



Hume on Personal Identity

David Pears

Hume Studies Volume XIX, Issue 2 (November, 1993), 289-300.

Your use of the HUME STUDIES archive indicates your acceptance of HUME STUDIES' Terms and Conditions of Use, available at

<http://www.humesociety.org/hs/about/terms.html>.

HUME STUDIES' Terms and Conditions of Use provides, in part, that unless you have obtained prior permission, you may not download an entire issue of a journal or multiple copies of articles, and you may use content in the HUME STUDIES archive only for your personal, non-commercial use.

Each copy of any part of a HUME STUDIES transmission must contain the same copyright notice that appears on the screen or printed page of such transmission.

For more information on HUME STUDIES contact
humestudies-info@humesociety.org

<http://www.humesociety.org/hs/>

Hume on Personal Identity

DAVID PEARS

The question that I discuss in this paper has often been raised and it has been answered in many different ways. "Why did Hume retract his theory of personal identity?" He puts it forward in the main text of the *Treatise* with his usual panache, and then takes it back in the Appendix. Why? Many different answers have been given to this question and what most of them have in common is the confidence with which they are given. Each interpreter sticks to his own interpretation in spite of the plausibility of its rivals, few of which are actually excluded by what Hume says. For this is an extreme case of the underdetermination of interpretation by evidence. I confess that in spite of all the recent work on this subject, I remain loyal to my own suggestion, which I first made in 1975.¹

I will start with a sketch of my interpretation. Then I will quote Hume's recantation and develop an argument for my interpretation, based on that text and on others. After that, I will show how my suggestion is related to some of its rivals and I will say something about their comparative merits. Naturally, there is a considerable amount of overlap between the more plausible answers to the question, why Hume recanted.

1. My Interpretation

In his Appendix, Hume confesses that his theory of personal identity seems to him to understate the strength of the connections between the

impressions and ideas that constitute a single person; or, as I would prefer to express his doubts, the impressions and ideas that constitute a single person's *mind*. He says that he would have been content if he could have attributed all those perceptions to a single substance; or alternatively, if he could have found some *real* connection between them. But all that he can find are connections based on the three natural relations: resemblance, contiguity, and causation. And those connections are, he implies, not enough.

It is certain that he is thinking of a fundamental difficulty. For he evidently sees some reason for doing something which is simply not possible within his system, namely finding stronger connections between perceptions than those allowed by the doctrine of the main text of the *Treatise*. But what is his fundamental difficulty? The suggestion that I put forward was this: when an impression of sense occurs in my series of perceptions, it does not hold its place in the series by possessing any property, or by standing in any specific relation to any other perception in the series (except, of course, the relation of being co-experienced). Consequently, when I review the series, I do not find that the sensation is anchored by anything other than the fact of its contingent occurrence at the point where it occurred. Nevertheless, there is a strong modal statement that can be made about it: It, *the very same sensation*, could not conceivably have occurred in any other series. This modal statement is based on the special character of the identity-line of a particular perception: the identity-line of a particular perception starts at the point where it occurs in a particular series, and no sense can be attached to the speculation that *it, the very same perception*, might have occurred in some other series. However, in Hume's system, perceptions are treated like stars which form constellations, but only form them contingently, and it is obvious that the discovery of additional contingent relations between perceptions would never provide any support for the strong modal statement. Yet in Hume's system the only possible relations between perceptions are contingent relations. The other alternative, that the self is a substance, would, perhaps, do the trick, but for Hume it is unthinkable.

It may seem fanciful to suggest that Hume's problem was the difficulty of accounting for the necessary ownership of perceptions within the framework of his system. But it is hard to see what other reason he could have had for requiring stronger relations between the perceptions of a single mind than those allowed in the main text of the *Treatise*. If he looked up at the sky, he would get a visual impression of blue, but he could hardly have supposed that an impression of that type *had to* occur at that point in the series that constitutes his mind. Nor could he have supposed that that particular impression, whatever its type, *had to* occur at that point. The only necessity here that is intelligible and plausible is the necessity that binds it to the series in which it *has* occurred.

If in the Appendix he was worried by his failure to account for this

necessity, it is not surprising that he did not give us a precise formulation of the problem. It is difficult to formulate it within the framework of his system, and his misgivings would naturally be deep but vague. That, indeed, is how they are expressed in his recantation.

