Personal Identity and the Concept of a Person

By John Perry

[Originally appeared in G. Floistad (ed.). Chronicles of Institut International De Philosophie. Volume IV, Philosophy of Mind (Philosophy: A New Survey. The Hague: Martinus Nijhoff. This version is Chapter 7 of Identity, Personal Identity and the Self, Hackett Publications, 2001]

Philosophers approach the concept of a person from two directions. In ethics and political philosophy it often is taken as primitive, or at least familiar and not requiring elucidation, but persistent inquiry and difficult problems make a deeper look inevitable. In discussing abortion, for example, one can hardly invoke principles about rights and welfare of persons concerned, without facing the question about which concerned parties are, in fact, persons, and what that means. One moves remorselessly from issues of rights and responsibilities to questions of consciousness, self-awareness, and identity---from the moral to the metaphysical.

From the other direction, no comprehensive epistemology or metaphysic can avoid the question of what persons---our primary examples of knowers and agents---are, and how they fit into the universe, whether as illusion, phenomena, or things in themselves. Answers to this question will have consequences in the ethical sphere.

These approaches meet in the problems of freedom and in the problem the recent history of which I discuss: personal identity. It is the identity of the knower over time that seems to be both the ground and the result of empirical knowledge, and identity of the moral agent that seems presupposed by notions of responsibility, guilt, decision, and freedom.

I shall discuss a number of contributions by philosophers to our understanding of personal identity. I shall follow a specific path through the literature, which means I shall have to ignore a number of contributions that lie to one side or the other. The discussion is mainly metaphysical and epistemological, but questions of ethical significance are posed.

1. Personal Identity from Locke to Shoemaker

Our path will begin with Sydney Shoemaker's seminal book, *Self-Knowledge and Self-Identity*, published in 1963. But before starting on the path proper, it will be helpful to glance at the historic sources of the problems Shoemaker discusses. Although questions

of personal identity are central to Idealism, from Kant to Royce, Shoemaker's book skips over this tradition (except, perhaps, as it enjoys a twilight existence in Wittgenstein's thought) and, like so much of twentieth century analytical philosophy, picks up the problem as it was left by empiricist and common sense philosophers of the seventeenth and eighteenth centuries. The most important of these was John Locke, who added a chapter on personal identity to his *Essay Concerning Human Understanding* in 1694. Identity of persons consists in continuity of consciousness, and this seems to be provided by links of memory: "as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person ..." (section 9). Thus Locke appears to analyze self-identity in terms of self-knowledge, and provides the theme of Shoemaker's book and the dominant topic in the discussions to follow.

Locke distinguished identity of person from identity of spiritual substance on the one hand and identity of human body ("identity of man") on the other. The second distinction he argues for with a striking thought experiment:

For should the soul of a prince, carrying with it the consciousness of the prince's past life, enter and inform the body of a cobbler, as soon deserted by his own soul, everyone sees he would be the same person with the prince, accountable only for the prince's actions ..." (section 15).

The use of thought experiments which are putative cases of "body-transfer" was to become a focus of discussion two hundred and seventy years later, but at the time Locke wrote his distinction between identity and person and identity of soul or spiritual substance was more controversial.

"Let anyone reflect upon himself, and conclude that he has in himself an immaterial spirit, which is that which thinks in him, and, in the constant change of his body keeps him the same: and is that which he calls himself: let him also suppose it to be the same soul that was in Nestor or Theristes, at the siege of Troy ... but he now having no consciousness of any of the actions either of Nestor or Theristes, does he conceive himself the same person with either of them? Can he be concerned in either of their actions? attribute them to himself, or think them his own, more than the actions of any other man that ever existed?" (Locke, 1694, sec. 14).

Locke also held that it is possible, for all we know, that consciousness can be transferred from one substance to another, so "two thinking substances may make but one person." (section 13). This outraged Joseph Butler:

"... in a strict and philosophical manner of speech, no man, no being, no mode of being, nor any thing, can be the same with that, with which it hath indeed nothing the same. Now sameness is used in this latter sense when applied to persons. The identity of these, therefore, cannot consist with diversity of substance" (Butler, 1736; see also Reid, 1785).

The idea that personal identity could be analyzed in terms of memory was used by twentieth century empiricists who attempted to analyze the self as a logical construction from momentary experiences. This project, more Humean than Lockean, requires a relation between those experiences which will group them into sets of "co-personal" experiences. Though Hume rejected Locke's memory theory (Hume, 1741) , and Locke himself did not hold a bundle theory, the two doctrines seem to fit together naturally: we use memory to hold the bundle together through time. The clearest expression of this view comes from H.P. Grice in his fine essay "Personal Identity" (1941). Grice labors to discredit the pure ego theory of the self, a descendant of the view that personal identity consists in sameness of spiritual substance, and to put in its place a "modification of Locke's theory of personal identity." We can understand Grice's subtle and sophisticated theory as the result of successive accommodations to counterexamples, starting with Locke's view. Reconstructing Locke's view within Grice's framework, we begin with experiences. Those that can be known by introspection to be simultaneous belong to the same total temporary state or *t.t.s.* (p. 88). Thus we may imagine the realm of experience broken into discrete bundles, each t.t.s. being the experiences belonging to a single person at a given time. Locke's theory then is seen as giving us a principle for stringing these bundles together through time, giving us persons as enduring entities. His view is simply that *t.t.s A* and *t.t.s B* belong to the same person if and only if the latter contains an experience that is a memory of some element of the earlier. But this permits Thomas Reid's famous brave officer paradox: the boy is the officer (for the officer remembers stealing apples), and the general is the officer (for the general remembers leading the charge), but, since the general doesn't remember anything the boy thought or did, the general is not the boy (Reid, 1785).

Grice's final theory goes roughly as follows. Consider the relation *t.t.s A* has to *t.t.s B* if either one could contain a memory of an experience contained in the other. Any set of experiences which is closed under this relation, and contains no subsets closed under it, we may call a Grice set. (A set x is closed under a relation R if anything that has R to any member of x is in x.) *T.t.s.*'s are members of the same Grice set if and only if they are stages of a single person. The theory gets around the brave officer paradox, and other

problems of Grice's own devising, by allowing indirect memory links, such as that between the general and the boy, to confer identity.

