DID LOCKE DEFEND THE MEMORY CONTINUITY CRITERION OF PERSONAL IDENTITY?

JOHAN E. GUSTAFSSON

Even though John Locke is the most influential thinker in the discussion of personal identity, his account is usually thought to have been proved false by Thomas Reid's simple 'Gallant Officer' argument. Locke is traditionally interpreted as holding that your having memories of a past person's thoughts or actions is necessary and sufficient for your being identical to that person. This paper argues that the traditional memory interpretation of Locke's account is mistaken and defends a memory continuity view, according to which a sequence of overlapping memories is necessary and sufficient for personal identity. On this view, Locke is not vulnerable to the Gallant Officer argument.

Although Locke never explicitly states such a criterion, he is traditionally interpreted as defending the following memory criterion of personal identity:[1]

> *The Memory Criterion of Personal Identity (MP)*: A person P_1 who exists at t_1 is identical to a person P_2 who exists at a later time t_2 iff P_2 is at t_2 conscious of (remembers) any of the thoughts or actions of P_1 at t_1.

In other words, you are identical to a person in the past if, and only if, you remember any of the past thoughts or actions of this person. George Berkeley came up with an objection to this criterion

[1] See, e.g. Flew (1951: 55), Mackie (1976: 178–9), Parfit (1984: 205), and Noonan (2003: 9).

that was later put in its most memorable form by Reid as his Gallant Officer argument:[2]

> Suppose a brave officer to have been flogged when a boy at school, for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life: Suppose also, which must be admitted to be possible, that when he took the standard, he was conscious of his having been flogged at school, and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging.[3]

On the traditional interpretation the officer would be the same person as the boy because the officer remembers having been flogged at school and the same person as the general because the general remembers taking the standard. By the transitivity of the identity relation the general would have to be the same person as the boy, but since the general does not remember any of the thoughts or actions of the boy, the general is not the same person as the boy. We have a contradiction and thus Locke's theory has to be false.

However, the memory criterion stands in contrast to Locke's main statements of his account which are expressed in terms of 'consciousness' rather than 'memory'. Locke writes:

> This may shew us wherein *personal Identity* consists, not in the Identity of Substance, but, as I have said, in the Identity of *consciousness*, wherein, if *Socrates* and the present Mayor of *Quinborough* agree, they are the same Person: If the same *Socrates* waking and sleeping do not partake of the same *consciousness*, *Socrates* waking and sleeping is not the same Person.[4]

[2] Berkeley's rendition appears in Berkeley (1950), *Alciphron*, Dialogue VII, § 8, p. 299.

[3] Reid (2002), Essay III, Chap. VI, p. 276.

[4] Locke (1979b), II. xxvii. 19.

Here Locke seems to state the following necessary and sufficient criterion of personal identity:

> *The Consciousness Criterion of Personal Identity (CP)*: A person P_1 who exists at t_1 is identical to a person P_2 who exists at a later time t_2 iff there is a consciousness C such that P_1 partakes of C at t_1 and P_2 partakes of C at t_2.

Advocates of the traditional interpretation have argued that in order to make sense of some other passages of Locke, 'consciousness' in the above quote would have to be interpreted in a way that makes CP equivalent to MP. Section 1 examines passages that have been taken as evidence for the MP interpretation and argues that they do not support the traditional interpretation. Section 2 argues that at least one key passage in Locke contradicts MP. Section 3 defends a memory continuity interpretation of Locke. This interpretation is then shown to be immune to the Gallant Officer case, in section 4.

1. Passages that seem to support the traditional interpretation

A passage that some might take to be conclusive evidence for an interpretation like MP is the following, where Locke comments on whether a person living today could be the same as the ancient Nestor. Locke writes: 'But let him once find himself conscious of any of the Actions of *Nestor*, he then finds himself the same Person with *Nestor*'.[5] From this we can conclude that Locke holds the following sufficient criterion of personal identity:

(1) A person P_1 who exists at t_1 is identical to a person P_2 who exists at a later time t_2 if P_2 at t_2 is conscious of (remembers) any of the thoughts or actions of P_1 at t_1.

[5] Locke (1979b), II. xxvii. 14.