The problem was formulated after his time as the problem of the necessary ownership of perceptions. It is discussed by Wittgenstein in *Philosophical Remarks*, Ch. VI.² One of the points made there is that there is no experienced feature of the pain that I feel which shows me that it is mine. It is simply axiomatic that if I feel it, it is mine; and though Wittgenstein does not take the next step and say that, if it is mine, nobody else could conceivably have had it instead of me, there is no doubt that he would have taken it, if he had gone on to examine that aspect of the identity of sensations. For he does say that it is logically impossible for one person to have a sensation that is had by another person, and he must have accepted the counterfactual corollary of this axiom.

Now suppose that Hume had been asked what he thought about Wittgenstein's way of dealing with the strong modal statement, that my pain could not conceivably have occurred in any other series. I think that his reaction might have been to say that it presupposes that the subject who feels the pain is a mental substance in which the pain inheres; but that, he would have reminded us, is not a viable alternative.

However, if Hume did react to Wittgenstein's suggestion in that way, he would be wrong. For Wittgenstein goes on to argue that it is not even necessary for the word "I" to occur in my report of my pain-sensation. All that is necessary is that the sensation should occur in a series associated with a particular body, my body. That association provides a line of escape from solipsism; and it also provides a line of escape from the dilemma which Hume faced, or, at least, thought that he faced in the Appendix to the *Treatise*.

2. The Text of Hume's Recantation and the General Argument for My Interpretation

...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 skeptic, and confess that this difficulty is too hard for my understanding. (T 636)

All commentators agree that Hume is not complaining that his two principles are inconsistent with *one another*; or even that he finds it difficult to

reconcile them. They are plainly *not* inconsistent with one another. So he must mean that the conjunction of the two principles is inconsistent with some third thing. However, he does not tell us what the third thing is. It has to be a feature of our idea of personal identity; and it has to be a feature of that idea which is in *obvious* conflict with the conjunction of the two principles. For if the conflict were roundabout or recondite, Hume would have told us what feature of the idea of personal identity he had in mind. But he does not tell us. So it must be something close at hand and obvious.

This is where my interpretation begins to get a foothold. When Hume complains that his theory of personal identity underestimates the strength of the connection between the perceptions in a single mind, his implication must be that he has not done justice to the evident unity of a single mind. Of course, if this side of his theory were changed, the change would affect its other side—its solution to the problem of reidentification after a lapse of time. But as many commentators have pointed out, his primary concern is with the unity of a mind, and his interest in reidentification is secondary.

Now the word “mind” is a count-noun, and so the problem of the unity of a mind has two distinct aspects: the question is not only what holds a single mind together, but also what holds it apart from other minds. Hume distinguishes those two questions from one another, but says that it is best to isolate the first one, the question of the mind’s unity, and to deal with it separately. This strategy evidently suits his method of inquiry, which is to sift the contents of his own mind, searching for the origin of the idea of personal identity and to assume that others will reach the same results when they sift the contents of *their* minds. However, the strategy carries a risk: the problem of the differentiation of one mind from another will be put aside and it may cause trouble later. If my interpretation is correct, that is what actually happened.

When Hume treats the question “What holds a single mind together?” in his potentially solipsistic way, his favorite astronomical analogy takes on a different character. If perceptions are like stars, the question that is now analogous to his question about perceptions is not “What is the principle of individuation of a constellation?” but “What is the principle of individuation of the universe?” Now, cosmologists define *the universe* as everything that is theoretically observable, and that indicates that the word is not a count-noun but the name of a self-aggrandizing unit. That is precisely how Hume views his own mind when he adopts his potentially solipsistic strategy: his mind contains everything that is theoretically accessible to him. But, of course, when he takes this point of view, the first and most obvious answer to the question “What holds my perceptions together?” is the Wittgensteinian answer, “The mere fact that I have them”: just as the answer to the parallel question “What holds everything in the universe together?” is “The mere fact that everything in it is theoretically observable.”

Naturally, I do not mean that there is nothing more to be said about the unity of a single mind. Barry Stroud³ is right when he argues that Hume set himself the task of tracing the idea of personal identity back to its origin in human experience; and it is entirely possible that a chaotic mind would never achieve any sense of its own identity. If that is so, then the Wittgensteinian answer to Hume's question about the unity of the mind is incomplete: it may be a contribution to the *analysis* of a self-ascription of a perception, but it would not explain the origin of a person's sense of his own identity, without which there would not be any self-ascription of perceptions. But, of course, it would not follow that the Wittgensteinian answer is not needed, and it does seem to be the first essential step toward understanding the *meaning* of the self-ascription of a perception. If that is forgotten, or if it is put aside while the relations that impose order on the contents of a single mind are investigated, we may well get the illusion that those relations ought to be sufficient in themselves to distinguish one mind from another. That too is something that actually happens in Hume's treatment of personal identity, as I shall show in a moment.