Between the publication of Grice's article and the publication of Shoemaker's book, Ludwig Wittgenstein exerted tremendous influence on philosophy, and Shoemaker's perspective and theoretical approach were very Wittgensteinian in some respects. In particular, Shoemaker makes heavy use of the concept of a criterion and of the asymmetries between first- and third-person reports. Shoemaker is also sympathetic to Butler and Reid, trying to bring out epistemological insights that motivate their criticisms of Locke, without adopting their metaphysics of immaterial substances. Shoemaker did not conceive of himself as building on Grice's work, and as in fact severely critical of aspects of Grice's new, and of much that Locke had said. Yet, perhaps ironically, a chief effect of Shoemaker's books was to precipitate an increasingly productive re-examination of their ideas.

2. *Self-Knowledge and Self-Identity*

A main theme in Shoemaker's book is that problems about self-knowledge have led philosophers to misconceptions about self-identity, and about the nature of selves in general. Self-knowledge is that which would be characteristically expressed in sentences containing the word "I." The problems have to do with the asymmetry between such statements and the third person statements which are, in some sense, equivalent to them. One who says "I see a tree," for example, will normally find a tree in his visual field, but not a person looking at one, or at least not himself, the person whose seeing he is reporting. And yet someone else who reports the same episode of vision in the third person by saying, for example, "Jones sees a tree," will have to see a person and identify that person as Jones, as well as seeing a tree. Now one who thinks that in the first instance, one must have seen, or somehow been aware of, or at least inferred the presence of, the tree seer as well as the tree seen, and must have identified the person so perceived or inferred as a person appropriately referred to with "I," is likely to be led to the conception of the self as a non-physical thing, simply because no physical thing seems available to fill this role. Shoemaker finds such conceptions in McTaggart, Russell, and others. But these theories, he thinks, all wrongly assume that in order to be entitled to say "I perceive an X" I must perceive more than an *x*. In fact, says Shoemaker, it is a distinguishing characteristic of first-person experience statements that their being true entitles one to assert them. The problem of identifying the perceiver as "me" does not arise, and so the mysterious thing so identifiable need not be found nor postulated. That there should be this entitlement, Shoemaker accounts for in two ways. First, that such

first-person statements are generally true when made is not contingent, but necessary. Second, it is simply a fact, indeed, a very general fact of the sort it is easy, as Wittgenstein had emphasized, to overlook, that we can teach individuals to use such sentences as "I see a tree" just when they see a tree, and in doing so we need not be and would not be providing them with criteria which they can use to identify themselves.

Similarly, in the case of a statement like "I remember going to the store," or "I broke the window," there are no first-person criteria which one must apply, to determine who is remembering, or went to the store, or broke the window. Now philosophers, like Locke and Grice, who are drawn to the view that personal identity consists of links of memory, may have been led to this view by supposing that we must have criteria of personal identity that we apply in our own cases, and finding nothing but memory that could play this role. Philosophers, like Reid and Butler, who emphasize the special and undefinable nature of personal identity, may have seen that no criterion is applied in our own case, but misinterpreted this to mean that identity is directly observed, and consists in the identity of immaterial substance. In both cases, philosophers have been led away from the view that persons are physical beings by the fact that one need not use a bodily criterion of personal identity in a first-person report of what one did, or remembers doing, in the past.

But if we see that rather than a non-bodily criterion being applied, or a non-bodily fact being observed, it is simply the fact that one is doing the remembering that entitles one to say that it is oneself who went to the store, we shall be free to agree with Shoemaker that identity of body is the fundamental criterion of personal identity.

Although Shoemaker criticizes the memory theorists severely, as being motivated by a mistaken epistemology, and defends bodily identity as the fundamental criterion of personal identity, he does allow that memory is a criterion of personal identity, and one that can conflict with the fundamental criterion. Early in the book he introduces the case of Brownson, a twentieth century version of Locke's cobbler and prince---a case that was to perplex and intrigue philosophers for years to come:

It is now possible to transplant certain organs ... it is at least conceivable ... that a human body could continue to function normally if its brain were replaced by one taken from another human body ... Two men, a Mr. Brown and a Mr. Robinson, had been operated on for brain tumors, and brain extractions had been performed on both of them. At the end of the operations, however, the assistant inadvertently put Brown's brain in Robinson's head, and Robinson's brain in Brown's head. One of these men immediately dies, but the other, the one with Robinson's head and Brown's brain, eventually regains

consciousness. Let us call the latter 'Brownson' ... When asked his name he automatically replies 'Brown.' He recognizes Brown's wife and family ..., and is able to describe in detail events in Brown's life ... of Robinson's past life he evidences no knowledge at all. (Shoemaker, 1963,pp. 23 - 24)

Shoemaker does not say that Brownson is Brown. But he does say that if people did say this, they would not be making a mistake, nor even necessarily deviating from our present criteria, or denying the primacy of the bodily criterion. They might simply be allowing it to be over-ridden by other criteria in some circumstances (p. 247).

At this point, some feel a certain frustration with Shoemaker's conclusions. If Brownson is Brown, or even if that is something we might decide was true without inconsistency, then personal identity is not bodily identity, and, it seems, persons are not simply live humans. What then is personal identity? Here the notion of a criterion of identity, and other notions and modes of argument reflecting a Wittgensteinian merger of epistemological and metaphysical questions, seem to obscure rather than illuminate issues. That memory is a criterion of personal identity means that it could not be discovered not to be good evidence for personal identity (p. 4). But that does not mean that memory is logically necessary or logically sufficient for personal identity. The same goes for bodily identity. So the identity of Brown is not settled: we have a conflict of criteria that usually don't conflict, and it appears we must leave it at that.

