

## PERSONAL IDENTITY AND THE PAST

---

MARYA SCHECHTMAN



**ABSTRACT:** In the second edition of the *Essay Concerning Human Understanding*, John Locke argues that personal identity over time consists in sameness of consciousness rather than the persistence of any substance, material or immaterial. Something about this view is very compelling, but as it stands it is too vague and problematic to provide a viable account of personal identity. Contemporary “psychological continuity theorists” have tried to amend Locke’s view to capture his insights and avoid his difficulties. This paper argues that the standard approach fails because it takes Locke to be a memory theorist, and does not focus enough on his claim that we need continuity of consciousness for personal persistence. An alternative reading of Locke is offered, emphasizing the role of self-understanding in producing continuity of consciousness. This alternative overcomes the difficulties with the standard approach, and shows how it is possible to attribute unconscious psychological elements to a person, even when personal persistence is defined in terms of consciousness.

**KEYWORDS:** personal identity, consciousness, memory, unconscious, self understanding.

**I**N THE SECOND EDITION OF THE *Essay Concerning Human Understanding*, John Locke takes up the question of what makes someone the same person throughout her entire life. His response to this question has served as the starting point for many of the views of personal identity represented in the philosophical literature today. Locke’s important contribution is to argue that the continuation of a person is independent of

the continuation of any substance—either physical (the body) or nonphysical (the soul). I am the same person as someone who existed in the past, says Locke, if and only if I can extend my current consciousness back to that person’s actions. This assertion is usually interpreted as a “memory theory” of personal identity—the view that whatever actions and experiences a person can remember are, for that reason, her actions and experiences.

In some respects, Locke’s view is extremely compelling, but at the same time a simple memory theory is totally implausible. Although Locke’s arguments that continuation of substance cannot serve as a viable account of personal identity are powerful, a view that implies that a person can have no experiences that he does not (or cannot easily) remember consciously seems far too strong. Locke might or might not be willing to bite the bullet and accept that no forgotten experiences can be ours, but most philosophers are not. The contemporary theorists who base their views on Locke’s insight (“psychological continuity theorists”) have thus altered his original account, trying to keep the basic insight while avoiding the counterintuitive implications. This has met with mixed success. Psychological continuity theories do fix some of the obvious difficulties with Locke’s original view, but they do so at a cost. These amended views lose much of the appeal of Locke’s original picture and undermine much of the original argument for a psychological account of identity.

In what follows, I argue that this cost is incurred because psychological continuity theories take a wrong turn in developing Locke's view. They concentrate too much on the notion of memory, and not enough on the notion of consciousness. As a result, they end up with views that neither fully capture Locke's original insight nor fully avoid the implausibilities of his view. I propose an alternative development of Locke's insight that emphasizes the importance of self-understanding. This view captures most of what Locke says about why consciousness is so central to personhood and personal identity without being committed to the implausible view that only experiences of which we are conscious can be ours. I begin with a brief review of Locke's view, the objections to it, and the development of psychological continuity theories to answer those objections. Next, I describe some of the deficiencies of psychological continuity theories, and outline my proposed alternative.

## LOCKE'S ARGUMENT

Locke's central contribution to work on personal identity is his insistence that identity must be defined in terms of sameness of consciousness rather than sameness of substance. It is not the continuation of either an immaterial soul or a body that constitutes the continuation of the person, he says, but rather the continued flow of consciousness. There are two basic elements of his argument for this claim. To make the claim intuitively plausible, he uses a number of hypothetical cases in which continuation of consciousness is separated from continuation of substance, showing that our judgment in such cases would be that the person goes with the consciousness. He also provides a more theoretical discussion about what it is to be a person, arguing that once we understand this clearly we see that identity must be defined in terms of consciousness rather than substance. It is helpful to review briefly each aspect of his discussion.

Locke gets us to see the force of his view through the use of a series of imagined cases. He asks us to imagine, for instance, the mental life of a prince "entering and informing" the body of

a cobbler, and argues that everyone would see that the resulting person is the same person as the prince rather than the cobbler (Locke 1979, 340). He suggests that we imagine someone who has the same soul as Nestor or Thersites at the siege of Troy, but without consciousness of any of their actions, and tells us that it is obvious that this person is no more the same person as Nestor or Thersites than he would be if his body happened to share some of the same matter that had once composed theirs (Locke 1979, 339). He asks us also to imagine a man who has two distinct consciousnesses sharing his body—one by day and one by night—with no communication between them, and says that it is clear that there are two distinct persons sharing one body in such a case (Locke 1979, 344–345).

This view of personal continuation is not, of course, uncontroversial. It is, however, widely accepted, and whether or not this is the final word on what it is to be a person, it undoubtedly captures one important strand of our thought about ourselves. The same basic intuition is expressed, for instance, by William James in *The Principles of Psychology* when he says,

The Soul, however, when closely scrutinized, guarantees no immortality of a sort *we care for*. The enjoyment of the atom-like simplicity of their substance *in sécular séculorum* would not to most people seem a consummation devoutly to be wished. The substance must give rise to a stream of consciousness continuous with the present stream, in order to arouse our hope, but of this the mere persistence of the substance *per se* offers no guarantee. (James 1950, 348)

This idea can also be seen when we recognize that there is a real sense in which we would view total, irreversible amnesia as a form of death. Faced with the prospect of such amnesia we might well distribute remembrances, write letters to loved ones, and in other ways act as if we were anticipating death.

