficulty at this point. Nor does he explicitly introduce any other motifs from the underplot. He sums up his reasons for rejecting his theory of personal identity in the following words: "In short, there are two principles which I cannot render consistent, nor is it in my power to renounce either of them, viz. that all our distinct perceptions are distinct existences; and that the mind never perceives any real connection among distinct existences. Did our perceptions either inhere in something simple and individual, or did the mind perceive some real connection among them, there would be no difficulty in the case. For my part, I must plead the privilege of a sceptic, and confess that this difficulty is too hard for my understanding." 8

This is not irony. Hume was a subtle writer, capable of irony which often passes undetected. So it could have been irony. But it is not in fact irony, and the proof of this is that it is entirely different from his later reaction to his reductive theory of causal necessity. In his Abstract of the Treatise he admits the sceptical character of that theory but claims that it is correct.

Why, then, did he recant about personal identity? In the remainder of this paper I shall go through the underplot trying to show that his real reasons for recanting are to be found there. But, naturally, I shall start from the reasons that he himself gives in the Appendix.

The explicit reasons, given in the Appendix, belong to the surface plot. He does not say anything that is not also true of ordinary material objects, and his argument could equally well be applied to the identity of cabbages. This is a very striking fact. An even more striking feature of his recantation is that he sums it up by repeating Axioms (4) and (5) and saying that he cannot renounce either of them and yet that he is unable to render them consistent. Now Axioms (4) and (5) belong to the second group that I distinguished at the beginning of this paper. They are concerned with the nature of the connections between the components of composite things, and they stipulate that these connections are always contingent. What then can he mean when he says that he cannot render them consistent? They do not even look inconsistent with one another. In fact, it would be plausible to argue that (5) merely gives the definition of the word "distinct" as it is used in (4). 9 If this is correct, Hume's point is that there are a priori connections between ideas, and associational connections between ideas produced, e.g. by constant conjunctions of impressions, but no third kind of connection called "real connection".

I think that the solution to this problem of interpretation is that Hume means not that the two axioms are inconsistent with one another, but only that taken together they are inconsistent with the fact that a person is a single unified being persisting through time. In other words, he takes this fact to imply a greater degree of unity than the two axioms allow. If he had been prepared to revise his
interpretation of the fact, he would not have had to recant. But he found himself unable to accept a more reductive interpretation of the fact, and so he recanted.

As far as the text goes, this interpretation fits the whole tenor of the passage. Moreover, he never says that the two axioms are inconsistent with one another, but only that they are inconsistent. But I must admit that even this is an odd way of expressing the view that I am attributing to him, and possibly the explanation is that the brevity of the Appendix has made it very inexplicit at this point.

In any case, this interpretation will be convincing only if it can be explained why the lack of a real connection between the components of a mind left Hume dissatisfied. After all, the two axioms only require the connections to hold contingently. So if I have a taste impression followed by a memory idea of Paris, it will only be a contingent fact that the second followed the first. But what is wrong with that? He can hardly have supposed that such a connection ought to be non-contingent. Admittedly, in this example the connection happens to be causal, but then his main thesis about causal connections is that they only hold contingently. In any case, non-causal examples could easily be found.

It seems that the only way to answer this objection and to give an intelligible reconstruction of his reasoning is to draw on the underplot. Minds differ from ordinary composite material objects in more than one way, and it is likely that some of those differences will provide clues to his line of thought.

For example, the ownership of impressions and ideas has several well-known peculiarities. If I have a sense impression, there is no room for any question about its owner—the owner must be myself. Nor can I speculate that the sense impression might have been yours instead of mine, or that it might have existed on its own, not belonging to anybody. Such speculations lack sense. Now these peculiarities of the ownership of mental objects have to be accommodated in any viable theory of personal identity. But how was Hume to accommodate them in his theory? How was he to weave these threads from the underplot into a surface plot whose dominant pattern was set by the analogy between mental objects and ordinary material objects?

One would expect that there would be some distortion at this point, and in fact there is. Instead of saying that, if I have an impression, it must be mine, could not have belonged to anyone else, and could not have existed on its own, he wants to be able to say that, if I have an impression, it could not have failed to occur in the series that is myself. But that would not be the same thing. I think that it seemed to him to be the same thing because he pushed the analogy between mental objects and material objects too far. If a cow belongs to a particular herd, it need not have belonged to it, and it—the very same cow—might have belonged to a different herd, or even lived on its own. This may have suggested to him that, if he had allowed that a certain sense impression which occurs in the series that is myself might not have done so, then he would have been forced to allow that it—the very same sense impression—might have occurred in a series that is someone else, or even existed on its own. In fact no such concession would have been forced from him. He could have pointed out that there is a limit to the analogy between mental objects and material objects.
HUME'S ACCOUNT OF PERSONAL IDENTITY

But because he failed to see the limit, he thought that the only way to accommodate the peculiar features of the ownership of mental objects would be to say that, when a series contains a particular mental object, it could not have failed to contain it. In short, he confused the following two modal propositions:

(i) This series might not have included S.
(ii) S might have occurred outside this series.

