Brent Emerson
13 May 1997

On Personal Identity

In Book 2, Chapter 27 of *An Essay Concerning Human Understanding*, Locke presents a view of personal identity that aims to clear the confusion surrounding what, exactly, it is that identifies someone as the same person over and over again. Throughout his discussion, he maintains one important foundational precept: "continued existence makes identity."(§29) Whether speaking of persons, vegetables, animals or substance, this is a guiding consideration: something maintains its identity (of whatever sort) through the continuity of that aspect of its existence.

That having been said, Locke considers three completely separate types of identity relevant to (and confused in) the question of what personal identity consists in. The first is *identity of substance*; two things can be said to be the same substance for Locke if and only if they consist of exactly the same bits of matter as one another. Identity of substance means perfect continued existence of the material body of something—a book resting on a shelf maintains its identity of substance until I come along and tear out a page, at which point it ceases to be the same substance it was before. The second type of identity is *identity of man*, which consists for Locke "in nothing but a participation of the same continued life, by constantly fleeting particles of matter, in succession vitally united to the same organized body."(§6) Identity of man means continued existence as the same organized living thing. Ultimately (§21), Locke leaves open the question of whether this means that being the same 'man' is being the same immaterial 'soul', being the same material 'animal', or both—identity of 'man' will be discussed later.

The most important of type identity for Locke is *identity of person*, which he suggests is the true nature of what we have called personal identity. Locke explains identity of person as "the sameness of a rational being: and as far back as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that *person*; it is the same *self* now it was then; and 'tis by the same *self* with this present one that now reflects on it, that that action was

done."(§9) Consciousness is the determining factor in Locke's view of personal identity; identity of person means continued existence of the same consciousness. I am the same person as myself of yesterday because I have memories of being that person. My body, having grown many new cells and shed and recycled others, is clearly not the same *substance* it was yesterday. Having maintained the same organization of a living being (being the same life-process it was yesterday) it is the body of the same *man*, and—having maintained the same consciousness—the body of the same *person*.

Locke spends several sections (§§10-17) emphasizing that personal identity should be understood as coincident with consciousness and *not* with substance. He considers several situations where the difference is emphasized, as in the removal of a body part—clearly the 'person' remains with the consciousness and not with the missing limb, now disconnected from the consciousness. Locke asks the question of whether one person may inhabit two different substances, as in the case of a person remembering being in a different body (or, for a less extreme example, as in the growth and change of the human body, which he neglects to mention); Locke concludes that the 'person' follows the consciousness and should be identified with both substances (each at the time the consciousness inhabits it). He considers whether one substance may be inhabited at different times by different persons, as in the case of one body (somehow an unchanging body, if it is to remain the same substance) claiming to be two different people; again, Locke suggests that we place the 'person' with the consciousness, taking the body to truly be two different people.

Locke clearly argues that personal identity should be identified with continuity of consciousness rather than continuity of substance; he defends the point well, and should probably be commended for making an important point in the history of western philosophy. However, there are some problems with the argument. The whole issue of the middle category of 'identity of man' is confusing—it seems to have been included merely to get around the question of lost memories, of gaps in the continuity of consciousness. Locke explains (in §20) that he is justified in using the term 'I' to refer to the agent of the forgotten actions of his body, even though it should then be

understood that he is referring to 'himself' as the same *man* and not the same *person.* This discussion makes unclear why 'I' should ever be used when not referring to the same *person*, which is strange in light of Locke's insistence everywhere else on the fundamental importance of the identity of person and nothing else. Locke may have included this category in order to make his philosophy seem less radically opposed to the law (wherein the sober person could be prosecuted for the crimes of the drunk person(§22)), but he could have gotten around this in some other fashion, perhaps by understanding the sober person's choice to drink as a choice to give his body to another, less careful person (the drunk person), thereby implicating the sober person in the crimes as well.

More important, though, is a larger objection. To bring his project to its fullest conclusion, Locke could have gone farther: he has succeeded, with several thought-experiments, in unseating both *substance* and *man* as the site of 'person'. Taken further, his argument can reveal *consciousness* also to be too narrow a site for 'person' and render problematic the category of the continuing 'self', reveal as incoherent the entire concept of 'personal identity.'

Let's consider a more complicated thought-experiment than Locke does. (I will use more modern terminology than Locke, but the same experiment could well be formulated in terms he himself uses.) Imagine someone (John) who is about to go on a long spaceflight to a remote planet. The trip is much too long for a human body to survive, so John's consciousness will be emulated in a computer and stored there for transport, to be re-instituted into another human body on arrival. John is naturally somewhat afraid of this complex operation, but he is assured that the process will seem no different than waking up after a period of sleep (and the trip will be instantaneous from his point of view), so he submits to the emulation. After the spaceship has left Earth, the people in control of the operation decide to kill John, since his consciousness is safely on the way to a new body anyway. When told of these plans, John is naturally alarmed. Is there any sense in which we can expect John to relax into his death, secure in the knowledge that he will soon 'reawaken' on a remote planet? Can we expect John to identify with his yet-to-be-instituted 'self' billions of miles away? Of course not. In what sense then can they be understood as the same

'person'? (To see this more clearly, imagine that John, instead of being killed after emulation, is allowed to live on Earth while a copy of his consciousness travels and is eventually reinstituted as $John_2$ on a remote planet. There is clearly no sense in which the two Johns share the same consciousness—other than their common memories, $John_1$ is as separate from $John_2$ as $John_1$ is from anyone else on Earth. Now, under these new circumstances, ask $John_1$ to volunteer for death—he will surely be at least as alarmed as he was before. Certainly $John_1$'s memories from before the emulation will 'live on' in $John_2$, but is there any sense in which $John_1$—the subjective notion of 'person' that is $John_1$—will be preserved in $John_2$? No, of course not, no more than it would be if John were killed and expected to 'reawaken in' his brother or his next-door neighbor.)

It seems from this experiment that all we need to do to restore a viable notion of 'person' is to require that the consciousness in which 'personhood' consists be continuous—certainly it is appropriate to consider John's life before the trip the life of a unified 'person'? Assume, then, that the spaceship crashes into the moon hours after takeoff and, with the emulation destroyed, John is not killed after all. Instead, he goes home, eats dinner, and prepares for welcome sleep at the end of a very hectic day. Lying in the dark, John has a terrible realization: he has again been placed in the situation that today frightened him so much. He is about to cede his consciousness (slipping into sleep) in anticipation of reawakening. When he reawakens, he will have exactly the same information that $John_2$ would have had awakening on the remote planet: only memories to connect him with the John of the past and convince him that he is the same person. John can't figure out why this is any different, simply because *there is no difference*. Every time John goes to sleep, he enters into essentially the same situation that seemed so technologically fantastic above. And if nightly sleep isn't universal enough, imagine John sitting all night at his desk, unable to sleep. Every few moments (perhaps) he realizes that a collection of memories is all he has to convince himself that he is, in fact, the same John he was a few moments ago.

Clearly, it makes less sense to characterize John as a unified 'person' than as a succession of *person-states*, one following another, a new one generated every day (or perhaps every instant) each convinced by memory that it is the same unified 'person' as the state that immediately

preceded it. As I have shown, however, each such ‘person-state’ has no such connection to the state that immediately follows it—the time-asymmetry of memory renders the relation between these states of consciousness intransitive. Since transitivity is a fundamental element of an equivalence relation like identity, these person-states clearly cannot coherently be assembled into a chain of anything resembling equivalence or identity—hence the idea of ‘personal identity’ is fundamentally incoherent.