The Lockean insight is also seen in the impulse to view Multiple Personality Disorder as a circumstance in which more than one person share a body. Of course, describing this disorder in this way is by no means uncontroversial, and later I examine some of the reasons we might

want to say that there is only one person per body, even in cases such as this. The point here, however, is to get clearer on the intuition behind the Lockean view, not to argue that there is a monolithic problem of personal identity to which it provides the unique solution. It is, I think, at least uncontroversial to claim that in Multiple Personality Disorder we are not presented unproblematically with a single person. The fact that these cases seem to present, at least sometimes, genuinely independent streams of consciousness, which may have no awareness of one another, seems reason enough to say that there is some very important sense in which distinct persons co-occupy a body just as Locke took his revolving day and night consciousness to do.

This, then, is the basic intuition behind Locke's view, and it is in many respects a familiar and compelling one. Locke further supports his account of identity by giving a more general definition of the concept of the person revealed in these cases. He tells us that *person* stands for "a thinking, intelligent Being, that has reason and reflection, and can consider it self as it self, the same thinking thing in different times and places" (Locke 1979, 335). To be a person is to have self-consciousness, viewing oneself as a persisting subject. This means that a person becomes the same person who has past experiences or undertook past actions when her consciousness includes them. "As far as any Intelligent Being can repeat the *Idea* of any past action with the same consciousness it had of it at first, and with the same consciousness it has of any present Action;" says Locke, "so far it is the same *personal self*" (Locke 1979, 336). His view in a nutshell, then, is that "*Personal Identity* consists, not in the Identity of Substance, but, as I have said, in the Identity of *consciousness*" (Locke 1979, 342).

Locke expands on this concept of person by telling us that *person* is a "Forensick Term appropriating Actions and their Merit; and so belongs only to intelligent Agents capable of a Law, and Happiness and Misery" (Locke 1979, 346). This characterization sets out two features that Locke takes to be unique to personhood, and that he uses to support his definition of personal

identity in terms of sameness of consciousness. First, persons are capable of a special sort of self-interested concern and second, they are moral agents, capable of taking actions for which they can be held responsible. Both of these capacities, he argues, are linked to consciousness. To care about the unfolding of our lives, we need to have a conscious conception of ourselves as having a future to care about. Moreover, it is through consciousness that we experience pleasure and pain, and so the extent of our concern is the extent of our conscious experience. He puts it this way: "*Self* is that conscious thinking thing, (whatever Substance, made up of whether Spiritual, or Material, Simple or Compounded, it matters not) which is sensible, or conscious of Pleasure and Pain, capable of Happiness or Misery, and so is concern'd for it *self* as far as that consciousness extends" (Locke 1979, 341). He also argues that moral responsibility requires consciousness. To be moral agents, we must be able to plan and to recognize that actions we take now have consequences in the future. Here he taps the intuitions unearthed by the kinds of hypothetical cases described above. He says, for instance, "for suppose a Man punish'd now, for what he had done in another Life, whereof he could be made to have no consciousness at all, what difference is there between that Punishment, and being created miserable?" (Locke 1979, 347). Again the case of Multiple Personality Disorder gives us a real-life example of what Locke has in mind here. Such cases have raised tricky questions in legal contexts about how to assign culpability when the consciousness of the personality who allegedly committed a crime is unavailable to other personalities in the same body.

Locke thus taps into a widely held concept of the person according to which a person is a self-conscious subject. On this picture, personal identity or continuation through time depends on the continuation of consciousness. In particular, if one is conscious of oneself as the same self who existed at some past time, this consciousness actually makes one the same person as that past self. The failure of such self-consciousness, on the other hand, signals the end of the person. Attempting to capture this conception of the

person, Locke thus tells us that we become the same person as some past person by extending our consciousness back to her experiences because in doing so we become aware not only of those experiences themselves, but of the fact that we are the ones who had them. This awareness gives us consciousness of ourselves as persisting beings, and so constitutes the fact of our identity with the past person—a fact that would not hold were we not conscious of her experiences.

Although there is undoubtedly something very compelling about Locke's view of personal identity, it is problematic as it stands. It is not entirely obvious just what the details of the view are, but it is generally read as a "memory theory"—the view that whatever experiences a person remembers are, for that reason, her experiences. Locke tells us that a person makes past experiences hers by extending her consciousness back in time to them, but as Thomas Reid suggests, "Mr. Locke attributes to consciousness the conviction we have of our past actions, as if a man may now be conscious of what he did twenty years ago. It is impossible to understand the meaning of this, unless by consciousness be meant memory, the only faculty by which we have an immediate knowledge of our past actions" (Reid 1976, 115). Although I have argued elsewhere that this is an oversimplified reading of Locke (Schechtman 1996, 105–112), it seems obvious that memory must be a large part of the picture.