Obviously, this does not imply MP. Because MP, of course, also implies the following necessary criterion:

(2) A person P_1 who exists at t_1 is identical to a person P_2 who exists at a later time t_2 only if P_2 at t_2 is conscious of (remembers) any of the thoughts or actions of P_1 at t_1.

There is not, however, any textual support for (2) in Locke's writings on personal identity.[6] In this section I will examine passages that have led earlier interpreters to conclude that Locke accepts (2), and then argue that these passages do not support (2).

Kenneth P. Winkler concludes that Locke accepts (2) from § 20 of Locke's chapter on identity where he writes:[7]

> But yet possibly it will still be objected, suppose I wholly lose the memory of some parts of my Life, beyond a possibility of retrieving them, so that perhaps I shall never be conscious of them again; yet am I not the same Person, that did those Actions, had those Thoughts, that I once was conscious of, though I have now forgot them? To which I answer, that we must here take notice what the Word *I* is applied to, which in this case is the Man only. And the same Man being presumed to be the same Person, *I* is easily here supposed to stand also for the same Person. But if it be possible for the same Man to have distinct incommunicable consciousness at different times, it is past doubt the same Man would at different times make different Persons.[8]

Here Locke seems to think that it is a mistake to conclude that I must be the same person who had the forgotten thoughts, albeit this is a mistake that is easy to fall into by confusing my identity as a person with the identity of the man whose body is both mine and

[6] The lack of support for (2) in Locke have been noted by Helm (1979), Atherton (1983: 276), Jenkins (1983: 124), and Loptson (2004: 54).

[7] Winkler (1991: 205).

[8] Locke (1979b), II. xxvii. 20.

that of the person who had the forgotten thoughts. Nevertheless, this does not imply that Locke thinks that I must not be the same person who had the forgotten thoughts. He might need more information to conclude one way or the other. In the final sentence of the quote Locke makes a further qualification before he deems it past doubt that the man would at different times make different persons.

It is important to note that in this final sentence Locke does not claim that if a man presently has no memory of an earlier part of the same man's life, then he does not make the same person now as the same man did during the earlier part of its life. He only grants that a man with distinct incommunicable consciousnesses at different times would make different persons. Locke seems to claim that:

(3) If a person P_1 exists at t_1 that partakes of a consciousness C_1 at t_1 and a person P_2 exists at t_2 that partakes of a consciousness C_2 at t_2 and C_1 is not identical to C_2 and C_1 is incommunicable to C_2 then P_1 is not identical to P_2.

But (3) does not imply (2). One might object that (3) would imply (2) given that to partake of the same consciousness as an earlier person just is to remember any of the earlier person's thoughts or actions. To presume this, however, would be question-begging as support for MP, since presuming it in conjunction with CP implies MP. So the quotation does not show that Locke accepts (2).

E. J. Lowe concludes that Locke accepts (2) from Locke's discussion of responsibility and the distribution of rewards and blame.[9] The passage that may seem to imply (2) is the following, where Locke answers an objection that a man is the same person when he is drunk and when he is sober:

> But is not a Man Drunk and Sober the same Person, why else is he punish'd for the Fact he commits when Drunk, though he be never afterwards

[9] Lowe (1995: 113).

conscious of it? Just as much the same Person, as a Man that walks, and does other things in his sleep, is the same Person, and is answerable for any mischief he shall do in it.[10]

From this one might get the impression that Locke holds that because the sober man does not afterwards remember the actions of the drunkard, they do not make the same person, which would indicate that Locke accepts (2). But this is not Locke's view. He clarifies this passage in a letter to William Molyneux:

For I ask you, if a man by intemperate drinking should get a fever and in the frenzy of his disease (which lasted not perhaps above an hour) committed some crime, would you punish him for it? If you would not think this just, how can you think it just to punish him for any fact committed in a drunken frenzy, without a fever? Both had the same criminal cause, drunkenness, and both committed without consciousness. I shall not inlarge any farther into other particular instances, that might raise difficulties about the punishing or not punishing the crime of an unconscious drunken man.[11]

Locke's point is that the drunkard committed the crime *without* consciousness. This is why the sober man should not be punished for them, which explains his analogy with the man who walks, and does other things in his sleep. The sleepwalker, like the drunkard, is not conscious of the things he does. Therefore it is not because the sober man afterwards does not remember the drunkard's crime, that he should not be punished for it. It is, according to Locke, because the drunkard was not conscious of his crime. The last part of the quote, 'the crime of an unconscious drunken man', seems especially hard to make sense of under Lowe's interpretation, which has to say that this means 'the crime of a drunken man who will not later remember his crime' rather than the much more plausible

[10] Locke (1979b), II. xxvii. 22.

