

The Problem with Memory Theories of Personal Identity

Introduction

In his book, *An Essay Concerning Human Understanding*, John Locke provides a non-substantial psychological account of personal identity.¹ For Locke, the sameness of consciousness is both necessary and sufficient for the relation of personal identity to obtain.² While this theory stands to be reasonable for most situations, there are several inconsistencies encountered when using memory as a phenomenon to determine personal identity. It is the case, as I will show, that a theory of personal identity constructed from a psychological phenomenon such as that of memory is only as dependable as the phenomenon of memory itself.

In this paper, I will first give a brief account of Locke's theory of personal identity. Following that, in my evaluation of the theory, I will discern the problems that one would inevitably face when using the phenomenon of memory continuity (as a consequence of the sameness of consciousness) as a criterion for the relation of personal identity to obtain. In conclusion, I would attempt to show how memory is not a reliable criterion for a theory of personal identity.

Locke's Theory of Personal Identity

Locke's theory of personal identity emerges from his own notion of the identity of objects in general. It is known that for any two objects under consideration to be the same object, they must have a certain set of identical properties (Leibniz's Law or the Indiscernibility of Identicals).

¹ Personal Identity, *Stanford Encyclopedia of Philosophy*. §4 <https://plato.stanford.edu/entries/identity-personal/>

² Locke, J. (1690). *An Essay Concerning Human Understanding*, Book II, Chapter XXVII, §§10, 17

Locke contends that for the objects under consideration to be identical they must have the same sort, time of origin, and place of origin. Here, by sort, Locke refers to the principle of constitution or kind of the object under consideration. Since different objects might have different definitions of the principle of constitution, it is the sort of the object that determines the relation of identity³.

Thus, by Locke's account when given two objects x and y , one may say that $x = y$ if and only if x and y have the same principle of constitution, and came to be at the same time and place.⁴ Stated more formally:

$$F_x = F_y \Rightarrow x = y \text{ where } F_x = (\text{time of origin of } x, \text{ place of origin of } x, \text{ sort of } x)$$

With this, Locke indicates that it is the uniqueness and continuity of an object that constructs the relation of identity for that object. He extends this notion of identity of objects to give an account of personal identity by providing the principle of constitution for persons.

Locke conjectures that it is in the characteristics of thinking, self-reflection over time, and intelligence of a person that the unique nature of a person can be found. He argues that all of these characteristics are explained by and inseparable from consciousness.⁵ As a result, it is the consciousness that a person possesses which explains their unique dispositions (with respect to their defining characteristics) both at a particular time and across time. Thus, Locke asserts that the principle of constitution of a person is determined by consciousness.⁶ By extension, since that which determines the principle of constitution of a thing also determines its identity, personal

³ Locke, J. (1690). *An Essay Concerning Human Understanding*, Book II, Chapter XXVII, §3

⁴ Locke, J. (1690). *An Essay Concerning Human Understanding*, Book II, Chapter XXVII, §1

⁵ Locke, J. (1690). *An Essay Concerning Human Understanding*, Book II, Chapter XXVII, §9

⁶ Locke, J. (1690). *An Essay Concerning Human Understanding*, Book II, Chapter XXVII, §9, 10, 17

identity is determined by consciousness. Thus, the continuity of consciousness implies the continuity of personal identity. To lay out this argument in steps:

1. The identity of an object is determined by that which explains its defining characteristics.
2. The defining characteristics of a person are thinking, self-reflection, and intelligence.
3. Consciousness explains these defining characteristics.
4. Consciousness determines personal identity.

With this, a relation of identity that is the core of Locke's theory of personal identity is established. In such an account for personal identity, one can say that a man A is the same person as a man B if and only if one of the two is conscious of being the other (that is, one of their consciousness extends backward to the other's consciousness).

It is important at this point to understand what exactly Locke means by the continuity of consciousness to fully understand his notion of personal identity. When Locke refers to a person as being conscious of a past self or talks of a person's consciousness extending backward, he indicates that the person possesses a memory of a past self.⁷ Thus, continuity of consciousness is nothing but links to the consciousness of the past self via memory of that self. In other words, if a person possesses the memories of their past and present actions, the actions belong to the same person, or the persons who performed those actions across time are the same.⁸

Thus, by Locke's account, one can say that a man A is the same person as a man B if and only if one of them recalls being the other, and if they have the same time and place of origin. This account of personal identity is a kind of a Memory Theory of personal identity.⁹

⁷ Locke, J. (1690). *An Essay Concerning Human Understanding*, Book II, Chapter XXVII, §9, 17