But it is not clear why Locke's theory, or Grice's modification of it, could not take hold here. Grice might agree with Shoemaker's main conclusions, but argue that his theory is consistent with them, and partly explains them. Memory is a criterion simply because personal identity consists in links of memory. Shoemaker argues that we cannot apply the memory criterion, or even have a concept of memory, without presuming a stable relation between bodily identity and links of memory. That is why the bodily identity criterion is fundamental. But Grice, it seems, could accept that bodily identity was the fundamental criterion of personal identity, and that the assumption of a close correlation between bodily identity and links of memory is a premise of the whole enterprise of talking about persons, without giving up the claim that personal identity consists in links of memory. Just this strategy was adopted by Antony Quinton in "The Soul" (1962), in which a version of the memory theory is defended.

A key consideration against the memory theory in Shoemaker's book is that since we employ no criterion of identity in first-person reports of past thought and action, we do not employ the memory criterion. But while being misled about this might have played a role in motivating the memory theory, it does not seem to provide a decisive objections,

as Shoemaker was to point out later himself. Grice's view is that if I remember an experience, it is mine. This is not to say that I use the fact that I remember it as a criterion for deciding that it is mine.

In spite of Shoemaker's criticisms of the memory theorists, and his reluctance to unequivocally allow Brownson to be Brown, the overall effect of his book on most philosophers was not to produce the conviction that personal identity is simply bodily identity. In the first place, the example of Brownson takes on a life of its own in the mind of the reader; to many Shoemaker's reluctance to straightforwardly identify Brownson with Brown underestimates the force of his own example. Second, Shoemaker's probing studies of various examples, claims, and positions, while not always proving the conclusions he draws, always impress one with the depth of the problems involved. Third, Shoemaker's point that we typically apply no criterion in first-person judgments about the past has seemed a point in favor of the memory theory, in spite of his own use of it as a contrary argument. Finally, Shoemaker does allow that memory is a criterion. Locke had part of the truth. Even though memory may not be enough to make Brownson unequivocally Brown, even Shoemaker admits it is enough to prevent him from clearly being Robinson.

3. Dividing Selves and Multiplying Minds

Before publication of Shoemaker's book, Bernard Williams had put forward a clever argument against the memory theory. In "Personal Identity and Individuation" Williams constructs the case of Charles, a twentieth century man who shows every sign of remembering the actions and experiences of Guy Fawkes:

"Not only do all Charles' memory-claims that can be checked fit the pattern of Fawkes' life as known to historians, but others that cannot be checked are plausible, provide explanations, and so on." (Williams, 1957, p. 9)

The case is designed to give us all the evidence we might want to say we have a case like that of Locke's cobbler. But, Williams points out, we are not forced to say that Charles remembers what Fawkes did, rather than merely that he claims to do so. And he comes up with an impressive argument to clinch the point:

If it is logically possible that Charles should undergo the changes described, then it is logically possible that some other man should simultaneously undergo the same changes, e.g., that both Charles and his brother Robert should be found in this condition. What should we say in this case? They cannot both be Guy Fawkes, if they were, Guy Fawkes would be in two places at once, which is absurd. Moreover, if they were both

identical with Guy Fawkes, they would be identical with each other, which is also absurd.... We might instead say that one of them was identical with Guy Fawkes ... but this would be an utterly vacuous maneuver, since there would be ex hypothesis no principles determining which description to apply to which. So it would be best, if anything, to say that both had mysteriously become like Guy Fawkes.... If this would be the best description of each of the two, why would it not be the best description of Charles if Charles alone were changed? (Williams, 1957, p. 9)

In a reply to an article of Robert Coburn's (1960), Williams makes the principle behind this argument explicit.

"The principle of my argument is ... that identity is a one-one relation, and that no principle can be a criterion of identity for things of type T if it relies only on what is logically a one-many or many-many relation ... 'being disposed to make sincere memory claims which exactly fit the life of ...' is not a one-one, but a many-one relation...." (Williams, 1960, p. 91)

Williams' "Reduplication Argument" provided an interesting challenge to those who wished to defend some version of the memory theory. But it also stirred interest in more general problems of identity and individuation, problems on which attention was also focused as a result of Peter Geach's provocative writings on identity (Geach, 1962, 1969). David Wiggins, in his pioneering study, *Identity and Spatio-Temporal Continuity* (1971), adopts a condition very much like Williams' requirement that a criterion of identity be one-one:

If *f* is a substance concept for *a* then coincidence under *f* must be a determinate notion, clear and decisive enough to exclude this situation: *a* is traced under *f* and counts as coinciding with *b* under *f*, and *a* is traced under *f* and counts as coinciding with *c* under *f* while nevertheless *b* does not coincide under *f* with *c*. (Wiggins, 1967, p. 38)

Wiggins distinguishes the concept of a person from that of a human body, and a person from his or her body. And he incorporates into his notion of a person the memory criterion of personal identity. But he claims, in opposition to Shoemaker's analysis of the Brownson case,

"... that no correct spatio-temporal criterion of personal identity can conflict with any correct memory criterion or character-continuity of personal identity."

Wiggins then accepts both the importance of memory and Williams' condition on criteria of identity. How can he escape the reduplication argument?

In Williams' notion of the memory "criterion" there is again a merging of epistemological and metaphysical considerations. Surely it is not *claiming* to remember that Locke or Grice thought constituted identity, but some relation of memory for which such claims are evidence. Wiggins is quite clear-headed about this. He generally keeps questions of what constitutes identity and questions of how it is known clearly distinguished. In the present case this is manifested by his adoption of a causal theory of memory, adopted from Martin and Deutscher (1966). Note it is much easier to imagine two persons sincerely claiming to have done what one person did in the past than it is to imagine two persons whose claims are both caused by the previous action in the way appropriate to be memories. The memory criterion, interpreted as a causal criterion, is much more plausibly one-one.

Even if this were enough to avoid the reduplication argument, it would not vindicate Wiggins' claim that the memory criterion, properly conceived, cannot conflict with the spatio-temporal continuity requirement. For in the Brownson case the causal requirement appears to have been satisfied.

Wiggins says that they cannot conflict because when the memory criterion is properly founded in the notion of causation the two criteria inform and regulate one another reciprocally

... indeed they are really aspects of a single criterion. For the requirement of spatio-temporal continuity is quite empty until we say continuity under what concept ... and we cannot specify the right concept without mention of the behavior, characteristic functioning, and capacities of a person, including the capacity to remember some sufficient amount of his past. (Wiggins, 1967, p. 46.)