[11] Locke (1979a), L1693, pp. 785–6.

reading 'the crime of a drunken man who is not aware of the things he does'.

One might object, however, that the following part of the letter, supports Lowe's reading:

> But drunkenness has something peculiar in it when it destroys consciousness; and so the instances you bring justifie not the punishing of a drunken fact, that was totally and irrecoverably forgotten.[12]

The latter part of this quote might suggest that the drunkard in Locke's example was aware of his actions while drunk and only later were these actions totally and irrecoverably forgotten. However, note that here Locke is not talking about his own example but rather the instances Molyneux brought up in the letter to which Locke answers.

Granted, Locke does write in two passages that the sober man does not remember the misdeeds. However, since a natural consequence of a period of unconsciousness is that one does not remember any actions performed during this period, this is consistent with the interpretation that Locke referred to an unconscious drunkard. Furthermore, both passages are immediately followed by material that suggests that it is really the unconsciousness during the performance of the misdeeds that is at issue. In the first passage, quoted above ('...though he be never afterwards conscious of it?'), Locke responds with the analogy to the sleepwalker, which would seem misplaced if it was the mere forgetfulness of the sober man rather than the unconsciousness of the drunkard that he had in mind. Similarly, in the second passage (§22) Locke claims:

[12] Locke (1979a), L1693, p. 785.

For ... the Drunkard perhaps be not conscious of what he did; yet Humane Judicatures justly punish him; because the Fact is proved against him, but want of consciousness cannot be proved for him. (II. xxvii. 22)

Here, while '...be not conscious of what he did' suggest forgetfulness, the 'want of consciousness' seems out of place if the forgetfulness of the sober man is what is at issue.

Finally, one might feel that the case of a drunk who does things without any present awareness at all is an extreme, rare, and too strange one. The case where someone who blacks out and cannot later remember the actions he did while drunk is much more common. Is not the case on my reading too extraordinary to be what Locke had in mind? To this I reply that Locke does not hesitate to use extraordinary examples. Even if a drunk who does things without any present awareness at all is a rare case, it is fairly commonplace relative to Locke's other examples in the same chapter, where we find, for example, 'a very intelligent rational Parrot' who does 'discourse, reason, and philosophize', a conscious little finger, and a contemporary person who is identical to a person who existed in antiquity. With this in mind, the fact that Locke on my reading presents a strange example should not make it a less plausible reading. So the quote that Lowe refers to does therefore not show that Locke accepts (2).

A further passage in §22that may seem to support (2) concerns the Day of Judgement. Locke writes:

But in the great Day, wherein the Secrets of all Hearts shall be laid open, it may be reasonable to think, no one shall be made to answer for what he knows nothing of; but shall receive his Doom, his Conscience accusing or excusing him.

Apparently, in Locke's view, no one will be punished on the Day of Judgement for any action that he on that day genuinely fails to remember having done. Any interpretation that rejects (2) faces a

hurdle: either God has to restore all memories of the resurrected or some persons will not have to answer for all of their misdeeds. But note that accepting (2) is not enough in order to avoid this problem. The traditional interpretation, MP, faces the same problem. If, on the Day of Judgement, I remember one of the thoughts of a person *P* at time *t* but not any of the misdeeds of *P* at *t* then I do not remember all my misdeeds even on MP. Hence God will have to restore my memories or I will not be made to answer some of my misdeeds. To avoid this hurdle one would have to demand that one remembers all thoughts and all actions of a person in the past for one to be identical to the past person. But such a theory is unattractive and further lacks textual support. Thankfully, the hurdle is not so high. If you are already in the business of resurrecting people from the dead, why not restore their memories too?

In a very influential paper Antony Flew argues that in Locke's main statements of his account 'consciousness' is equivalent to 'memory' as this can be seen from the following passage in §25:[13]

> Could we suppose any Spirit wholly stripp'd of all its memory or consciousness of past Actions, as we find our Minds always are of a great part of ours, and sometimes of them all.