The thesis, that Hume is here exaggerating the analogy between mental objects and material objects, must not be taken to imply that all types of material object have criteria which make their numerical identities independent of the numerical identities of material objects of any other type. That is not so. Although most types of material object are not identity-dependent in this way, some types are. For example, a particular brick is not dependent for its identity on the wall in which it has been incorporated, and it could have been incorporated in a different wall. But suppose that I point to the branch of a tree, and say that it—the very same branch—might have grown on a different tree. Here we have a material example of identity-dependence, and my speculation would lack sense. All that I can imagine is that another tree might have grown a branch exactly like this one, and that, at the same time, this tree might not have grown this one. So this kind of identity-dependence is not confined to mental objects.

However, it does seem to extend to all types of mental object, and it is, perhaps, especially puzzling in this area. Its puzzling character may be brought out through a contrast between a branch of a tree and a sense impression. An explanation of the identity-dependence of a branch would go something like this. In the case of a branch we could always adopt a new criterion, according to which its numerical identity would be tied to the matter out of which it is formed. Then the speculation that lacked a sense could immediately be given one. It would mean that that matter might have been absorbed by a different tree, and might have grown out of it in the form of a similar branch. Of course, someone might object that this would still not be a case of the very same branch growing on a different tree, because he might persist in treating the numerical identity of the tree as a necessary condition of the numerical identity of the branch. But there would be no mystery about this. We would have three discernible things to juggle with, the matter of the branch, its form, and its relation to a particular tree. These three things could be used in various ways to produce alternative criteria of numerical identity for branches. The relation to a particular tree is only one thing, and it is easy to see what is going on when someone refuses to allow the numerical identity of a branch to be independent of this relation.

But the whole affair is more mysterious in the case of a sense impression. For though two of the things are used in this case—the form (or quality) of the sense impression and its relation to a particular person—the third thing—its matter—is not used. Consequently, we do not have such a clear idea of what we would have to do in order to give a sense to the senseless speculation. Would we merely collapse the concept of the numerical identity of a sense impression into exact similarity? Or would we have to wait until we were in a position to base a new criterion of numerical identity on the matter of the nervous system?
To return to Hume—it may seem hard to believe that he could have exaggerated the similarity between sense impressions and material objects, such as bricks, to quite such an extent. For the facts about the ownership of mental objects are familiar, and so they might seem to be less malleable than this. But half thought out analogies are very powerful in philosophy, and there is ample evidence in the Treatise for this account of what was going on in his mind. For example, in the chapter on the Immateriality of the Soul, he says: "... since all our perceptions are different from each other, and from everything else in the universe, they are also distinct and separable, and may be considered as separately existent, and may exist separately, and have no need of anything else to support their existence. They are, therefore, substances, as far as this definition explains substance". This is not an isolated passage, and it marks out a line of thought which Hume was prepared to follow in the text of the Treatise, but not in the Appendix. The analogy with bricks, or perhaps with physical atoms, is being openly pushed to the extreme. In this way sense impressions seem to acquire a degree of independence which makes it impossible to explain the fact that one of mine could not have been one of yours.

This strange way of treating mental objects is connected with his account of the way in which memory helps a person to answer a question about his own earlier identity. He says that memory discovers, but does not constitute, personal identity, except in so far as it multiplies resemblances between a person's mental objects. This suggests that my memory puts me in touch with a number of mental objects, about which I then ask whether they belong to the series that terminates on myself at the present moment, and that I am supposed to answer this question by applying the criteria of resemblance and causation. But this description of my procedure presupposes that memory is an impersonal way of collecting data, which are then examined and assigned to myself or some other person. Did Hume then deliberately use "memory" to signify a faculty which puts me in direct touch with earlier mental objects that belong to any person, myself or another? Apparently not. For there is no evidence to support the view that in his discussion of personal identity he is intentionally presupposing a predicament that is not ours. Of course, he allows for errors of memory, but not for what Shoemaker calls "quasi-memory". On the other hand, in this discussion he does not even mention any of the ordinary ways in which we discover the objects in other people's minds. So there really is an unintentional presupposition that memory is an impersonal way of collecting data. How should we interpret it?

It should probably be connected not with any carefully thought out theory about the way that the data are, or might be acquired, but, rather, with his general picture of the world of mental objects. When he pushed the analogy between mental and material objects too far, it was natural for him to write as if we could establish the existence of mental objects without prejudice to the question of their ownership, as can be done with cattle. Then, if he retained his uncritical assumption that memory is the only source of the relevant data, it would be assigned a role which it could not possibly perform. For how could I rely on my memory for the existence of an earlier mental object, while re-
Hume's Account of Personal Identity

jecting the inference that it was mine?

Only too easily, it might be retorted, if the concept of memory is changed. But since there is no evidence that Hume was deliberately changing the concept in the required way, it is more likely that he slipped into the impersonal treatment of memory because he exaggerated the analogy between mental and material objects without fully realising the consequences.