In the final analysis, Wiggins says, we should

...analyze person in such a way that coincidence under the concept person logically requires the continuance in one organized parcel of all that was causally sufficient and causally necessary to the continuance and characteristic functioning, no autonomously sufficient part achieving autonomous and functionally separate existence." (Wiggins, 1967, p. 55.)

Thus, as I understand it, Wiggins allows that Brownson is Brown, the brain being the "organized parcel."

Wiggins points out, however, that building causality into the memory criterion does not totally preclude the reduplication problem:

Suppose we split Brown's brain and house the two halves in different bodies ... there is memory and character and life in both brain transplants .... In this case we cannot simply disregard their (claimed) memories. For we understand far too well why they have these memories. On the other hand, if we say each is the same person as Brown, we shall have to say Brown I is the same person as Brown II. (Wiggins, 1967, p. 53.)

Wiggins reasons that we cannot take both Brown I and Brown II to be the same person as Brown, for they are not the same as each other. And he reasons that even if half of the brain is destroyed, and the other half transplanted, we do not have identity:

"... one of the constraints which should act on us here is the likeness of what happens to the surviving half in this case to what happens in the unallowable double transplant case ...." (p. 56)

Wiggins here agrees with the key move in Williams' original argument. If there were two survivors, we could not say they both were the original. But both would have just the relation to the original that a sole survivor would have. So the relation the sole survivor would have cannot be identity, or enough to guarantee identity. Now one might criticize this by pointing out that the relation differs, in the latter case, in that there is no competitor. And Shoemaker, in an article to be discussed below, does take just this attitude towards the reduplication case: causally based memory without competition is sufficient for identity. It is natural to reply, on Wiggins' behalf, that this added element of lack of competition does not seem the right sort of difference. Why should who I am be determined by what is going on elsewhere in the world---the presence or absence of a competitor to the identity of the person whose thoughts and actions I remember? This line of thinking will lead us naturally to the insistence not only that the criterion or principle of identity for persons (and perhaps for anything) be logically one-one, but that it be, in some sense intrinsically so than as a result of an ad hoc stipulation that competitors defeat identity. Here, however, there is a problem. It is not clear that there are any such intrinsically one-one empirical relations. As Richard Gale (1969) points out, it is not even clear that Williams' favored criterion of bodily continuity is logically and intrinsically one-one. Can we not imagine a situation in which there are two bodies, either of which by itself would be clearly reckoned as a later stage of a given body?

Another alternative is to allow both of the survivors to be the original. This is assumed to be incoherent, due to the logical properties of identity, by Williams, Wiggins, and Shoemaker. In "Can the Self Divide" (Perry, 1970), however, I argued that if we were careful we could allow this, without incoherence---or that at least we could say

everything we wanted to say, giving each of the survivors full credit for the past of the original.

Like Wiggins', my views were set within a general approach to individuation developed as a response to Geach's thesis of relative identity, the thesis that there is no such thing as identity, but only different kinds of "relative identity," and that objects can be identical in one of these ways and not in another. (Geach, 1962, 1969.) My point of view was derived from Frege (1994) and Quine (1963), and emphasized the distinction between identity, a relation that is a part of logic and which every object or entity of any kind or type has to itself, and various relations, unity relations, which were closely related to identity but which were different for various kinds and types of objects.

The undeniable phenomenon motivating doctrines of relative identity is the relativity of individuation. Imagine a checkerboard. We can think of it as eight rows, as eight columns, as 64 squares, or in a variety of other ways. That is, we can individuate it, break it up into individuals for the purpose of description, in different ways. To these different ways of thinking of the same hunk of reality, there seem to be different relations that correspond to identity. Imagine pointing to a checkerboard, saying "this is the same as that." If one is thinking of it as rows, the sentence will only be true if the pointings are side by side, the same distance from the bottom of the board (roughly). If we are thinking in terms of columns, one pointing must be above the other. This suggests that identity is a different relation, depending on whether we are talking about rows or columns. It appears that we need to distinguish between row identity and column identity. This crude example captures one motivation people have had for accepting Geach's doctrine, indeed, the conclusion seems almost forced upon us. If I point at the very same places, side by side, saying first, "This is the same row as that," and next, "This is the same column as that," what I say first will be true and what I say second will be false. So the relations asserted to obtain between the identified individuals must, it seems, be different. (Geach has more sophisticated arguments, of course.)

If we think of rows and columns as sums of squares grouped according to different relations---being above and being beside--- and consistently follow through on this, all of these difficulties, and the motivation for relative identity, will disappear. Being beside and being above are not two kinds of identity, but relations between squares used to construct two different kinds of objects. The problem with "This is the same as that" is not that it hasn't been said what sort of identity is at stake, but that the objects referred to have not been fully identified. And "This is the same row as that " and "This is the same column as that" do not assert different relations of the same objects, but the same relation,

identity, of different objects, a row in the first case, and two columns in the second. Thus, like Wiggins, I was unconvinced by Geach's doctrine. In many ways Wiggins' view is that of well-behaved relative identity, however; he does not emphasize, and in some cases (where temporal parts are needed to make the distinction) seems not to allow, the distinction between the unity relations and identity. (See Shoemaker, 1970.)

This distinction, however, allows us to see a fundamental flaw in Williams' reduplication argument. Williams claims that any criterion of identity must be logically one-one. Now it seems perfectly clear that the evidential relations we have for identity need not be one-one. If *x* looks exactly like you, that is good evidence that *x* is you, not because of logic, but because of the rarity of what are called "identical twins." What Williams has in mind are clearly the relations that are constitutive of identity, the relation that parts have if they are parts of the same person (as I would put it). That is, his principle is that unity relations must be one-one, because identity is.