Let me sum up so far. I have suggested, but not yet given much argument for, an interpretation of Hume's reasons for recanting his theory of personal identity. The suggestion is that, when he put aside the question "What holds one mind apart from another?" and concentrated on the internal unity of a single mind, he was storing up trouble for himself later. For if he succeeded in giving a naturalistic description of the origin of a single mind's sense of its own unity, he would be likely to assume that the same naturalistic description would serve to show what holds one mind apart from another. But that would be an illusion. In fact, the relations between perceptions to which he appeals when he is explaining the mind's sense of its own unity do not succeed in drawing the boundary between one mind and another. That might make it look as if his theory of mind was on the right lines, but that his particular development of it was faulty. But the trouble was more radical: for even if he improved his theory and succeeded in giving a list of relations which always hold between the contents of a single mind and never between the contents of two distinct minds, that still would not put him in a position to explain the strong modal statement that none of my perceptions could have conceivably belonged to anyone else. In other words, his superficial mistake is a sign of the profound mistake that underlies it.

3. Detailed Argument for My Interpretation

Hume says that the idea of personal identity is derived not from an internal impression of a separate self, but from three relations which connect the ordinary impressions and ideas of a single mind. The three relations are

Resemblance, Causation, and Contiguity. But contiguity in this case is *contiguity in time*, not as elsewhere in the *Treatise*, *contiguity in time and space*.

Consider first the negative part of this explanation, the rejection of the internal impression of a separate self. This is the rejection of a theory which nobody finds plausible, but its main interest for the exegesis of Hume's ideas lies in the role which he says *it would have* played if he had not been forced to reject it. He says that it would have solved a problem which he does not bother to identify—the problem that would also have been solved if he had been able to find real connections between the perceptions of a single mind. Now, according to my interpretation, that problem was the problem of the necessary ownership of perceptions, and it created two related difficulties for him. First, instead of adopting the axiom, "If I feel it, it is mine," he tried to base ownership on contingent relations between perceptions; but as I shall show in a moment, his relations hold not only between the contents of a single mind but also between the contents of different minds. It is this difficulty that would be circumvented if the contents of each mind really were undetachably related to a separate nuclear self, which would provide a unique identification of the owner. The second, more radical difficulty was that the three relations which, according to him, contingently connect the perceptions of a single mind, cannot possibly serve as a basis for the strong modal denial of alternative ownership; and it may be that this difficulty too would vanish if he could show that each person's perceptions were anchored by a separate nuclear self. But, of course, he found this idea unthinkable.

The *argument* that he uses against the internal impression of a separate self is also interesting. He is not content to point out that nobody actually has such an impression; he adds that the suggestion that they do have it involves "a manifest contradiction and absurdity." For since the self is constant and uninterrupted, the impression of the self ought also to be constant and uninterrupted but no impression can satisfy that requirement. However, the requirement is excessive: my eyes are the constant source of input to my visual field but it is not necessary that my impressions of my eyes should be constant too—it is enough that they are available.

It is worth observing that the suggestion that I have an internal impression of a separate self is absurd for another reason. Even if I thought that I had identified an internal impression of my self, there would be no way for me to verify my identification. In the physical world, I can verify that I see with my eyes, but in the mental world there are no resources for verifying that the object of a given internal impression really is the subject of all my mental achievements. The very suggestion is absurd.

The relation between this negative part of his theory of personal identity and his later recantation is not entirely clear. It is certain that he found the suggestion that the self is a mental substance unthinkable. It is probable that

he thought that if it had been a viable suggestion it would have provided a way of differentiating two minds, when his own theory broke down, i.e., in cases when the perceptions of two minds are connected by a relation, which, according to his theory, ought only to hold between the perceptions of a single mind. But there is also another possibility. He may have thought that the real necessity, which seemed to him to be required by the way in which the perceptions in a single mind must hang together, might somehow be provided by the hypothesis of a mental substance.