A simple memory theory, however, is not plausible on its own. It is at the same time too weak and too strong. This theory is too weak because it seems that it takes more than simply remembering an experience to make me the person who did it. As many objectors have pointed out, memory is good evidence that a past experience is ours, but on its own it does not seem quite enough to make it ours. To use a variation on a science fiction case that appears often in the literature, the fact that a neurosurgeon may develop a technique whereby she could implant in my brain the recollection of an experience had by her grandmother does not now make me the person who had that experience (i.e., her grandmother). The memory theory is too strong because even when we are thinking of ourselves as

conscious subjects rather than substances, it seems obvious that we can and do forget experiences that are nonetheless ours. If the mind of the prince were, for example, to enter the body of the cobbler but along the way lose the memory of what the prince ate for breakfast or repress the memory of a rather unfortunate interaction with the vice chancellor, we would not want to deny that these were, nonetheless, experiences of the prince and so of the person inhabiting the cobbler's body. To provide a plausible account of identity, then, the simple memory theory needs some development.<sup>1</sup> This is the task undertaken by psychological continuity theorists.

## PSYCHOLOGICAL CONTINUITY THEORIES

Psychological continuity theorists take on the goal of providing an account of personal identity over time that captures what seems so right in Locke's observations but is more plausible in the details. There have been several strategies for dealing with the implausible aspects of the simple memory theory; I review some of the most fundamental. First, consider the objection that merely remembering some past experience does not seem enough to make me the person who had it. One response to this difficulty is to require that there be more than one memory connection between a present and past person if we are to judge that they are the same person. In the example where a neurosurgeon implants one memory from her grandmother's life in someone's brain, a large part of the reason we are unwilling to make a judgment of identity is that there is only a single, out-of-context connection between the post-operative person and the pre-surgery grandmother. Psychological continuity theorists thus generally require that there be "enough" connections between people at two times if we are to say they are the same person. There is, of course, a great deal of difficulty in individuating connections, let alone in finding a nonarbitrary number of connections to determine identity. The general idea, however, is that there must be some threshold level of memory connection between a present person and a past

one that is required to make them the same person.

Even if a person remembers a great many experiences of a past person, however, it may seem that the relation of memory is too weak to support identity—at least if *memory* is taken to mean no more than having an experience of recollection. For this reason, it has been very common for psychological continuity theorists to require that for a past experience to be attributed to a person, he must not only recall the experience, but the recollection must be properly caused by the original experience. The idea of proper cause is generally spelled out in terms of the continued functioning of a single brain in which a memory trace is laid down at the time of the original experience.

To address the sense that memory theories are too strong, a different but related set of changes is made to the simple memory theory. A first step in overcoming this difficulty is recognizing that there is a fundamental difference between the types of cases Locke amasses to make the case that memory is crucial to personal identity and the kinds of cases marshaled to show that the memory theory seems so implausible. The cases Locke describes all involve a person experiencing wholesale memory loss, and so having her conscious connection to the past severed entirely. The cases where it seemed obvious that a lapse of memory does not result in a change of person were cases in which only a few memories were lost, but others stayed in place, providing a connection to the past time, if not to those particular past events. Once again, then, the issue is how many memory connections to a past time are present. In this context, however, the significance of the number of connections suggests that there are really two distinct questions at issue here: one is the question of whether a person in the present is the same person as some person in the past (or whether a particular person survives into the future). The second is the question of whether some particular past action or experience is attributable to a person in the present. What our discussion seems to show is that the intuition that memory is crucial to issues of identity applies most clearly to the first question.

Locke's general claim seems plausible when we take him to be saying that, for a person to continue into the future, there must be someone in the future who remembers some (enough) of her experiences, and therefore takes herself to be that person. It seems less plausible when we take it to say that, for a person to be the person who had some particular experience or undertook some particular action, she must remember having that experience or taking that action.

One natural way of responding to this implausible feature of a simple memory theory is thus to separate the questions of what makes a person at one time the same person as a person at another time from the question of what makes a particular action or experience the action or experience of a given person.<sup>2</sup> It is then possible to offer a view according to which a person at the present time is the same as some person in the past if the present person has enough memories of the past to allow for a conception of herself as a continuing subject, without being committed to the idea that it is memory that makes a particular action or experience attributable to a person. It will then be necessary, of course, to offer some account of attribution as well.

Psychological continuity theorists follow this strategy either implicitly or explicitly, separating these questions, and taking the question of identifying persons at two different times with one another as prior. To answer this question they take as their basic starting point a reading of Locke, which says that a person at time  $t_2$  is the same person as a person at an earlier time  $t_1$  just in case the person at  $t_2$  remembers some of the experiences of the person at time  $t_1$ . They then amend this view to try and overcome any residual implausibility. First they allow that a person need not remember even one experience from every part of his life. There may be some phases of the past—perhaps those very remote in time—that one does not directly remember at all, but that nonetheless seem to be part of one's life. Psychological continuity theorists account for this fact by requiring not direct memory, but overlapping chains of direct memory to make a person at one time the same as a person at another. If I currently have some memories of my 30th

birthday, and if the person whose experiences I remember on that day had some memories of my 20th birthday, and the person whose experiences are remembered on that day had memories of my 10th birthday, and that person had memories of my 6th birthday, then I now am the same person as that person celebrating my 6th birthday, even if I have no direct memories of that day whatsoever. These theorists also suggest that memory is not the only psychological connection that can contribute to identity, but that the persistence of beliefs, values, desires, or the connections between intentions and the later actions that carry them out can also serve as identity-constituting connections between persons at different times.