It is hard to understand how Flew sees Locke's supposed equivalence of 'consciousness' and 'memory' in the above passage. Perhaps Flew has in mind the common use of phrases of the form '*X* or *Y*' on which *X* and *Y* are equivalent expressions. But there is also another natural reading of phrases of the form '*X* or *Y*', namely, the disjunctive. Since a disjunctive reading of the above passage is also plausible, it does not show that Locke accepts (2).

[13] Flew (1951: 55).

2. A problem for the traditional interpretation

To see why Locke does not accept (2) and therefore not MP we need to look at his example of Nestor in §14 again.[14] It begins as follows:

> Let any one 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 *Thersites*, at the Siege of *Troy*, (For Souls being, as far as we know any thing of them in their Nature, indifferent to any parcel of Matter, the Supposition has no apparent absurdity in it) which it may have been, as well as it is now, the Soul of any other Man: But he, now having no consciousness of any of the Actions either of *Nestor* or *Thersites*, does, or can he, conceive himself the same Person with either of them?

In the discussion that follows Locke explains why your having the same soul as Nestor does not imply that you are the same person as Nestor. And then Locke writes: 'But let him once find himself conscious of any of the Actions of *Nestor*, he then finds himself the same Person with *Nestor*'.

It should be clear from the introduction to this example ('Let any one reflect upon himself ...') that 'him' in the second quote could be anyone. It should also be clear from the end of the first quote ('But he, now having no consciousness of any of the Actions either of *Nestor* or *Thersites* ...') that this person is not at present conscious of any of the actions of Nestor. Let that person be me. I am at present not conscious of any of the actions of Nestor. Suppose that last week, for one brief moment, I found myself conscious of one of the actions of Nestor. Then the second quote implies that

[14] Peter Loptson offers another argument against the memory criterion interpretation in Loptson (2004: 61). A weakness of his argument is that it only affects the memory criterion for personal identity at one time. It does not affect the memory criterion for personal identity over time.

I am the same person as Nestor. But if Locke accepted (2) this would yield a contradiction. Since I am at present not conscious of any of the actions of Nestor, (2) implies that I am not the same person as Nestor, which is a contradiction.

We find then that Locke either presents a counterexample to his own account or he does not accept (2). In the light of the absence of textual support for the claim that Locke accepts (2), it would be extremely uncharitable to conclude that Locke contradicts himself. We should instead conclude from the Nestor example that Locke does not hold (2). Therefore he does not accept MP, since MP implies (2). The traditional interpretation seems to be mistaken.

3. The memory continuity criterion of personal identity

So if MP is not Locke's view, we need another interpretation of his theory of personal identity. In this section I will argue that Locke defended the following memory sequence criterion of personal identity:

> *The Memory Continuity Criterion of Personal Identity (MCP)*: A person P_1 who exists at t_1 is identical to a person P_2 who exists at a later time t_2 iff there exists a memory sequence from P_1 at t_1 to P_2 at t_2,

where memory sequence is defined as follows:

> *Memory sequence*: There exists a memory sequence from a person P_1 that exists at t_1 to a person P_2 that exists at a time t_2 iff there exists a sequence of a person (or persons) at different times starting with P_1 at t_1 and ending with P_2 at t_2 such that for all adjacent times in the sequence the person at the later time

remembers any of the thoughts or actions of the person at the earlier time.

Note that a memory sequence from P_1 at t_1 to P_2 at t_2 need not include any intermediary elements. If you suddenly woke up one morning t_2 remembering one of Nestor's actions at t_1 then there is a memory sequence connecting you and Nestor, for example, a sequence with just two elements, you at t_2 and Nestor at t_1. Furthermore note the symmetry that if there exists a memory sequence connecting P_1 at t_1 with P_2 at t_2 there also exists a memory sequence connecting P_2 at t_2 with P_1 at t_1.

A similar interpretation has previously been suggested by Jane Lipsky McIntyre.[15] Peter Lopston is not convinced by this interpretation as he claims McIntyre has not provided sufficient support from the Lockian text for the interpretation.[16] However, in his recent paper on the 'Fatal Error' passage, Don Garrett argues that a memory continuity interpretation of Locke is strongly suggested by Locke's discussion of two cases in which personal identity fails despite identity of thinking substance or man.[17] In this section I will go further and show that the memory continuity interpretation follows implicitly from a couple of passages in Locke.