Let us now look at the positive part of his theory. The most striking deficiency of this part of his account of personal identity is that the three relations he uses do not draw the boundaries between two minds in the right place. It is obvious that two perceptions, A and B, each in a different mind, a and b, might be temporally contiguous and might also resemble each other. So causation ought to step in at this point and secure the assignment of A to a, and B to b. Now "causation" here means "direct causation," and Hume ought to be able to point to the cause of A in a's mind and to the separate cause of B in b's mind. However, causation in the mind lacks something which is essential to causation in the physical world, namely contiguity in space. For perceptions have no physical location. This produces an unfortunate effect: it is possible that causation, when it is judged in this way—i.e., without contiguity in space—might be found to link a single cause in a's mind both with A in a's mind and with B in b's mind. If this happened, the sets of perceptions which were collected by Hume's three relations could not be mapped onto separate minds as separate minds are distinguished in real life.

The source of this fault in Hume's theory has been pointed out by many commentators. For example, Don Garrett argues that since spatial contiguity cannot play any role in Hume's account of mental processes, two perceptions, each in a different mind, might share all their relations—in which case Hume could offer no reason for attributing one to one of the two minds, and the other to the other one.⁴ This is, indeed, a consequence of Hume's faulty account of mental causation, but I doubt if it is the consequence that worried him. It is more likely that he was worried by the consequence that I have just described: his theory does not draw the boundary between one mind and another in the right place. In any case, it seems to me that this fault is only the tip of the iceberg. For it is really only the first indication of two fundamental errors in his theory. If direct mental causation has to dispense with contiguity in space, it must rely on some substitute for contiguity in space: and the only plausible substitute is sameness of mind—i.e., if one mental event is going to cause another directly, they must be within the same mind. That is the most plausible substitute for contiguity in space as it figures in physical causation. But if this is the situation, sameness of mind will require some independent basis, and the only plausible independent basis is association with the same body.

Let me digress briefly in order to relate this diagnosis of Hume's difficulty to a suggestion made by Paul Grice and John Haugeland in an unpublished article on Hume's retraction of his theory in the Appendix to the *Treatise*. My diagnosis relies on the fact that direct mental causation involves sameness of mind and cannot, therefore, be used to explain sameness of mind without circularity. Grice and Haugeland argue, in a roughly parallel way, that causation presupposes personal identity and cannot, therefore, be used to explain it. However, the parallelism between the two diagnoses is not exact. For *they* are thinking of causal processes in the physical world, and their idea is that these processes will be based on the observations of different observers, who would need to be assessed for reliability. That assessment is essential to the verification of a physical causal process, but it could not be carried out until one mind had been separated from another; and the separation would be impossible unless the concept of personal identity were already in place. That is the diagnosis of Grice and Haugeland.

The main difference from my diagnosis is that mine is not concerned with physical causal processes, but with mental causal processes. That surely is Hume's concern in the Appendix. Also, if he *had* been concerned with physical causal processes in the Appendix, the varying reliability of different observers would be much less important than their locations in physical space. So the Grice-Haugeland diagnosis might be developed until it made contact with mine in the following way: different observers can be located in physical space only through their bodies, and, in the case of mental causation, the right substitute for spatial contiguity of cause and effect is belonging to the mind which is associated with a single body. But this is an elaborate development of their interpretation and that marks the distance between their diagnosis and mine.

To return to the mainstream of my argument, the first fundamental fault in Hume's system which surfaces at this point is the use of relations which hold contingently within a single mind to explain what counts as a single mind. The most that such a theory can achieve is an account of the origin of a person's *sense* of his identity, but even that would be incomplete without an account of his sense of his difference from other people. This fault is related to the fault criticized by Wittgenstein in the passage to which I referred at the beginning of this paper. Perceptions are ascribed to their owners not because they possess any kind of order in their owners' minds but simply because their owners have them.

The second fundamental fault which surfaces at this point is a misunderstanding of the basis of the axiom, "If I feel a pain, it is mine." Against anyone who offers Hume's explanation of the assignment of different perceptions to different minds as the correct explanation, it is necessary to trot out this axiom. It is necessary, but it is clearly not sufficient. For he will immediately want to know what the basis of the axiom is: he will want to know

what general features of human experience make it natural for us to adopt the axiom. Now Hume evidently assumes that the key to this problem, to which he alludes in the Appendix, is the organization of the perceptions of a single mind, and so, if he had considered the axiom, he would have searched for its basis among the relations which hold between those perceptions. But that would have been a mistake. The key to the problem of what we now call "the necessary ownership of sensations" lies not in the relations which unify a mind but in the criterion of identity of a sensation. The reason why one of mine could not have been one of yours is that the identity-line of a sensation starts at the point where someone has it, and no sense can be attached to the speculation of alternative ownership.