The final form of the psychological continuity theory's answer to the question of what makes a person at time  $t_2$  the same person as a person at time  $t_1$  is that it is an overlapping chain of psychological connections including memory and other connections between the person at  $t_2$  and the person at  $t_1$ . The answer to the second question—the question of what makes a particular action or experience that of a particular person—is often implicit in psychological continuity theories and rests on the answer to the first. The idea is simple and natural: the criterion outlined tells us whether a person at  $t_2$  is the same as a person at  $t_1$ . If they are the same, then whatever experiences are attributable to the earlier person at  $t_1$  are also attributable to the person at  $t_2$ . If, that is, the cobbler body before us is determined to be the same person who was previously in the prince body, then he is the person who ate breakfast in the prince's body or had that bad interaction with the vice chancellor, whether he remembers it or not.

This strategy, of course, requires that we have some way of attributing experiences to persons at a time. This sort of attribution is usually taken for granted, but the underlying idea is pretty clear. It is assumed that there is a primitive notion of unity of consciousness at a time. At any particular moment, there is some set of experiences experienced as co-conscious, and all of these belong to a single person. If a later person remembers some but not all of the experiences that were co-conscious earlier, she thereby con-

nects himself to the entire earlier person, and makes all of the experiences that were co-conscious at that time hers. This position is in the background in most psychological continuity theories, but is made explicit in an early version of this view put forth by H. P. Grice. Grice defines what he calls a "total temporary state" or "t.t.s." To give an idea of what he means by this, he says that a t.t.s. is "composed of all the experiences any one person is having at any given time. Thus, if I am now thinking of Hitler and feeling a pain, and having no other experiences, there will be occurring now a total temporary state containing as elements a thought of Hitler and a feeling of pain" (Grice 1976, 86). He then defines personal identity over time in terms of overlapping chains of memory connections between t.t.s.'s. The same strategy can also be found in versions of the psychological continuity theory offered by Sydney Shoemaker (1984), John Perry (1976), David Lewis (1983), and Derek Parfit (1984).<sup>3</sup>

Psychological continuity theories thus start with the basic Lockean insight that persons are essentially self-conscious entities, and that their identity over time should be defined in terms of the extension of consciousness. They read this as the claim that personal identity should be defined in terms of memory connection, and alter this view to avoid some of the more obvious objections. They end by offering a view according to which a person at one time is the same person as a person at an earlier time if the present person is connected to the earlier one by overlapping chains of sufficient numbers of psychological connections. When people at two different times have been determined to be the same person, any experiences or actions that are attributable to the one person are also attributable to the other.

## PROBLEMS WITH PSYCHOLOGICAL CONTINUITY THEORIES

Psychological continuity theories present a rather tidy solution to the difficulties described, but in the end they are not that much more satisfying than a straight memory theory. These theories, too, seem both too weak and too strong,

if in a different way from the memory theory itself. They seem too weak because it is not obvious that the relation in terms of which they define identity really does any better than Locke's original view at quieting the sense that memory alone is not enough to constitute identity, and may even do worse. In the finished psychological continuity theory, what we have are a collection of independent "persons-at-times" that are cobbled together through memory and other connections with the requirement that there be some critical mass of memory connections and some (usually physical) causal connection between present memory experiences and past experiences of which they are memories. Although the requirements added to the simple memory theory do in some sense make a stronger connection between the different temporal parts of a person's life and between a person and the experiences that are attributed to her, it is not clear that it is stronger in the right way—that is, that is stronger in a way that will answer the original worries.

First, consider the requirement that there be some critical number of memory connections in place before we make a judgment of personal identity. This amounts to a difference in degree rather than kind from the simple memory theory. What we have here is just a collection of memories, and it is not clear that that is enough where one memory was not. It is not obvious how the addition of connections besides memory could help here either, because these give us even less conscious access to the past. This, of course, is why the requirement of causal connection is added, but it is not evident that that does the work psychological continuity theorists want it to, either. What this amounts to, really, is a requirement of sameness of substance. In addition to having the experience of remembering some set of experiences in the past, it requires that there be a causal pathway, through a continuing body, from the experiences to the recollections. There is no indication that this changes the phenomenological character of the experience at all, only that it guarantees that a single body (or at least brain) is present for both the experience and the recollection.

This requirement does make us feel better about the worry that memory alone is not enough to constitute identity, but how does it do so? Essentially by appealing to the intuitions supporting a sameness of substance view—the very intuitions that Locke's arguments were meant to overcome. For this reason, Locke's original arguments can be brought into play against this means of developing his insight. Why should it matter to us that the memory experiences in question be in one substance rather than another if it does not change the character of consciousness? If the memory experiences alone are not sufficiently strong to constitute identity, why should their being placed in a substance continuous with the one that had the experiences be a "consummation devoutly to be desired"? Or, put another way, why on this view should the fact that this lump of stuff here will have memories of these experiences in the future be more comforting than the fact that those who come after me will remember me when I am gone? I do not deny that it may feel more comforting. There is no doubt that one strand of our thought on personal identity places that identity in the continuation of substance, and there is much to be said for sameness of substance views. Locke, however, has identified another strand of thought according to which what matters is the character of our experience. To solve difficulties with that view by adding a requirement of sameness of substance is not to develop the basic insight, but to reject it. If the simple memory theory provides too weak a relation to constitute identity, then the psychological continuity theory does as well.