My strategy will be first to find textual support for a sufficient condition for consciousness identity, and then show that there is also textual support for the claim that Locke regarded this condition as necessary. Since Locke held that personal identity consists in sameness of consciousness, we will then have a condition Locke regards as both necessary and sufficient for personal identity.

Loptson holds that Locke did not provide a necessary and sufficient criterion for sameness of consciousness. Furthermore, we

[15] McIntyre (1977: 126).

[16] Loptson (2004: 51).

[17] Garrett (2003: 110).

cannot, according to Lopston, have an 'independent notion of what numerical sameness of consciousness can be'.[18] Contrary to Loptson, I will argue that Locke has implicitly provided a necessary and sufficient criterion for sameness of consciousness.

There is, as argued in earlier sections, strong textual support for the claim that Locke held both CP and (1). The conjunction of CP and (1) implies that:

(4) If a person P_1 exists at t_1 that partakes of a consciousness C_1 at t_1 and a person P_2 exists at a later time t_2 that partakes of a consciousness C_2 at t_2 and P_2 is at t_2 conscious of (remembers) any of the thoughts or actions of P_1 at t_1 then C_1 is identical to C_2.

From (4) and the transitivity of identity we then have that Locke accepts the following sufficient criterion of consciousness identity:

(5) If a person P_1 exists at t_1 that partakes of a consciousness C_1 at t_1 and a person P_2 exists at a later time t_2 that partakes of a consciousness C_2 at t_2 and there exists a memory sequence between P_1 at t_1 and P_2 at t_2 then C_1 is identical to C_2.

This is because (4) implies that every person in a memory sequence partakes of the same consciousness as the person at the next and previous times in the sequence. By the transitivity of identity it then follows that the first person, P_1 at t_1, partakes of the same consciousness as the last person, P_2 at t_2.

Thus, the existence of a memory sequence from a person who partakes of consciousness C_1 to a person who partakes of C_2 is sufficient for Locke for the identity of C_1 and C_2. To see that it is also necessary we need to examine what makes Locke, in §23, conclude that a consciousness is distinct from another consciousness.

[18] Loptson (2004: 62–3).

For granting that the thinking Substance in Man must be necessarily suppos'd immaterial, 'tis evident, that immaterial thinking thing may sometimes part with its past consciousness, and be restored to it again, as appears in the forgetfulness Men often have of their past Actions, and the Mind many times recovers the memory of a past consciousness, which it had lost for twenty Years together. Make these intervals of Memory and Forgetfulness to take their turns regularly by Day and Night, and you have two Persons with the same immaterial Spirit, as much as in the former instance two Persons with the same Body.

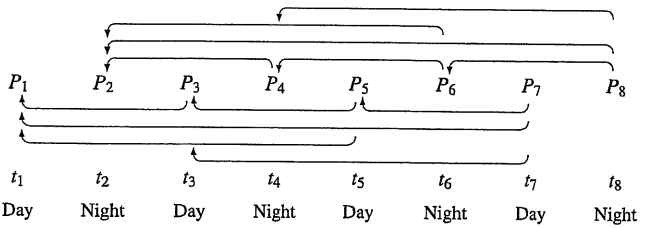

Figure 1: Intervals of memory and forgetfulness taking their turns regularly by day.

Figure 1 illustrates a case like the one Locke describes, which makes 'intervals of Memory and Forgetfulness to take their turns regularly by Day and Night'. In the figure there are eight different points in time alternating between day and night with a person existing at each. An arrow denotes that the person the arrow points from remembers the thoughts or actions of the person the arrow points to. Locke presents the case with intervals of memory and forgetfulness as an example of a case where there are 'two distinct incommunicable consciousnesses acting the same Body, the one constantly by Day, the other by Night' (ibid.). That there is not any night person that remembers any of the thoughts or actions of any day person is apparently sufficient for Locke to conclude that there is not any night person that partakes of the same consciousness as any day person. It follows from the definition of memory sequence that another sufficient condition for there not being any night person who remembers the thoughts or actions of any day person is that

there does not exist any memory sequence from a night person to a day person.

