

## John Locke, Personal Identity, and *Memento*

Basil Smith

In his *Essay concerning Human Understanding*, John Locke famously offers an explanation of personal identity. In particular, he holds that our conscious memories constitute our identities.<sup>1</sup> Christopher Nolan's *Memento* (2000) tests this theory of personal identity. In the film, Leonard Shelby (Guy Pearce), an insurance investigator from San Francisco, suffers short-term memory loss as a result of an assault on his wife, Catherine (Jorja Fox), and himself. But now, without his memories, he can hardly function. He insists that his attackers have destroyed his ability to live. Leonard asks: "How can I heal if I cannot feel time?" The question for us, however, is what can *Memento* tell us about personal identity? I address this question while attempting to show that, in some measure, Locke and *Memento* offer similar sets of messages. In particular, I argue that they both provide evidence that memory constitutes personal identity. This is not to say that they offer exactly the same messages or that the messages they agree on are not counterintuitive on many fronts. The point of the comparison, rather, is to delineate what this theory of personal identity implies and how it leads to a theory of survival without identity.<sup>2</sup>

### Locke on Personal Identity

To begin, Locke defines what he means by a *person*. He says that a person is always conscious of what he thinks. The person "can consider itself as itself, the same thinking thing, in different times and places." But, so far, this definition does not say what personal identity is. To answer this, Locke notes that, insofar as consciousness always accompanies thinking, identity is a matter of consciousness "extending backwards to unite thought and action" (267). In other words, such identity is constituted by "being able to repeat the idea of any past action" in a series and, thus, is a matter of

*memory*. If something is not retrievable by consciousness ever again, it is not part of that person anymore (268). But this admission entails that personal identity is not static or unified but, rather, a complex set of memories that continually changes. To further argue for his positive thesis, Locke offers some negative observations, supported by various puzzle cases. In fact, his strategy is to offer such cases to indicate what identity is not, which, in turn, provides indirect evidence that his theory is true, that identity is a matter of conscious memories. These puzzle cases should be familiar to anyone versed in science fiction.

To bolster his thesis that personal identity is conscious memory, Locke argues that neither a soul nor a body is necessary for such identity.<sup>3</sup> To show this, he postulates that, if one consciousness had used many souls or bodies, as in the case of a contemporary man recalling the memories of an ancient philosopher, that consciousness would be who he is. In this case, it seems that, so long as a person is conscious of, and can remember, a linear series of memories, even if that consciousness is contained in different souls or bodies, those memories constitute that person (270). This seems correct, for, although we cite souls or bodies as evidence of personal identity, the identity itself *is* the series of memories. By contrast, Locke also notes that, if any single soul or body were to be host to different consciousnesses over time—as when, say, one man uses a body or soul by day and another uses the same body or soul by night—then we would say that that soul or body was not one person but two different people. In other words, since there would be two series of conscious memories, there would be two people using that one soul or body (274). This again seems correct, for, in such cases, we would say that there were two persons, not one, who made use of one soul or body. These puzzle cases indirectly suggest that personal identity is a matter of conscious memory.

Locke offers his theory of personal identity for two reasons. The first reason is that, since we are constantly changing as persons, we need an account of what makes us the same person over time. If we did not have such an account, it would be difficult to explain why our lives matter so much to us. In other words, our concern for such identity is "founded on a concern for happiness," which is easier to obtain if we are persons (278). The second reason is to provide a proper understanding of our *responsibility* for our actions. Locke says that, if we think of identity in this way, our consciousness can become concerned and accountable, in that it "owns and imputes to itself past actions" (277). It follows that, when any consciousness happens to lose any memories, when any memories are irretrievable, then they are

no longer part of that person. Locke insists that whatever past actions consciousness cannot reconcile to its appropriate present are as though they have never been done. To his credit, however, he concedes that there is often no way for us to say when this is.<sup>4</sup>