Indeed, the psychological continuity theory may be even worse off in this regard because of the method by which it attributes particular actions and experiences to persons. On the simple memory theory, one needed to be directly conscious of an action or experience for it to be hers. Although, as we have discussed, this seems like too strong a requirement in the end, at least we had some explanation for why this relation was supposed to be important. Those experiences of which we are directly conscious are experiences that must necessarily affect our well-being. Their nature matters to us because we experience them.

They are also, at least according to Locke, tied to responsibility in this way, because we can know them to be our actions or experiences, we have a responsibility to and for them that we could not otherwise have. The Lockean insight thus seems to rest on the special relation we have to experiences while we are conscious of them. According to the psychological continuity theory, however, there are many experiences—and even whole life phases—that are counted as mine even though I no longer have any consciousness of them at all. They are no more connected to my present consciousness than they would be by a sameness of substance view. The original appeal of Locke's theory is thus lost on this view.

None of this is to deny that memory—at least conceived as a simple recollection (even appropriately caused)—is too weak a relation to seem intuitively plausible as the relation which constitutes personal identity. It is just to say that psychological continuity theories do not remedy the deficits of the simple memory theory in a way which counts as a development of Locke's insight. One response to the recognition of the weakness of the memory theory, of course, is simply to reject the Lockean insight in favor of a sameness of substance view, and many have done just that. If, however, we think that there is still something valuable in the original arguments against sameness of substance views, the way to speak to the weakness of the memory theory would be to develop an account of psychological continuity defined in terms of conscious connections richer and deeper than memory, or to deepen our account of memory itself. Simply adding more of the same or continuity of substance will not do the trick.

A seemingly contradictory set of intuitions also suggests that the psychological continuity theory is too strong, particularly in its account of attribution. Even though this theory allows that experiences and life phases of which we are no longer conscious can be ours, attribution still depends on an initial act of consciousness. It is because we are connected through overlapping chains to some past time at which we were conscious of an action or experience that the life phase and its experiences become ours. This is

problematic because it leaves no room in the view for the impact and attribution of unconscious (or nonconscious)<sup>4</sup> experiences to a person. If the person was not conscious of an experience at the time to which she is currently connected by overlapping chains, then the experience cannot be hers. It seems clear, however, that experiences of which we are not conscious can be part of our psychological lives. To name just two species, dispositional states and repressed states seem as if they can contribute to identity every bit as much as consciously entertained states, but they are ruled out as attributable to the person on this view.

It might be protested that psychological continuity theories are in fact perfectly plausible on this matter. All they require is that an experience have been conscious at some time. Then, if the time at which the experience was conscious is linked via other conscious connections to the present, the unconscious state is made part of the present person through its attribution at the past time. This will probably capture many of the states we wish to attribute to a person despite the fact that she is not conscious of them, but it will not entirely solve the psychological continuity theorists' problem. First of all, this view does not give us a way to distinguish between psychic elements that are at one time a part of consciousness and then lost forever and those which remain as unconscious states. Because the state's connection to the present is only a courtesy via the connection of the past person to the present person through other states, there is no difference on this view between an experience that was part of consciousness and then faded entirely and, for example, a repressed state that is still actively at work in a person's psychological life, although no longer part of consciousness. There is an important difference between an experience that is mine because I experienced it in the past but have now forgotten it entirely, and one that is mine because I have repressed it and am still suffering the symptoms of that repression, and there is no clear way to capture this difference in the psychological continuity theory.

Moreover, it seems likely that some experiences or features of our psychological lives may



be ours even if we are never explicitly conscious of them. We can be affected by a great deal that does not make its way into the realm of consciously entertained experience. Although this claim is itself plausible enough, it may seem that to make it I have to revert to the perspective behind sameness of substance views as I earlier accused psychological continuity theorists of doing. Things are a bit more complicated in this instance, however. The idea that unconscious states can be part of who we are arises in the context where we are thinking of ourselves as psychological subjects, and not just as substances, and this is, among other reasons, because many states of which we are not conscious are implicated in the relations of concern and responsibility that Locke identifies. There is a vast difference between the case where an experience was had by a particular substance (soul or body) that is currently cut off from it—as in Locke’s hypothetical cases—and a case where we say that a person has an (again, in a general, non-technical sense) unconscious memory, or belief, or desire. In the latter case, there is an assumption that the state is not simply wiped out of the psychological economy, but is still playing a role. It will impact a person’s well-being and the nature of her consciousness even if she is not directly aware of the state itself and cannot become so. It can, moreover, inspire the feelings of guilt or shame or pride associated with moral agency even if the source of these feelings remains obscure. The relation we have to our unconscious states is thus generally far more intimate and far more relevant to our present consciousness than states that we would consider “not ours” on the Lockean view—states that befell this substance but have no impact on current experience at all. It thus seems both necessary and possible to find a way to make unconscious states attributable to a person within a broad Lockean perspective. The psychological continuity theory, however, does not seem to have a means of doing so.

What this discussion of the failing of the psychological continuity theory reveals is that what originally looked like a clash of intuitions about whether an experience a person does not remember can be his experience is really about con-

sciousness rather than memory itself. Memory became important in the discussion as the medium through which consciousness is continued over time. The real conflict of intuitions, which makes development of the Lockean insight so difficult, is the conflict between our sense that the person (understood as psychological subject) should be identified with conscious mental activity only and the sense that the psychological subject is more extensive and includes non-conscious mental activity as well. Both ideas have their appeal. Freud has pointed out many *motives* for identifying ourselves with our consciousness—in doing so we can avoid acknowledging characteristics, fantasies, desires, and experiences that are, for one reason or another, unpalatable. Locke, on the other hand, tries to present *reasons* for identifying the person with conscious psychological life. The reasons involve first the fact that it is the continuation of conscious experience that seems both necessary and sufficient for our own continuation, and second the ways in which our consciousness of experiences underlies the special kind of self-interested concern and moral agency that distinguish persons from other kinds of creatures.