Locke seems to hold that if there is a set of persons at different times S_1, for example, the day persons, and another set S_2, for example, the night persons, and there does not exist any memory sequence from a person at time in S_1 to a person at a time in S_2 then no person in S_1 partakes of the same consciousness as a person in S_2. This implies the following special case where each set has only one member:

(6) If a person P_1 exists at t_1 that partakes of a consciousness C_1 at t_1 and a person P_2 exists at a later time t_2 that partakes of a consciousness C_2 at t_2 and there does not exist a memory sequence from P_1 at t_1 to P_2 at t_2 then C_1 is not identical to C_2.

We have then that Locke accepts (5) and (6), which implies the following necessary and sufficient condition of consciousness identity:

> *The Memory Continuity Criterion of Consciousness Identity (MCC)*: C_1 is identical to C_2 iff there exists a person P_1 at a time t_1 that partakes of a consciousness C_1 at t_1 and there exists a person P_2 at a later time t_2 that partakes of a consciousness C_2 at t_2 such that there exists a memory sequence from P_1 at t_1 to P_2 at t_2.

Finally, MCP follows from CP and MCC.

4. Surviving Reid's Gallant Officer

To illustrate the difference between MCP and MP we shall finally look at how MCP handles Reid's Gallant Officer. Since the officer

remembers having been flogged at school there exists a memory sequence from the boy to the officer and therefore the boy is the same person as the officer. Likewise, because the general remembers taking the standard, there exists a memory sequence from the officer to the general. Therefore the officer is the same person as the general. Since the general remembers a thought or action of the officer and the officer remembers a thought or action of the boy, there exists a memory sequence from the boy to the general. Therefore the boy is the same person as the general. The fact that the general does not remember any of the thoughts or actions of the boy does not, on the MCP interpretation, imply that the boy is not the same person as the general, since there still exists a memory sequence from the boy to the general. Thus Locke, on the MCP interpretation, handles the Gallant Officer case without contradiction.

Philosophers sympathetic to Locke's approach to personal identity have usually modified the theory to be able to handle cases like the Gallant Officer by the standard move of basing their account on a relation of a continuity of memories (or other psychological connections) rather than just memories. But if the MCP interpretation is right, Locke had already made this move.[19]

Royal Institute of Technology, Stockholm

References

Atherton, M. (1983), 'Locke's Theory of Personal Identity', *Midwest Studies in Philosophy*, 8: 273–93.

Berkeley, G. (1950), *The Works of George Berkeley: Bishop of Cloyne*, eds. A. A. Luce and T. E. Jessop, iii (London: Thomas Nelson).

[19] Thanks to Sven Ove Hansson, Jens Johansson, Martin Peterson, Tor Sandqvist, and an anonymous referee for *Locke Studies*.

Flew, A. (1951), 'Locke and the Problem of Personal Identity', *Philosophy*, 26: 53–68.
Garrett, D. (2003), 'Locke on Personal Identity, Consciousness, and "Fatal Errors"', *Philosophical Topics*, 31: 95.
Helm, P. (1979), 'Locke's Theory of Personal Identity', *Philosophy*, 54: 173–85.
Jenkins, J. J. (1983), *Understanding Locke: An Introduction to Philosophy through John Locke's Essay* (Edinburgh: EUP).
Locke, J. (1979a), *The Correspondence of John Locke*, ed. E. S. de Beer, iv (Oxford: Clarendon Press).
Locke, J. (1979b), *An Essay Concerning Human Understanding*, ed. P. H. Nidditch (Oxford: Clarendon Press).
Loptson, P. (2004), 'Locke, Reid, and Personal Identity', *The Philosophical Forum*, 35: 51–63.
Lowe, E. J. (1995), *Locke on Human Understanding* (New York: Routledge).
Mackie, J. L. (1976), *Problems from Locke* (Oxford: Clarendon Press).
McIntyre, J. L. (1977), 'Locke on Personal Identity: A Reexamination', *Philosophy Research Archives*, 3: 115–44.
Noonan, H. (2003), *Personal Identity* (London: Routledge, 2nd edition).
Parfit, D. (1984), *Reasons and Persons* (Oxford: Clarendon Press).
Reid, T. (2002), *Essays on the Intellectual Powers of Man*, ed. D. R. Brookes (Pennsylvania: Pennsylvania State University Press).
Winkler, K. P. (1991), 'Locke on Personal Identity', *Journal of the History of Philosophy*, 29: 201–26.