Locke incurs two problems with his theory of personal identity that must be mentioned before proceeding. The first of these, as Thomas Reid notes, is that his theory seems to deny the *transitivity* of such identity. Transitivity is the logical relation that, if A is B and B is C, then A is C. Imagine a boy who stole apples and was punished, who later won an award as a young officer, and who is now a retired general. The young officer remembers the boyhood events, and the general remembers the young officer events but cannot recall the boyhood events. Reid notes that, for Locke, this lack of conscious connection indicates that the boy and the general are not the same person and, thus, transitivity is denied. He insists that, “if there is any truth in logic,” the boy and the general *are* the same person.<sup>5</sup> Yet this objection is not decisive, for many of the stages from the boy to the general *overlap*, so there is a sense in which the general and the boy are the same person after all. But Locke does not say this, for, if there really are *no* memories that the boy and general share, there is no reason to say that they are the same person. So he admits that personal identity is *not* transitive in the way required here.

The second problem with Locke’s theory is that conscious memories “presuppose, and therefore cannot constitute, personal identity.” In other words, memories seem to be united in a series only because most of them really happened to the same person. If this were not the case, so the objection goes, merely having deluded memories of being someone else would make one into that person.<sup>6</sup> But, given this possibility, it is more plausible to say that personal identity is constituted by the “thinking substance” (a mind or soul) and is not a matter of conscious memories at all. This objection too is less than decisive. Derek Parfit, a contemporary philosopher, notes that it assumes that the conscious memories in any series really did happen to the same person. To remedy this, Parfit insists, a notion of memory that does not presuppose this is easily developed.<sup>7</sup> If conscious memories are thought of in this way, we may suppose that they are in a particular series, but not that they happened to any particular person. These two objections indicate that personal identity is not transitive, that our conscious memories don’t have to be true, and that there is no unified or static self underneath memories.

In what follows, I will ignore these common criticisms of Locke on

personal identity because, as we have seen, they do not work—but also for a more compelling reason. This is that *Memento* threatens personal identity in a different way. The film suggests, in the character of Leonard Shelby, the problem of *fusion*, or the problem that two conscious series of memories might be combined. In point of fact, what is unique about *Memento* is the way in which it poses this problem. It is not just that Leonard may have been fused from two consciousnesses, but also that there is no way to discern, either from the inside or from the outside, which elements of the two former persons now exist in the resultant person, and, thus, that nobody knows which beliefs from which series are true or false. Leonard may be constituted by two series of conscious memories in an uneasy mix. But what is really troubling is how his plight may mirror ours.

### The Meaning of *Memento*

When Leonard and his wife, Catherine, are assaulted in their home, Catherine is killed, and Leonard is struck on the head, losing the ability to make new memories, which eventually drives him to seek vengeance. Yet this is all on the surface, for underneath this plot lurks the problem of what constitutes personal identity over time, or the issue of what makes a person the same person at two distinct times. Leonard insists that “we all need mirrors to remind ourselves of *who* we are” and endeavors to find or create some memories for himself. *Memento* asks us to question who Leonard is after he is unable to make new memories and, by extension, to ask that of ourselves. But the film is not neutral on the answer to these questions, suggesting that we revise many of our presuppositions about personal identity. In particular, it asks us to abandon the notion that personal identity is transitive, that our memories must be true, but also that such identity is not static or unified. In rough outline, then, *Memento* suggests just what Locke argues concerning personal identity and what Parfit adapts. To explain how this is so, it will be necessary to examine the plot in detail.

*Memento* is confusing in two ways. The first way is that the film is shot in a disorienting fashion. It has both color and black-and-white scenes. The color scenes are presented in reverse chronological order. They take place, moreover, over a short period of time, such that the end of each new scene is repeated as the beginning of the next one. Interspersed with these color scenes are black-and-white scenes of a single telephone conversation. This conversation is presented in normal chronological order and occurs before the color scenes. Since the color scenes are presented backward and

the black-and-white ones forward, in the beginning of the film, the last chronological black-and-white scene gradually turns into the first chronological color one. *Memento* is also shot with numerous flashbacks, of two principal types. The first type is those that are shot twice, with significant differences, and they suggest different pasts. The second type must be pure fantasy and could not occur in any past.<sup>8</sup> The point of all these flashbacks is to disorient the audience as much as possible.