Our discussion of how psychological continuity theories grow out of Locke’s view, and how they fall short of their goals, has given us the tools we need to develop Locke’s insight in a more satisfying way. From Locke we have the reasons, outlined in the previous paragraph, for identifying the person with the conscious self. To develop his insight, we thus need a view that incorporates the role of self-conscious awareness in the constitution of identity, and also recognizes the link between identity and issues of self-interested concern and moral responsibility. From psychological continuity theories, we get the strategy of separating these two elements, offering distinct but interrelated accounts of personal continuation and the attribution of particular actions and of experiences—the first depending on conscious connections between the different parts of a person’s life, and the second allowing for the attribution of experiences a person does not remember. Their theory of attribution does not, however, seem to do the work it needs to. Ulti-

mately, it loses the relation to concern and responsibility that is so important in the Lockean view, and it does not say enough about the difference between experiences that are really dead to us and those which, though not part of consciousness, are an active part of our psychological lives. To give a more satisfying development of the Lockean insight, then, we need to give a better account of how and when states of which a person is not conscious can be attributed to her. I offer such an alternative in the next section.

### AN ALTERNATIVE ACCOUNT

The task of developing the Lockean insight is now twofold. First, it is necessary to give an account of the kind of connection to the past that a person must have to develop the type of self-conception that constitutes personal identity. Second, it is necessary to give an account of the attribution of particular actions and experiences that allows for the attribution of unconscious elements as well as conscious ones. It is easiest to begin with the first task. Locke tells us that we make ourselves selves, and so determine our identity, by forming a self-conscious conception of ourselves as persisting subjects. The exact nature of this self-conception is never really spelled out, however. In the hands of psychological continuity theorists, it becomes the requirement that we have a sufficient number of memories of past experiences to connect us to a past time. Undoubtedly, this is a very important part of how we come to have our self-conceptions, but as a full account, it is rather thin. All that is required to have the appropriate sort of self-conception—the kind that determines personal identity and continuation into the future—is to have knowledge of what some collection of experiences is like from the inside. It is enough, on this view, just to have access to a certain number of memories. There is no further requirement on how these memories are to cohere or to be associated with present states.

The alternative development of the Lockean view I suggest adds to the recognition of the importance of memory and brute self-consciousness a recognition of the importance of being

intelligible to ourselves. To have the kind of self-conception that constitutes personal identity on the view I am urging (call it the “self-understanding view”), one must not just know about some collection of past experiences and think of them as hers, she must see her life as unfolding according to an intelligible trajectory, where present states follow meaningfully from past ones, and the future is anticipated to bear certain predictable relations to the present. This does not mean that a person’s life course is entirely under her control, only that she can see connections between how things were, how they are, and how they are likely to be. Having a self-conception does not just amount to knowing that one has a past and will have a future, but also involves seeing these as inherently interconnected and rich with implications for one another.

This understanding changes the fundamental nature of the self-awareness that constitutes us as persons. It is no longer a passive knowing that we have had experiences, but a more active attempt to make sense of those experiences and understand where they are leading us. Fully understood, Locke’s fundamental insight is that as self-conscious entities we are interested in the character of our experience, and also in what we should do and what kind of person we should be. What this means, however, is that we are constantly self-monitoring, keeping track of how we are feeling, what we are doing, and what we are like. This self-monitoring is mostly implicit. There are many occasions where we introspect and consciously consider the trajectory of our lives and how its episodes fit together, but usually we are caught up in the activity of living, and this work goes on in the background. On the self-understanding view, it is this self-monitoring that gives us our sense of continuation and coherence as a self, and so provides the kind of self-conception and relation to a particular past that constitutes personal identity.

The basic picture of this self-monitoring is nicely developed by Raymond Martin in his book *Self-Concern*. There Martin develops the notion of a “perceiver self.” We experience the world, Martin says, as if one part of the self were split off from the flux of events as an observer, watch-

ing and recording the stream of our experience. Martin argues that the perceiver-self is an illusion, and of course in some sense it must be; there is no homuncular entity within people who is the observer of their experience. Nevertheless, as Martin indicates, the sense that there is such a self is a robust and pervasive element of experience, and a central feature of human psychological organization. He also suggests that in thinking about personal survival it is the continuation of this self in which people are interested; they think they have survived if the perceiver-self continues, and that they will die if it does not.

Obviously, there are many issues to be settled about the nature and function of the perceiver-self before any forceful claims can be made about its role in personal identity. Martin says a great deal on this subject, and there will doubtless be much more discussion to come.<sup>5</sup> For present purposes, however, it is not necessary to put too much metaphysical weight on the concept; it can simply be used to represent a widespread and familiar picture of psychological continuation and personal survival. A somewhat more whimsical version of this picture is found in Michael Frayn's novel *Headlong*, where the protagonist undertakes a common sort of internal dialogue to convince himself to do what he knows to be wrong:

Odd, though, all these dealings of mine with myself. First I've agreed to a principle with myself, now I'm making out a case to myself and debating my own feelings and intentions with myself. Who is this self, this phantom internal partner, with whom I'm entering into all of these arrangements? (I ask myself.)