So much for the second of the two possibilities whose unrealisability he laments in the Appendix: "did the mind perceive some real connection among them [our perceptions] . . . ". The first one still remains to be considered: "did our perceptions . . . inhere in something simple and individual . . . ". Why did he say that, if this possibility were realised, "there would be no difficulty in the case"? What made him wish that he had been able to accept this kind of theory?

This question is unlikely to be answerable in as straightforward a way as the question about his other wish. The theory that a mind is "something simple and individual" is such a panacea that the wish that it were an acceptable theory is likely to be overdetermined. Nevertheless it is surprising to find how very general Hume's stated reasons are. The considerations that he adduces apply not only to minds but also to composite material things. According to him, the only kind of identity enjoyed by all these composite things is the inferior, fictional kind. But if the problem is so widespread, why does he confess his inability to solve it only in the case of minds? Here too there must be something at work in the underplot. But what?

There are two things that may have been working in his mind at this point. He may have felt that, if he could not show that the impressions and ideas of a person were presented to a single thing, consciousness would remain unexplained. Or, alternately, he may have felt that, if he could not show that a mind is a simple continuant, he could not give an adequate account of the way in which it acquires and retains knowledge of the world in which it is placed. The first of these two problems is concerned with the phenomenology of what happens within a mind, while the second is concerned with its mechanism and its place in nature. Unfortunately, Hume does not say which of the two problems was exercising him. He gives clear reasons for rejecting the theory that the mind is "something simple and individual". If he had identified the self with any detectable component of a mind, the impression of the self would have been an impression of that component, and, therefore, an impression of reflection. But it is absurd to identify a composite thing with any of its components, and a single subject could hardly get an impression of itself in that way. Why, then, did he find this theory so attractive? He does not tell us, and the choice between the two interpretations must be based on indirect evidence.

The prevalent interpretation is the first one, which claims that he was hoping to find a way of explaining the phenomenon of consciousness. But there is strong circumstantial evidence against this interpretation. He never shows any sign of thinking that he needs to point to a single subject in order to explain consciousness. He took it for granted that one component of a mind can reflect another, and apparently did not feel that impressions of reflection posed any problem. He allowed himself to make use of intentionality without trying
to explain it, and he felt no doubts about its range because he treated impres-
sions of memory as unproblematic. In short, he did not see that there might be
difficulties about the synthetic unity of consciousness, and his course appeared
to be set towards the kind of theory that was later developed by William James.
In fact, he did not go so far as to identify the self with the passing thought. But
without violating the principles of his empiricism, he could have explained the
idea of the self as an idea of an expanding series of mental components. These
components would be identified through their positions in the series, and if a
number of them were all tokens of the same type-idea, their contents would be
identical. But the content of the idea of the self would increase as the series
expanded. No doubt this is a complicated theory, but at no point does Hume
show any signs of being troubled by the general problem of consciousness or by
the special problem of reflexivity.

His difficulty seems to begin when he has to explain how a mind acquires and
retains any knowledge of the world and of its place in it. For how can he explain
the working of memory in his system? If I wonder whether a man sitting
opposite me in a train is the one who sat opposite me yesterday, I would take
my own identity for granted. But if he questioned whether I was the man who
had occupied the same seat on yesterday's journey, I would have to verify my
claim that I was, perhaps by producing memory impressions which could be
checked. Similarly, the identity of a star could be established through a photo-
graph taken by a rocket-borne camera, or the identity of the camera could be
established through the images on its film. In both cases it is necessary that we
should be able to argue either from the identity of the recorder to the identity
of what it records, or in the reverse direction. This evidently requires that there
should be general agreement between the record and the recorder's independently
established history. But how could memory meet this requirement in Hume's
system?

That depends on how liberally his system is constructed. He occasionally
speculates about the physical basis of the mind, and, if he had allowed himself
to use that kind of material in his discussion of personal identity, he might have
given an adequate account of memory. But in fact he does not use it, and the
implication is that he can solve the problem entirely from the resources of the
mind. This restriction puts memory in an impossible position. Just as I, body
and mind, am related to the two appearances of the man in the train, so too I,
the remembering subject, ought to be related to any sense impressions that occur
in my mind. But this will not work. For in the train there are independent ways
of checking my identity, but there are no independent ways of checking the
identity of the remembering subject in Hume's restricted system. His wish that
the subject were single is merely the wish that this did not matter. Another
reason why the analogy does not work is that there is no material left over for
incorporation in an account of the causal mechanism of the remembering
subject. His wish that the subject were simple is merely the wish that this too
did not matter. However, it does matter that his picture of the mind impels him
towards a theory which makes the identity of the subject independent of any
checks and its operation independent of any mechanism. He knows that such a
HUME'S ACCOUNT OF PERSONAL IDENTITY

theory is a philosopher's dream, but he does not retrace his steps in order to find out which of them led him into the impasse.

FOOTNOTES

2 ibid.
4 Treatise, Book I, pt. 4, sec. vi; p. 301.
5 Principles of Psychology, p. 342.
6 Treatise, Appendix: p. 331.
8 ibid., Appendix; p. 331.