However, *Memento* is confusing in a second way, which is more important here. There are two versions of the plot. It is prudent to describe these versions not as we see them but as they occur chronologically. In the first version, Leonard and Catherine are assaulted at home, and the latter dies. During this assault, Leonard is struck on the head, suffers short-term memory loss, and, thus, cannot make new memories. But he remembers the crime and hopes to avenge his wife, the police report having convinced him that there was a second assailant. Because of his phone conversation (the subject of the black-and-white scenes), Leonard infers that the second assailant is named John G. He proceeds to track this person down—by taking Polaroid pictures of everything he will soon forget, by making copious notes, and by tattooing important facts on his body. But he is haunted by the irony that, in his former job as an insurance investigator, he had denied coverage to one Sammy Jankis (Stephen Tobolowsky), who had been, he had suspected, faking the very condition that he now suffers from. Unfortunately, Sammy had a diabetic wife who could not bear the loss of her husband as a person. In her despair, she allowed Sammy to inject her with insulin, killing her, and Sammy was then put in a mental institution.

Leonard is also used by Teddy (Joe Pantoliano), a former cop who worked his case and who presumably is his interlocutor in the black-and-white telephone conversation scenes. Teddy has Leonard kill any number of drug dealers by leading him to believe that those persons are his John G. Leonard eventually realizes this deception and is not pleased. Importantly, he makes a note about Teddy, saying: “Don’t believe his lies.” He spots Teddy in his car and then has a tattoo made of the license plate number, which will later suggest that Teddy may be his John G.<sup>9</sup> But at the time Leonard *knows* that Teddy is not his man and that he will forget this later. He asks himself: “Do I lie to myself to be happy?” He knows that this lie to himself may result in his committing murder, yet he lies anyway. Leonard also meets Natalie (Carrie-Anne Moss), a barmaid and girlfriend of a drug dealer whom he has just killed. She suspects that Leonard has killed her boyfriend but still uses him to protect herself against Dodd (Callum

Rennie), another drug dealer. She notices that Leonard has a tattoo of a license plate number on his leg and runs the plate for him. Leonard matches up the license plate number with new information and infers that Teddy is his John G. He subsequently lures Teddy to a warehouse and kills him.

In the second version of the plot, events are quite different. In an important conversation between Leonard and Teddy, a different past is suggested. Teddy admits that he was the cop assigned to his case and admits that he now uses Leonard to kill drug dealers. He tells Leonard that, by doing so, he has given him “a reason to live.” Teddy insists that Catherine survived the assault, which fact is hinted at in two ways. In a flashback, after the assault, Catherine blinks. Moreover, the date of death listed on the police report is much later than the date of the assault. Teddy also says that Catherine was the diabetic, and this too is hinted at. In a flashback, we see her being injected with insulin, and then the same scene is replayed with her not being injected. Leonard has apparently transposed elements of his past with the past of Sammy Jankis. In effect, he has *projected* his own memories onto Sammy and *invented* false ones for himself. This explains why he says of his wife that she “was perfect to me.” He is then committed to the mental institution, and this too is hinted at. In a flashback, we briefly see Leonard sitting in the institution, from which he later escapes. Teddy lastly tells Leonard that there really was a second assailant, his John G., and that Leonard has already killed that person.

Understandably, when Teddy tells him all this, Leonard is dismayed at having been turned into a killer. In fact, he does not believe that Catherine survived the assault, that she was a diabetic, or that it is possible that he has transposed any memories. Nor does he believe that he has projected his own memories onto Sammy or that he has invented false ones for himself. Despite this, it seems that he is willing to manipulate the evidence, to create a puzzle for himself, merely to justify his ongoing quest to avenge his wife, even if that puzzle is already solved. He even deceives himself into believing that Teddy is his John G. and, thus, sets up the latter to be murdered. Leonard later meets Natalie, whom he protects from Dodd, but who accidentally puts him onto Teddy, whom he then kills. This version of the plot seems to be what Christopher Nolan intended, but this is hardly the end of the story. Even if this version of the plot is the intended one, it still leaves many elements unsettled, and there may be no way to reconcile them. Yet, for the purposes of this essay, this is no matter. Given these outlines of two different versions of the plot, we can still address our main concern, which is what the film can tell us about the issue of personal identity.