Well, who am I talking to now? Who is the ghostly audience for the long tale I tell through every minute of the day? This silent judge sitting, face shrouded, in perpetual closed session? (Frayn 1999, 126–127)

The “perceiver-self” then should be thought of as a stable observer who views and records the passing flux of experience, and recognizes it as part of a single life; it need not be an actually persisting agent, or even a truly continuous psychological process, but it is a process that gives rise to the background sense of a stable self of the sort whose existence seems crucial to personal identity on the view we have been exploring. My

suggestion is that it is this sense of a stable perceiver-self, rather than a simple knowledge that one has had experiences in the past, that constitutes the continuation of consciousness that constitutes personal identity over time.

One way to make this rather abstract conception more concrete is to connect this self-monitoring to certain capacities. One quite simple implication of having such a self-conception is that a person can generally answer questions such as, “Why do I feel this way?” or “Why am I doing this?” should they arise. If she cannot answer them, she should be motivated to look for an answer. Answers to questions of this sort usually involve a number of factors. If one cannot easily make sense of the way one is feeling or choosing to act, however, it is natural to look for external explanations, considering the environment carefully to see if there is some hitherto undetected factor that is exerting an influence. In this way one might notice that it is the gathering clouds outside that are making one gloomy or anxious, even if one had not noticed them before; or that one's mother is constantly sending off subtle signals of disapproval and that this can explain one's guilt.

Sometimes, however, scrutiny of our environment and of our conscious internal states still leaves us baffled about why we feel or act as we do. This unintelligibility threatens our integrity as self-conscious subjects—in the subject, as in the world more generally, there should be no events that are simply uncaused. This does not mean that we must fully understand all of our feelings or motives, but only that we should not be at a loss as to where to start in such self-understanding. This really would undermine the difference between punishment and being created miserable. If we are at a loss, and no overlooked external factors can be found, it is natural to look for occult internal causes—nonconscious memories or impulses. Descartes offers an early example of this strategy with respect to his perceptions in the course of his *Meditations on First Philosophy*. He recognizes that he had previously taken his perceptual images to come from external objects because he was not aware of bringing them about himself. In the context of

the meditations, however, he is supposing that there are no external objects, and so wonders if the images might come instead from some unknown part of himself. Although he ultimately decides that they come from an external world via God, his reflections demonstrate nicely the kind of dynamic I have in mind—first, the basic idea that it is our responsibility not only to know the contents of our consciousness but to understand their origins, and second the willingness to consider that we could, unknowingly, be the cause of our conscious experience. In a context more directly related to the current discussion, Freud’s “discovery” of the unconscious follows just this logic, and it is, of course, this insight which is the inspiration of the self-understanding view with its emphasis on the importance of being intelligible to ourselves.

In postulating inner sources for our experiences and actions we recognize that there are parts of ourselves that we are not conscious of, but that are yet part of our psychological lives in an importantly intimate way. These experiences impact the nature of our conscious experience, and hence should be part of the purview of our self-interested concern. They also influence what we do, and lead to experiences of guilt, shame, or pride, and so are connected to our moral agency. To have an identity-constituting self-conception, then, a person must demand (at least implicitly) a kind of coherence and intelligibility to the course of her life. The past should not only be remembered; it should help to explain the present, which in turn should help predict the future. To achieve this intelligibility, we often need to allow for the impact of psychological features and experiences of which we are not directly conscious. The suggestion is thus that the attribution of individual states should be on this basis. Those memories or desires or motivations whose existence as part of the psychological economy must be postulated to make sense of a person’s experience or the course of her life will be considered her experiences. In this way, we can make room for the attributions of experiences that are nonconscious without violating the fundamental Lockean insight. These experiences still affect us along the dimension of plea-

sure and pain, and are still connected to our capacities for moral agency. They are part of what determines how our lives will unfold, and what our experience will be like.<sup>6</sup>

The proposed alternative to the psychological continuity theory is thus a view that develops Locke’s idea that to be a person is to understand oneself as a persisting being in terms of the demands we make that our lives be intelligible. To be a person on this view is implicitly to keep track of the unfolding of one’s life. The particular type of self-concern that Locke takes to be definitive of personhood, as well as the capacity for moral agency, depends on our not just knowing ourselves to persist, but actively seeking to understand how our lives come to be pleasant or unpleasant, learning lessons from the past and applying them to the future. The stream of consciousness that we count as personal continuation involves understanding how the connections between past, present, and future work for us—not just seeing the present as connected to the past, but as flowing from it. This sometimes depends on the recognition of psychological states that play a role in determining our conscious experience, although they are not themselves conscious. It is important to understand that for such states to be attributed to a person, she need not consciously reflect on her history and accept their existence. This is worthy work and may be well worth undertaking for many reasons. For us to say of a person that a nonconscious experience is hers, however, all that is required is that it in fact be necessary to make her psychological life intelligible, whether she recognizes this fact or not.