But in fact unity relations need not, and often do not, share the logical properties of identity. It is more convenient to think about this in terms of the traditional conception of identity as an equivalence relation (reflexive, symmetrical, and transitive), for the notion of one-one becomes awkward when comparing relations between parts (the unity relations) to relations between wholes (identity). Now many unity relations are not, even as a matter of fact, much less as a matter of logic, equivalence relations. In general, where K's are a certain kind of entity with spatio-temporal parts, the formal "*x* and *y* are parts of a single K" gives us the unanalyzed unity relation for K's. Now consider, for example, highways. The roadbed of the Golden Gate Bridge and the portion of U.S. 101 that goes by Candlestick Park are parts of a single highway, as are the roadbed of the Golden Gate Bridge and that part of California Highway 1 that goes by the Pillar Point Fishing Pier. But the part of 101 that goes by Candlestick and the part of 1 that goes by Pillar Point are not parts of a single highway. (The situation is the familiar one of different highways merging to cross an expensive bridge.) Now it might seem that if the unity relation for highways is not an equivalence relation, then highway identity must not be either a counterexample to the one should provide a counterexample to the other. But one finds that an attempt to produce a counterexample is blocked by a failure of reference. "The highway that crosses the Golden Gate Bridge," or "This highway" as said on the bridge, fail to refer, for there are two highways that cross the bridge. Thus mechanisms of reference act as fuses, which by failing keep the logical shortcomings of unity relations from being passed on to identity.

In "Can the Self Divide," (Essay 3) this idea was worked out in some detail, in a way that allowed us to say, without contradiction, and without abandoning any of the traditional properties of identity, that each of the survivors of a reduplication case did all of the things the original had done, and that he was to do all of the things each of them did. The abstract point that unity relations need not share the logical properties of identity has been more convincing than the particular solution proposed, however. Criticisms by David Lewis (1976) and Terence Leichti (1975) have weakened my faith that my intuitions about what to say in a reduplication case were as inevitably the product of careful reflection, and that my scheme embodied them in so completely an unobjectionable way, as I had thought. I would now prefer to speak of "individuative crises" occasioned when unity relations that have been reliably equivalence relations (though not logically) cease to have that character, to which we can respond in a number of ways, the present concept underdetermining the matter (See Essay 4).

"Can the Self Divide" was one of three papers (Essays 3, 5, and 8) in which I defended Grice's memory theory. It seemed to me that Grice had been clearer about the structure of identity than his successors, and that since a careful distinction between identity and unity was built into his account, the reduplication argument did not touch it. This still seems to me correct, even if we adopt the view that what to say in a case of reduplication is left indeterminate by our concept of personal identity rather than being as intuitively clear as I had supposed. In the later papers, I argued that Grice's point of view, when stripped of its goal of logical construction, leads to a plausible causal theory of personal identity, and that an account can be given, within this framework, of the importance of personal identity. In my thinking on each of these matters the distinction between the unity relations and identity loomed large; I thought it was a necessary first step to clarity on these issues. In reviewing the literature for this article, I find my earlier attitude rather unfounded, and think it must have led me to be insufficiently appreciative of others, particularly Shoemaker and Wiggins, who manage to make pretty much the same points without explicitly appealing to the distinction. Perhaps emphasis on the distinction between unity and identity is not so necessary a first step as I had thought!

Another approach to the reduplication case is taken by Roderick Chisholm (Chisholm, 1976). Chisholm wondered how we might face the prospect of splitting, like an ameba. He concludes,

There is no possibility whatever that you would be both the person on the right and the person on the left. Moreover, there is a possibility that you would be one or the other of

these two persons. And finally you could be one of those persons and yet have no memory at all of your present existence. (Chisholm, 1976, p. 179)

Chisholm draws on Shoemaker for support. He says he agrees with Shoemaker's contention that first-person psychological statements are not known to be true on the basis of criteria. He thinks a consequence of this is,

... it makes sense to suppose ... that you are in fact the half that goes off to the left and not the one that goes off to the right even though there is no criterion at all by means of which anyone could decide the matter. (Chisholm, 1976, p. 182)

These reflections on reduplication come at the end of an article whose main object is to defend a version of Bishop Butler's claim that there is a "loose and popular" as well as a "strict and philosophical" sense of identity. Personal identity, unlike the identity of ships and carriages and trains and rivers and trees and in general "compositia" or evolving systems thereof, is identity in the strict and philosophical sense. Identity in the loose and popular sense is typically vague, open ended, defeasible and, ultimately, a matter of convention, of how we choose to talk. In puzzling cases, decision by courts or other agencies is appropriate. But none of this is applicable to personal identity, according to Chisholm. He considers Peirce's (1935, p. 355) example of someone who is to be operated upon, without anesthetic, with a drug administered beforehand which wipes out memories during the operation, and one administered after that restores these but leaves no memories of the operation. Chisholm has no doubt that it is the person in question who will feel pain during the operation, but he considers someone---perhaps someone tempted by Grice's theory---who is not so sure. He says it ought to be obvious to such a person that the adoption of a convention, a way of talking, or a practice by a judge or a whole community, cannot in the least affect the question he is worrying about.

In his reply to Chisholm's paper at the Oberlin Colloquium (Shoemaker, 1969), Shoemaker begins to develop a line of thinking which goes significantly beyond his book, and introduces ideas and problems that dominated the study of personal identity for the next decade. "What we need to clarify," he says, "is the nature of that interest we have in personal identity, and in particular that special concern that each of us has for his own future welfare." (Shoemkaker, 1969, p. 117) Shoemaker entertains the idea that it might be appropriate for one who knows he is to undergo fission, to anticipate the experience of both offshoots, while not supposing that he would be identical with either. These themes are developed in an important paper Shoemaker was to publish three years later.

4. Persons and Their Pasts

In "Persons and Their Pasts," published in 1970, Shoemaker gives a much more sensitive and sympathetic treatment of the memory theory than he had in *Self-Knowledge and Self-Identity*. He says that he is defending Locke's view that persons have, in memory, a special access to facts about their own past histories and their own identities, and he is also defending the non-trivial nature of Grice's claim (suitably interpreted) that "one can only remember one's own experiences." This would be trivial if there were some general mode of access to past experiences, our own others', and "remembered" were simply a title for the subset of experiences so known that happened to have been ours. It is nontrivial if memory (of some sort) is an independently specifiable mode of knowing of past experiences and what we mean, or part of what we mean, by calling an experience "ours" is that it is remembered. In that case, the limited access we have in memory would be constitutive of the notion of a single person. Shoemaker considers two criticisms of the memory theory. One is that it is circular, a charge originally made by Butler; Shoemaker had earlier made a version of this criticism himself. The other is the reduplication argument, not in the form in which Williams originally advanced it, but as put forward by Wiggins and Chisholm, with appropriate causal links between the survivors and the original.