That is all I want to say by way of explanation and of argument for my interpretation, except for one more point. A complete diagnosis of the problem that Hume faces in the Appendix might have to say more about the association of a person's mind with a particular body. How, for instance, does this association provide a solution to Hume's problem which is a problem about the mind, not the body?

This is connected with a point that I made near the beginning of this paper. I suggested that Hume would have accused Wittgenstein of tacitly relying on individual substances, but that would be an unfair accusation, because what Wittgenstein relies on is the mind's association with a body. But, of course, that would only give Hume a further ground for complaint: for when I say that I have a thought, what I mean has nothing to do with my body—only with my mind.

The only way to answer this would be to analyze self-reference in detail, which is evidently beyond the scope of this paper. So I will simply make an *ad hominem* point. I do not see how Hume could possibly rehabilitate his account of direct mental causation without using the association of minds with bodies; and at a more fundamental level, I do not see how he could distinguish one mind from another without using their association with bodies.

4. Alternative Interpretations

I have already explained how two alternative interpretations are related to the one that I have been arguing for in this paper—one proposed by Garrett and the other proposed by Grice and Haugeland. Both are concerned with the first of the two fundamental mistakes I see in Hume's theory of personal identity, the mistake that he makes in his specification and explanation of the relations which, according to him, give a single mind its unity. The mistake lies in his use of causation *without* including in its basis contiguity in space, and *without* finding my substitute for contiguity in space. The result is a flawed account of the unity of a mind, which, Garrett argues, could lead us to attribute two indistinguishable perceptions to either or both of two minds.⁵

I argued that the consequence that worried Hume was that his theory drew the boundary between one mind and another at the wrong point. He ought to have used sameness of mind as a substitute for the unavailable spatial contiguity. I then presented the Grice-Haugeland diagnosis as a criticism of the circularity of using sameness of mind to explain causality and causality to explain sameness of mind, and finally I argued that we can escape from this circle only by appealing to the mind's association with a body.

But there is one alternative interpretation which I mentioned earlier in this paper, but have not yet examined, Barry Stroud's.⁶ I should say immediately that I misunderstood his interpretation in my recent book on Hume,⁷ taking it to be concerned with the observation of physical causal processes, like the Grice-Haugeland interpretation. In fact, it is concerned with mental causation. Stroud argues that Hume's use of mental causation is not really circular, because he is not trying to *analyze* the idea of personal identity, but only to trace the idea back to its origin in human experience. This enterprise, Stroud says, is successfully carried out by Hume up to a certain point. That point is marked by a proviso: "provided Hume confines himself to the contents of his own mind." Beyond that point, his success comes to an end because he cannot explain how perceptions originally came to be grouped in the sets that we treat as individual minds.

As I said earlier, I agree with Stroud's claim that Hume is trying to give a naturalistic account of the origin of the idea of personal identity, and not an analysis of the idea. But, like Garrett, I do not see why that requires him to explain how perceptions originally came to be grouped as they are. It seems to me that he is not required to explain why our perceptions start their careers in different minds, each attached to a different body. All that he is required to explain is the way in which this fact about our perceptions influences our conceptualization of them and our thoughts and beliefs about them. However, even this lesser explanatory task must involve an analysis of our use of the word "mind" as count-noun. But that is precisely what is lacking in Hume's naturalistic account of the origin of the idea of personal identity. So his account is lame in one foot: it fails to deal with the physical basis of the differentiation of one mind from another. The result is that his success, so long as he confines himself to the scrutiny of his own mind, is severely limited—not only that, but it also inevitably generates the kind of trouble that he takes to be so serious in the Appendix to the *Treatise*.

REFERENCES

1. David Pears, "Hume on Personal Identity," *Questions in the Philosophy of Mind* (London: Duckworth, 1975), 208–223.
2. Ludwig Wittgenstein, *Philosophical Remarks*, edited by Rush Rhees and translated by Raymond Hargreaves and Roger White (Oxford: Blackwell, 1975).
3. Barry Stroud, *Hume* (London: Routledge and Kegan Paul, 1977), Chapter VI, "The Idea of Personal Identity," 118–140.
4. Don Garrett, "Hume's Self-Doubts about Personal Identity," *The Philosophical Review* 90 (1981): 337–358.
5. Ibid.
6. Stroud.
7. David Pears, *Hume's System* (New York: Oxford University Press, 1990), 148–150.