⁸ Locke, J. (1690). *An Essay Concerning Human Understanding*, Book II, Chapter XXVII, §16

⁹ John Locke, *Stanford Encyclopedia of Philosophy*. <https://plato.stanford.edu/entries/locke/supplement.html>

Inconsistencies with Locke's Theory of Identity

The first objection to Locke's Theory of Identity comes from Scottish philosopher Thomas Reid. Reid gives the example of a person at three stages in life to enunciate his issues with Locke's Theory.¹⁰ Consider a young boy who stole candy from a store. Suppose this boy grows up to become a great lawyer, and then after many years, retires from his position. If the man who has just retired remembers himself as a lawyer, by Locke's account, he is the same person as the lawyer. Similarly, if the lawyer remembers stealing candy from the store, the lawyer is the same person as the young boy. However, if the retired man does not remember the incident from his childhood, according to Locke's account it must mean that he is not the same person as that child. Herein lies a contradiction, for it is not possible for two things to be identical to a third thing and non-identical to each other (i.e., $b = a$, $c = a$, $b \neq c$ is a contradiction) as identity must be transitive.

While Locke never explicitly gives a response to this objection, Reid's objection falls short when the interpretation of Locke's theory is modified in a very small way by using indirect memory to link persons instead of direct memory. Consider carefully the previous example of the man to understand this better. It is known that the retired man is the same person as the lawyer as he possesses memories of being the lawyer. It is also known that the lawyer is the same person as the young boy as he possesses memories of being that young boy. Thus, by extension (appealing to the transitivity of identity), the retired person is also the same as the young boy — since the retired person has an indirect memory of being the young boy.¹¹

¹⁰ Reid on Memory and Personal Identity, *Stanford Encyclopedia of Philosophy*. §3 <https://plato.stanford.edu/entries/reid-memory-identity>

¹¹ Locke, J. (1690). *An Essay Concerning Human Understanding*, Book II, Chapter XXVII, §17

While this first objection is solved by modifying the theory a little in terms of its fine details, there still exist some issues with the theory which cannot be rectified so long as the theory uses memory as a tool to link the consciousness of various selves. That is to say that so long as Locke's theory of personal identity makes use of memory (be it direct or indirect), it is still susceptible to all the same problems that the human memory is susceptible to, which further complicates the matters for such a theory of personal identity.

For instance, consider the phenomenon of memory gaps. Take the example of a man who drank a lot. If the man does not remember a specific time frame from the previous night, it would be the case according to Locke's theory that the drunken man and the sober man are not the same person. However, this seems very absurd, as intuitively the man was just in different states of consciousness, and not two distinct persons. In response to this objection, Locke held his view strongly and indicated that he thought they weren't the same persons.¹²

Furthermore, as is hinted by modern psychology, the human mind is capable of generating quasi-memories to fill in the gaps of its own awareness.¹³ If such a thing is true, it is not inaccurate to say that human memories are rendered unreliable. One can find many such similar situations which demonstrate the inconsistency of human memory. In such a case it can be questioned as to which memories should be considered genuine. Moreover, there must be a criterion for choosing which of these memories should hold in order for the relation of identity to obtain depending on which memories are genuine. However, determining whether a memory is genuine requires knowledge outside of the person's principle of constitution, which, in an ideal

¹² Locke, J. (1690). *An Essay Concerning Human Understanding*, Book II, Chapter XXVII, §20

¹³ Personal Identity, *Encyclopedia Britannica*. Section on Traditional Criticisms <https://www.britannica.com/topic/personal-identity>

theory of personal identity, should not be required - one should be able to say that one is the same person in the present as another person in the past without consulting something external to the person himself.

Conclusion

The objections to Locke's theory of personal identity are not limited just to the theory's appeal to memory as a phenomenon that links consciousness across time. Even the use of other aspects consciousness (not just memory) or of consciousness itself can be faulty in certain scenarios. For example, consider a person who is in an unconscious state such as sleeping or a coma. Such a person is unaware of his previous self in a waking state in any way. According to Locke's theory, it would be the case that this person is not the same as the person who was awake, even if the idea of memory is left out of Locke's theory of personal identity since they have no consciousness of being the person in the past. Thus, Locke's theory leads us to an absurdity in such cases as well.

As such, these objections together indicate any theory of personal identity that is determined by a phenomenon of consciousness such as memory (or even consciousness itself) can only be as reliable as the phenomenon itself. This indicates that this is perhaps not the best way to proceed in order to develop a dependable theory of personal identity, and shows that there is a big drawback to Locke's theory of personal identity - perhaps even to all similar theories of personal identity.