Those who charge the memory theory with circularity acknowledge a strong conceptual link between personal identity and memory, but see this as simply the upshot of the fact that personal identity is a logically necessary condition for memory. If it is a part of our concept of memory that one can only remember events one witnessed or participated in, then it is hardly surprising that memory is a sufficient condition for identity with a past witness or participant. But the analysis of personal identity in terms of memory would be circular.

Shoemaker suggests an analysis of memory for the purposes of considering this charge, which goes more or less as follows:

X remembers event *e* if and only if

1) *X* is in a cognitive state *S*;

2) *Y* was aware of *e* when it happened, in virtue of being in cognitive state *S*';

3) Cognitive state *S*' corresponds to *S;*

4) *Y*'s being in *S*' and *X*'s being in *S* are elements in an *M*-type causal chain;

5) X = Y.

The cognitive state mentioned in 1) is intended to be the sort of state one could be in whether remembering or only seeming to; to distinguish apparent from real memory we need the rest of the analysis. Clause 2) captures part of what is called the "previous awareness condition": if one remembers an event, one must have been aware of it at the time it occurred. By using "*Y*" instead of "*X*" in its statement, the part of the analysis that seems to lead to circularity, the condition that the previously aware person be the remembered, is split off for separate consideration. Clause 3) makes the plausible point that what is remembered must correspond to what one perceives or experiences, though exactly what this involves is not explained at any length. Clause 4) requires that the present memory activity be caused by the earlier perceptual activity in the right way, that is, in the way that it is usually caused in memory. Clause 5), finally, is the isolated condition split off from the previous awareness condition, the element in the analysis that makes the use of memory to analyze identity seem circular.

Shoemaker then introduces two new notions. *x* quasi-remembers *e* if conditions 1) -- 3) are satisfied. *x* quasi-remember *e* if conditions 1) -- 4) are satisfied. Thus the statements that one can only quasi-remember one's own past experiences, or that one can only quasi-remember one's own experiences, would certainly not be trivial. If we find that either of these notions assigns the same past event and experiences to a person as does the "unstripped" notion of memory, then we can say that the additional clauses are really just redundant. If, for example, we find that one quasi remembers just those past events that one would be said to remember, then one can say that clause 5) is really not necessary for the analysis of memory: memory is just quasi-memory. 5) would be true, but now we could look on its truth as a consequence of the nature of memory as given by 1) -- 4) and the non-circular analysis of personal identity in terms of that notion of memory.

Shoemaker concludes that quasi-memory is not a very interesting notion. The strong conceptual link between memory and personal identity is brought out in two principles. The first is the (unstripped) previous awareness condition. The second is what Shoemaker calls preservation of immunity to first person misidentification. This notion is a descendant of the idea in *Self-Knowledge and Self-Identity* that one needs no criterion for first-person identification. Used in the book as a basis for criticism of the memory theory, this immunity is now seen as something the memory theory goes some way towards explaining. Shoemaker brings in the helpful notion of memory "from the inside." When I remember a past thought or action from the inside, then I can identify myself as the past thinker or doer, without identifying the thinker or doer as someone who fits a certain description or satisfies certain criteria. Now, insofar as we can understand quasi-memory

at all, neither the previous awareness condition nor the principle of preservation of immunity to misidentification seem to hold for it.

For quasi-memory, however, the picture is quite different. When we add the causal requirement, we get a notion almost indistinguishable from ordinary memory. Virtually any situation I can imagine in which the conditions for quasi-memory are met is a situation in which the conditions for memory are met. This strongly suggests that personal identity can be analyzed in terms of quasi-memory, that clause 5) can be seen as a consequence of this analysis, and that the memory theory need not be circular.

The need for qualification comes from the second criticism of the memory theory that Shoemaker considers. The way the world is, *M*-type causal chains neither branch nor merge. Given this orderly behavior, quasi-memory seems indistinguishable from memory. But it is imaginable that *M*-type causal chains should not behave in such an orderly way; this is just what we imagine when, with Wiggins, we imagine halves of a brain transplanted to different bodies, or with Chisholm, we imagine splitting like an ameba. In these cases Shoemaker supposes that we

"cannot identify both of the physiological offshoots of a person with the original person, unless we are willing to take the drastic step of giving up Leibniz's Law and the transitivity of identity..." (p. 28)

Given such ill-behaved causal chains, I could quasi-remember from the inside an experience or action that wasn't mine. For this reason, the analysis of personal identity in terms of quasi-memory is not totally straightforward. But we can get a logically sufficient condition for personal identity: quasi-memory with no branching. Basically, Grice and Locke are vindicated.

Towards the end of the article, Shoemaker picks up the question he raised in the reply to Chisholm. In a case in which there has been branching, Shoemaker thinks that neither of the branch persons is identical with the original. But each would quasi-remember the experiences and actions of the original person. Now which of these facts is important? That they do quasi-remember, or that they are not identical? As Shoemaker puts it,

"If I [quasi-remember] from the inside a cruel or deceitful action, am I to be relieved of all tendency to feel remorse if I discover that because of fission someone else quasi-remembers] it too?" (Shoemaker, 1970, p. 284).

Shoemaker thinks not. It is the quasi-memory that is important, not the lack of identity. As against this, we might appeal to such facts as that identity is a necessary condition of responsibility for past actions. But then Shoemaker could simply repeat the identity-

stripping investigation for the concept of responsibility, and argue that the operative concept is really quasi-responsibility. No concept has identity more "built into it" than that of survival. Shoemaker thinks that if one is to fission, one will not be identical with either of the survivors. And yet,

"The prospect of immanent fission might not be appealing, but it seems highly implausible to suppose that the only rational attitude toward it would be that appropriate to the prospect of immanent death..." (p. 284)

The idea emerging here is that personal identity is important to us because it involves certain relationships to past and future persons, rather than these relationships being important because they constitute identity. This idea was to undergo explicit statement and dramatic development in an important article by Derek Parfit published shortly after Shoemaker's. Before looking at that, however, we must look at an article by Bernard Williams (1970) published the same year as "Persons and Their Pasts" (1970), in which, once again, the argument against according memory too much importance in personal identity is made subtly but forcefully.