### Locke, Parfit, and Memento

In the foregoing, I have explained how Locke argues that personal identity is a matter of our conscious memories over time, that Parfit adapts this argument, and that *Memento* tests this general theory. The importance of the film is that it is another puzzle case, although one of a unique sort. Is Leonard Shelby the same person as he was when he could make new memories? If that original person is gone, is he now someone different? To answer these questions, let us distinguish Leonard 1 and Leonard 2, corresponding to the two versions of the plot, cited above. Leonard 1 is the person who suffers an assault, who is trying to avenge his wife who has died in the assault, and so on. He has a linear series of conscious memories with only one gap, that produced by the assault. He makes new memories, which connect onto his old series and are then forgotten. But, actually, Leonard 1 may not sound like a person at all, for his personhood is *not* bound by chronology. The new parts of his series of memories are almost immediately forgotten, regardless of *when* this occurs. But, if we assume that personal identity is a matter of consciousness of memories, dropping any chronological requirement for it should not be that counterintuitive.<sup>10</sup>

Leonard 1 is not a multiple person or a person with overlapping but distinct identities. He may seem like a multiple person, for his conscious memories are born and then die, every few minutes, compounding his personhood over time. If this were correct, he would become and then cease to be many persons, with the only overlap being what he can manage with his professed “conditioning,” whatever that amounts to.<sup>11</sup> But, although Leonard 1 suffers his memories’ being born and then dying continually, this is harmless here. In fact, it is not as though his entire personhood is born and then dies, with a distinct person taking its place each time. This is because, although his entire series of conscious memories may be reproduced and eliminated every five minutes, in the next five minutes that *same* series comes back to him, except for those few memories that were formed in the previous few minutes. Leonard 1 returns every time, with the exception of those recent conscious memories, and, thus, is just who he believes himself to be. Since this is so, he is not a multiple person, and there is no mystery about his personal identity. Thus, Leonard 1 does not shed light on the issue of personal identity.

The issues are different when we look at Leonard 2. He is the person who suffers an assault and who accidentally kills his wife, who did not die in the assault, with insulin. He then transposes his own past with that of

Sammy Jankis, in that he projects his memories onto Sammy and invents false ones for himself. He then enters a mental institution, escapes, and tries to avenge his supposedly murdered wife. Leonard 2 is a more troubling case for the issue of personal identity, for a myriad of reasons. The main worry is that he should have a linear series of conscious memories but does not. To see why this is, let us create another distinction. Leonard 2a is the person who existed before the assault. This is the person who lived with his wife but whose series of conscious memories halted soon after the assault. Leonard 2a suffers a kind of death when his entire set of memories is infected, is fused with that of another person. In point of fact, there is no way to ascertain the details of this fusion, either from the inside or from the outside, or to decide which of his present memories are true and which false. But this is not really death, for *some* person seems to survive.

Leonard 2b comes into being at this point. He is the person who is created by the assault. This occurred when Leonard 2a took his own conscious memories and projected them onto Sammy Jankis and then invented false ones for himself. But *who* is Leonard 2b? In point of fact, Locke would not find this an easy question to answer, for this person has not a series of conscious memories but only an uneasy mix of two, which are fused together. In Locke’s theory, there does not seem to be an answer to this question. Parfit insists: “Any two people fused together will have different characteristics, different desires, and different intentions.” The trouble is that some of these states will be compatible and some not. It follows that, in any fused person, when that person is stable enough to have a consistent set of characteristics, some of both persons will be sacrificed. Parfit notes that, afterward, the resultant person will not be wholly similar to either and, thus, that this may seem like a kind of death.<sup>12</sup> In other words, such identity fusion may strike us as death, for our personhood changes. But such partial survival is not really the end either. Leonard 2b is such a person, although the question now is whether such personhood is worth having.