There is a longstanding idea that memory plays a crucial role in the constitution of personal identity over time. This idea is tied to the Lockean insight about the importance of continuity of consciousness for personal identity. There is also a sense that a person can obviously have experiences she does not remember. This conflict of intuitions seems to be more than simply a conflict over whether persons are to be viewed as psychological subjects or substances. It is also internal to the view of persons as subjects. Psychological continuity theorists attempt to recon-

cile this conflict by developing a view where a sufficient number of memory connections between one moment and the next defines the basic continuity of the person, and attribution of forgotten experiences takes place through the relation of personal identity over time. Although this is a move in the right direction, it fails to capture the Lockean insight in many respects, and also fails to capture the way in which unconscious states can be part of who we are. An alternative account—the self-understanding view—recognizes the original conflict as between intuitions that identify the psychological subject with conscious states and intuitions that see psychological life and the subject itself as involving much that is unconscious as well. We can accommodate this latter understanding of the person and still retain Locke’s insight if we simply recognize the fact that unconscious states have a powerful effect on consciousness and self-consciousness. The self-consciousness that Locke is after requires more than memory, it requires a certain level of self-understanding, and this, in turn, leads us to the attribution of psychological elements, which help explain how our conscious experience comes to be as it is. The memory theory on its own is obviously too simple to explain what is distinctive about being a person, but the idea that we are what we are—and who we are—because we understand ourselves in a certain way is not. On at least one important conception, to be a person is, as Locke says, to recognize oneself as a person, and here identity is indeed determined by self-understanding. Given that we are ultimately trying to define beings as complex as ourselves, however, it should be no surprise that this self-understanding involves more than a simple act of recollection.

## NOTES

1. I do not mean to imply that these are the only difficulties with a memory theory of identity. There are many other famous objections raised in the literature. These are, however, the objections that are important in what follows.

2. In a slightly different context, I have called these, respectively, the “reidentification question” and the “characterization question” (Schechtman, 1996, 1–2).

3. These theorists all differ from one another in a variety of details of course. Parfit, in particular, differs

from the others in one of the details I have discussed here. He does not require any particular causal connection between recollections and the experiences they recollect to make them identity constituting. This is a significant deviation from the standard view, but because it only makes it harder for Parfit to speak to the difficulties I raise in the next section, I ignore it in what follows.

4. I have no technical sense of “unconscious” in mind here. For now when I talk about “unconscious” and “nonconscious” psychological states, I mean only states that are not consciously entertained, but are nonetheless part of the overall psychological economy. I will say a bit more about this in a few pages.

5. Martin believes that this assessment of the conditions of survival is ultimately a mistake because the perceiver-self never actually persists (and people do sometimes survive). An alternative way of reading these facts would be to say that, to capture what is accurate in this intuitive sense of survival, it is essential to think of the continuation of the perceiver-self as a phenomenological rather than metaphysical fact. That is, if a person can experience the perceiver-self as continuous and persistent she will have the kind of psychological continuation that, in at least some moods, is taken to constitute personal survival. Resolving these issues is, of course, a matter for a different series of papers.

6. It is worth noting that it may be absence of this kind of intelligibility as much as the unity of body that stands behind our hesitations to say that there is more than one person present in cases of Multiple Personality Disorder. In most such cases, there are periods of blackout and a variety of other puzzling events that keep any of the streams of consciousness from providing a really intelligible life narrative. From the outside, moreover, it often seems clear that if the streams of consciousness could have access to each other, the intelligibility of each would be greatly increased. In this respect, this is more like a case of a person having experiences of which he is not conscious than a case where they are not his at all. This account can thus capture both our inclination to say these cases involve more than one person sharing a single body and our inclination to say that this is not so.

## ACKNOWLEDGMENT

I am grateful to Grant Gillett for helpful comments throughout the process of writing this paper.

## REFERENCES

- Descartes, R. 1992. *Meditations on first philosophy: In which the existence of God and the distinction of the human soul from the body are demonstrat-*

- ed, ed. and trans. G. Heffernan. Notre Dame: Notre Dame University Press.
- Frayn, M. 1999. *Headlong*. New York: Picador.
- Grice, H. P. 1976. Personal identity. In *Personal identity*, ed. J. Perry, 73–95. Berkeley and Los Angeles: University of California Press.
- James, W. 1950. *The Principles of psychology*, Vol. 1. New York: Dover.
- Lewis, D. 1983. *Philosophical papers*, Vol. 1. Oxford, UK: Oxford University Press.
- Locke, J. 1979. *An essay concerning human understanding*, ed. P. Nidditch. Oxford, UK: Clarendon Press.
- Martin, R. 1998. *Self-Concern: An experiential approach to what matters in survival*. Cambridge/ New York: Cambridge University Press.
- Parfit, D. 1984. *Reasons and persons*. Oxford, UK: Clarendon Press.
- Perry, J. 1976. The importance of being identical. In *The identities of persons*, ed. A. Rorty, 67–90. Berkeley and Los Angeles: University of California Press.
- Reid, T. 1976. Of Mr. Locke's account of our personal identity. In *Personal identity*, ed. J. Perry, 113–118. Berkeley and Los Angeles: University of California Press.
- Schechtman, M. 1996. *The constitution of selves*. Ithaca, NY: Cornell University Press.
- Shoemaker, S. 1984. A materialist's account. In *Personal identity*, ed. S. Shoemaker and R. Swinburne, 67–132. Oxford, UK: Blackwell.