5. The Self and the Future

In "Persons and Their Pasts," Shoemaker says that Brownson is Brown, and that his former reluctance to conclude this was a result of overlooking the causal component in the notion of memory---an element which, as we have seen, was emphasized by Wiggins, who himself seems to have accepted that Brownson is Brown. Though remaining more certain of the puzzling nature of the questions raised than of his own conclusions, Bernard Williams remained unconvinced, and in "The Self and the Future" (1970) he argued against the possibility of body transfer and the considerations about memory and personal identity that seem to allow it. Since I have discussed this paper at length in Essay 6, I have ommitted the discussion of it that was originally included in this survey.

While I argue in Essay 6 that Williams argument against Shoemaker and the possibility of body shifting is not convincing, his subtle and stubborn argumentation forces to one's attention what might be called the phenomenological difficulties of accepting one's identity as the sort of thing which could be a matter for decision. Chisholm (1969), it will be recalled, found it simply bizarre to imagine that one's identity could be a matter for decision, a matter that would be decided by convention or litigation or even by social practice. Though not drawn, as Chisholm was, by something like a pure ego or immaterial substance theory, and more in a mood to remind us of difficulties than to

establish conclusions, Williams shares Chisholm's attitude towards the suggestion that personal identity could be a matter of convention.

"There seems to be an obstinate bafflement to mirroring in my expectations a situation in which it is conceptually undecidable whether I occur ... The bafflement seems, moreover, to turn to plain absurdity of we move from conceptual undecidability to its close friend and neighbor, conventionalist decision ... as a line to deal with a person's fears or expectations about his own future, it seems to have no sense at all." (Williams, 1970, p. 61.)

While Chisholm was drawn to a metaphysical solution to these problems, Williams seems to think that there is refuge in bodily identity. It seems to me, though, that he has put his finger on something that is not just a baffling consequence of one theory of personal identity, that which emphasizes memory, and is avoided by others, but rather on something that is simply baffling. Let any empirical relation *R* be the candidate for the unity relation for persons. Then some philosopher is clever enough to construct a case in the area of *R*-vagueness, that is, a case where our concepts leave it indeterminate whether *R* obtains or not, even given all the facts. Then we will have an indeterminate case for the theory that maintains that *R* is the unity relation for persons. It may be easier to construct cases for links of memory than for links of body. But it seems that all of our concepts are formed, as Wittgenstein said, within the context of certain very general assumed facts; by imagining those facts to be otherwise, we can create cases the concept was not designed to handle.

We have seen, in the course of discussions, a shift of attention from persons and their pasts to persons and their futures, a shift called for by Shoemaker in his reply to Chisholm (1969), and initiated in the thoughts at the end of "Persons and Their Pasts" (1970). A key concept in such a enquiry is that which Williams calls "the imaginative projection of myself as participant in [a future situation] ." (Williams, 1970, p. 59; see also Williams, 1973) We can perhaps think of this as the future-oriented analogue of Shoemaker's "memory from the inside." Shoemaker's question, whether one should look forward to fission as death, given his belief that one will be identical with neither of the products of fission, is then a proposal to consider imaginative projection of ourselves as participants, when we realize that no participant in the situation will be identical with us. Made sensitive by Williams of the baffling aspects of such questions and proposals, let us return to them.

6. Survival Without Identity

If one is asked why one feels bad about an event of the previous evening, and responds, "Because I am the one that committed the outrage," the identity asserted between the present speaker and the participant in the earlier event seems to be bearing an important explanatory role. But in "Persons and Their Pasts," Shoemaker is on the verge of displacing identity from this explanatory role, putting in its place the "identity-stripped" concepts of quasi-memory, quasi-fear, quasi-responsibility, and the like. The importance of identity derives from the importance of these relations, which in our well-behaved world, with no *M*-fission or *M*-fusion, can be taken as constitutive of identity. The suggestion that identity is after all not so crucial is also considered by Terence Penelhum, in *Survival and Disembodied Existence* (1970), with special reference to what we really want when we hope for survival after death. But the step of pushing identity to the background was made most boldly and unequivocally by Derek Parfit, in his profound, imaginative, and influential article, "Personal Identity":

Judgments of personal identity have great importance. What gives them their importance is the fact that they imply psychological continuity.... If psychological continuity took a branching form, no coherent set of judgments of identity could correspond to, and thus be used to imply, the branching form of this relation. But what we ought to do ... is take the importance which would attach to a judgment of identity and attach this importance directly to each limb of the branching relation. ...judgments of personal identity derive their importance from the fact that they imply psychological continuity.(Parfit, 1971)

Parfit thinks that there are cases in which there is no correct answer to a question about personal identity. He refers to the examples of Locke (1694, section 18) , Prior (1966) Bennett (1967) , and Chisholm, but in particular to that of Wiggins. "My brain is divided, and each half housed in a new body. Both resulting people have my character and apparent memories of my life. What happens to me? (Parfit, 1971, p. 5) To say he does not survive seems odd, Parfit argues: "How could a double success be a failure?" For him to be only one or other seems arbitrary. But to say he survives as both is to violate the laws of identity, Parfit assumes. His solution is that we do not need to have identity to have survival, or at least not to have what is important in survival: "We can solve this problem only by taking these important questions and prizing them apart from the question about identity." (Parfit, 1971, p. 9) When we do this the results are dramatic. While identity is an all or nothing affair, the various identity stripped relations that constitute it when well-behaved, and are what really matters in any case, are often quite

plausibly regarded as matters of degree. This is a matter of importance, not only in analytical metaphysics, but in the way we think of ourselves in real life:

"Identity is all-or-nothing. Most of the relations which matter in survival are, in fact relations of degree. If we ignore this, we shall be led into quite ill-grounded attitudes and beliefs." (Parfit, 1971, p. 11)

Among these are the principles of self interest, and regrets about one's eventual death. He argues:

"Suppose that a man does not care what happens to him, say, in the more distant future... We must say, 'Even if you don't care, you ought to take what happens to you then equally into account.' But for this, as a special claim, there seem to me no good arguments... The argument for this can only be that all parts of the future are equally parts of his future. But it is a truth too superficial to bear the weight of the argument." (Parfit, 1971, p. 26)

Parfit notes that in certain extreme puzzle cases---a network of "persons" who periodically fission and fuse, for example---we are naturally led to think not in terms of continuous persons, but in terms of more or less connected selves, reserving the word 'I' for the greatest degree of psychological connectedness. This way of thinking could be applied even in normal cases, and would embody a recognition that it is psychological connectedness that is what matters, and this would help in avoiding the ill-grounded attitudes and beliefs mentioned above. In "Later Selves and Moral Principles" (1976), Parfit argues that thinking in this way, or recognizing the possibility of doing so, undercuts certain arguments against utilitarianism.

In Parfit, we might say, Shoemaker's analytical tool of identity stripping has become an approach to life.

The question of the importance of identity seems to me greatly illuminated by general questions about identity and individuation, and in particular by the perspective sketched in the Essay 4 above. Indeed, as soon as one adopts the perspective that identity is a logical relation, one is implicitly committed to the derivative importance of identity, although not necessarily to Parfit's claim that what matters are relations of degree.

After all, there are many conceivable ways of individuating the world---of choosing unity relations with which to unify our objects. That is, many are conceivable from the point of view of constraints imposed by logic, although most fanciful alternatives would not be possible ways for beings like ourselves (however individuated) to experience or deal with the world. Each way of individuating gives rise to a class of objects, members

of which are identical to themselves in as literal and unsullied a sense as I am to myself. Thus, for example, we could think in terms of a kind of object which consisted, during any baseball game or inning thereof in which the San Francisco Giants participate, of the Giants' shortstop for that period. This would be a discontinuous object composed of stages of ordinary men, stages of Le Master and this season (1982). We could give rules for referring to and assigning predicates to these objects, adjusting things to preserve the indiscernibility of the identical. Let us call such entities longstops. Then the longstop in the game gets an error or strikes out just in case the shortstop does. But the present longstop may have struck out in the last inning, even if the present shortstop didn't---if Metzger replaced a slumping Le Master, for example.

Last inning's longstop is identical with this inning's longstop in as pure a sense of identity as anything is identical with itself. But the identity is unimportant. That the longstop was injured last inning, and that the very same longstop is now playing, gives us no reason to expect limping. That the longstop who is playing now made a good play last inning gives us no reason to cheer when he comes to bat.

Clearly, the importance of the identity of objects of a given kind depends on the unity relation. The choice of a unity relation to be a part of our scheme, and so the presence of objects of the corresponding kind in the scheme, reflects its importance. The importance of identity is in this sense derivative; how could it be otherwise?

But we can also ask why a given unity relation is important--- worth fashioning identity out of. In particular, if memory, or some more general kind of psychological continuity or connectedness is important, and the source of the importance of personal identity, why is this so? What is so important about it? I think this question is the one which if often on the mind of philosophers who resist the idea that personal identity is analyzable at all. For we can make the point about the derivative importance of identity from an even more general principle. If, as Locke supposed, personal identity may be analyzed, must not the analysans explain the importance of the analysandum? The idea that any such explanation of the special importance identity has for us must be absurd, leads to the claim that identity is unanalyzable and primitive. Butler (1736), for example, thinks that if personal identity is analyzable, then it is not strict identity after all but something else, and that if this were so it would be

"... a fallacy upon ourselves to charge our present selves with anything we did, or to imagine our present selves interested in anything which befell us yesterday, or that our present self will be interested in what will befall us tomorrow."

While I think that there is no distinction to be drawn between strict and loose identity of the sort that Butler imagines, if we hold that personal identity is analyzable, it seems his challenge must be met. Of Parfit's analysis we might ask: "Why is it important, and why do we care in a special way, about what will happen to someone tomorrow who is psychologically directly connected with me?" Now it is no longer open to us to say the most natural thing, that it is because psychological connectedness is sufficient for identity, and so he will be me if he is so connected with me. We have concluded that such an appeal to identity is not ultimate, but gives way to the explanation in terms of connectedness.

I tried to deal with this problem in "The Importance of Being Identical," (Essay 8) which appeared in Amelie Rorty's anthology *The Identities of Persons* (1976). The attempt led to conclusions which I found peculiar at the time, and still find peculiar, but of which I reconvince myself each time I reflect upon the matter. It seemed clear that a theory of personal identity should be causal; I adopted a descendant of the memory theory that fully relied on the fact that memory is a causal notion. Now in general, attempts to explicate concepts in causal terms make reference to the normal mode of causation. It is not enough for me to remember a past event that the event have caused my present memory impression; it must have done so in the right way. If I spill soup on my grandmother as a child, and am told of it so often as an adolescent that as an adult I have clear memory-like impression, then my spilling has caused the impression. But I do not remember, for it was not caused in the right way.

The account of why identity should be important was built around the fact that we know what to expect from ourselves in the normal case, and can expect continued commitment to the values we have. But it is hard to see why an atypical causal chain that provided the same guarantees should not be just as good, even though, as seemed and seems clear to me, if it is atypical enough it doesn't provide identity. I came to the conclusion that it shouldn't matter:

Suppose the following. A team of scientists develops a procedure whereby, given about a month's worth of interviews and tests, the use of a huge computer, a few selected particles of tissue, and a little time, they can produce a human as like any given human as desired... I have an incurable disease. It is proposed ... that a duplicate be created ... and simply take over my life... He would not be me. The relation between my terminal and his initial states is too unlike the [normal causal relations which preserve psychological continuity between earlier and later stages of humans] to be counted, even given the vagueness of the concept of a person, as an instance of it. But ... I would have

the very same legitimate reasons to act now so as to secure for him future benefits as I would have if he were me. (p. 83)

I meant to include by this the full appropriateness of "imaginatively projecting myself" into the benign imposter's future experiences. Such a position still seems to me a natural and inevitable outcome of Locke's original idea.