There is no real answer to who Leonard 2b is—only that he bears degrees of resemblance to both his former persons. The message here is that “survival itself can have degrees” yet also that this sort of identity is, in fact, worth having.<sup>13</sup> Mark Rowlands notes that this revelation can change our expectations, for, as soon as we drop the prejudices that personal identity is transitive, that our memories must be true, and that there must be a static and unified self, we realize that no one is ever identical with himself over time but, rather, that “we are all just survivors, very close survivors, of the persons we were a moment ago.”<sup>14</sup> This realization allows us to drop our

vain hopes for anything more from personal identity and to see that we should alter our attitudes about it. But the moral here is that cases of personal identity fusion are not so unusual. The difference between Leonard 2b and us is that he has projected his former memories onto another person and created false ones for himself, which we presumably do not do. But this is an inessential difference, for we are still forced to *reinterpret* our own pasts (our memories aren't like photographs or written chronicles; they're hazy, fragmented, partial), and this activity of reinterpretation again issues in a mixture of true and false memories that become our personal identity. In other words, our true and false memories determine who we will be.<sup>15</sup>

Leonard 2b, as well as we, have fused personal identities, at least to some extent. There are two disturbing consequences of this. The first consequence is that, just as Leonard 2a has a special concern for his future, we have a special concern for our futures. Since Leonard 2a counts on not becoming Leonard 2b, and since we count on not becoming anything we do not choose to be, it is *rational* for both to have a special concern for their respective futures. But, if Leonard 2a comes to believe that he will be the fused Leonard 2b, and if we come to believe that we will be fused, then neither of us will count on being the same person in the future. If Leonard 2a and we suspect that our future persons will be entirely unlike us, then neither will have any rational interest in those future persons.<sup>16</sup> But, plainly, this is a problem, for we all care about our futures and do so rationally. However, this problem of concern for our future personhoods may not be so serious after all. We still may have many conscious memories that traverse the former and later persons. Since this is so, we can be optimists about our futures, in the hope that these memories will justify our special concern for them.

The second consequence of our having fused personal identities, at least to some extent, concerns the attribution of responsibility. If we are fused persons to any degree, this renders any attribution of responsibility difficult. The memory theory of personal identity, and its successor in terms of survival, is supposed to explain such responsibility, but does not seem to. Leonard 2b is a mix of Sammy Jankis and invented memories, just as we reinterpret our pasts and become a mix of true and false memories. This consequence spells trouble, for, in such cases of fusion, there seems to be no way to say *who* is really responsible for any action. If we are fused persons, it is difficult to know what *element* of us bears responsibility for our actions. This problem, although difficult, is not completely intractable, for we still do attribute responsibility to fused persons. In attributions, we

just attempt to decrease the amount of responsibility that the present person bears in inverse proportion to any increase in the fusion he suffers. In the end, however, perhaps there can be only conjectures about these difficulties concerning our concern for our futures or about the attribution of responsibility.

### Conclusion

So, what can we really glean from this comparison of Locke and *Memento* on the topic of personal identity? The first lesson is that personal identity is constituted by our having a series of conscious memories, at least usually. In most cases, this memory theory seems to be the only plausible candidate for a conception of personal identity. This is so even though there is no chronological constraint on any linear series of conscious memories. The second lesson is that any series of conscious memories can be fused with another, such that the result is a mix of two persons or a mix of the true and false. But, then, it follows that our personal identity is not really identity at all but, rather, a matter of survival, which, in turn, admits of degrees. This may seem like a radical notion, but it is so only because we are used to thinking of personal identity as transitive, of our memories as true, and of identity as static and unified. The third lesson is that, given this memory theory of personal identity, our special concern for our futures, and the attribution of responsibility to such persons, may turn out to have problems after all.<sup>17</sup>

### Notes

1. In point of fact, many philosophers call this theory of personal identity the *memory theory*. But this is misleading because memories are of many kinds of mental states, such as desires, beliefs, and even long-term goals. It follows that this theory of personal identity in terms of memory is really about psychological continuity over time.

2. John Locke bears this in mind, for he notes that his speculations on personal identity are "apt to look strange to some readers." This is so, he says, only because of the "ignorance of the nature of that thinking thing that is in us" (*Essay concerning Human Understanding* [1690], ed. Walter Ott [New York: Barnes & Noble, 2004], 278 [page numbers for subsequent quotations will be given in the text]).

3. It does not matter for this account if it is souls or bodies that are said to be necessary for personal identity. This is because the postulation of either as neces-

sary for such identity is open to the same criticism, that either souls or bodies may provide evidence for personal identity but are irrelevant to what it actually is.

4. Locke distinguishes between the *man* and the *person*. He says that, since these are usually the same, “human laws punish both,” and rightfully so. He insists that God will have the solution for our errors, for “in the great day, when the secrets of the heart shall be laid open, no one will be made to answer for what he knows nothing of” (278).

5. Thomas Reid, *On the Intellectual Powers of Man* (1785), reprinted in *Personal Identity*, ed. John Perry (Berkeley: University of California Press, 1975), 115.

6. Anthony Flew, “Locke and the Problem of Personal Identity” (1951), reprinted in *Locke and Berkeley: A Collection of Critical Essays*, ed. C. B. Martin and D. M. Armstrong (New York: Doubleday, 1968), 159.

7. Derek Parfit, “Personal Identity” (1971), reprinted in Perry, ed., *Personal Identity*, 209.

8. In particular, in the end of the film, we see a flashback wherein Leonard has “John G. raped and killed my wife,” yet also “I’ve done it,” tattooed on his chest, with his wife alive, lying beside him. But, if his wife is alive, he would not have either tattoo. This must be part of a dream, perhaps one that mixes his desire for vengeance with his desire to see his wife alive.

9. In this version, Leonard is correct that Teddy lies about many things: about his wife dying during the assault; about Sammy Jankis; and about his not yet having killed John G. Leonard is lied to by Teddy and, thus, is justifiably angry. In this version, Leonard has many reasons to kill Teddy after all.

10. There is a chronological requirement for personal identity on this account simply because all that matters now is continuity. This is easily imagined, e.g., in cases of hibernating for generations and then waking up. In such cases, chronology is broken, but, given the continuity, we would still expect to wake up as ourselves.

11. Mary Litch seems to understand Leonard in this way. In particular, she says that, after the assault, “there are *too many* distinct contenders” to be him (*Philosophy through Film* [New York: Routledge, 1992], 77). But, plainly, even if Leonard is reproduced every few minutes, it does not follow that his personhood is fractured by this.

12. Parfit, “Personal Identity,” 212.

13. *Ibid.*, 215.

14. Mark Rowlands, *Sci-Phi: Philosophy from Socrates to Schwarzenegger* (New York: St. Martin’s Griffin, 2005), 118.

15. Leonard 2b and we are still in different positions, however, for his fusion is worse than ours. Even so, his case suggests that, often, we do not know just how bad the fusion is. The trouble is that, without this sort of knowledge, we might be more fused than we think we are and that, because of that, we cleave to a false personal identity. Michael Baur interprets Leonard 2b in this way, yet with more detail

(see his “We All Need Mirrors to Remind Us Who We Are: Inherited Meaning and Inherited Selves in *Memento*,” in *Movies and the Meaning of Life*, ed. Kimberly Blessing and Paul Tudico [Chicago: Open Court, 2005]).

16. Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984), 307.

17. Theodore Sider notes that, even if the memory theory of personal identity does have such problems, that does not help any other theory. This is especially so, he says, because the problems are the same for other theories. See his “Personal Identity over Time,” in *Riddles of Existence: A Guided Tour of Metaphysics*, ed. Earl Conee and Theodore Sider (Oxford: Oxford University Press, 2005).