

# Locke's Theory of Identity

DAN KAUFMAN

## 12.1 Introduction

It would be difficult to overstate the importance of Locke's theory of identity in Book Two, Chapter 27 of his *Essay*. His theory of personal identity, in particular, in terms of its philosophical content, its originality, the way it reshaped a significant philosophical issue, and its influence on Locke's contemporaries and *our* contemporaries, stands as one of the truly exceptional contributions to philosophy and the history of philosophy.

However, without some friendly nudging, Locke may never have written it, and it is the story of the origin of 2.27, as a paradigmatic case of a lucky historical accident, with which I wish to begin. It is well known that 2.27 was not in the December 1689 First Edition of the *Essay*, and that it made its first appearance in the 1694 Second Edition. What is less well known is how this came to be. With the exception of a note from 5 June 1683 – in which Locke gives hints of the theory of personal identity that would be fleshed out in 2.27 – and a brief discussion in 2.1 (whose main point is not personal identity but rather an attack on Descartes' notion that the nature of mind is thought), Locke did not write much on identity prior to 2.27. Identity is not discussed in the early drafts of the *Essay* – “Draft A” (1671), “Draft B” (1671), and “Draft C” (1685), though Draft C does contain some discussion of consciousness, pleasure and pain, and personal identity. Nor is identity discussed in any detail in the First Edition of the *Essay*. On 2 March 1693, William Molyneux wrote to Locke, and suggested that Locke say something about two issues neglected in the First Edition, namely the eternal truths and the *principium individuationis* or “the principle of individuation.” On 23 August 1693, just a few months after Molyneux's request, Locke replies: “you

will herewith receive a new chapter Of identity and diversity, which, *having writ only at your instance, 'tis fit you should see and judge of it before it goes to the press*" (Corr. 1655; emphasis mine). This suggests that 2.27 was written in a few months; and the result reflects that, for all of its brilliance, it is occasionally sloppy and perhaps worse. On the other hand, the *most general* themes of Locke's theory of personal identity were lurking, scattered, but by no means presented in any systematic or detailed manner, by the early 1680s.

Among the reasons for my assertion of the importance of 2.27, four warrant mention here. First, 2.27 is the earliest detailed account of identity written by a well-known proponent of the "new mechanical philosophy." Second, Locke is the best representative of what I take to be a more general shift among early-modern philosophers, one that stands in contrast to the "traditional" problems of individuation and identity as addressed by scholastic philosophers. Third, unlike nearly all of the important and interesting views of early-modern philosophers, many of which are treated as interesting but quaint and naïve views of people who did not know better, Locke's theory of identity not only provoked a strong reaction from his contemporaries and near-contemporaries, it continues to influence philosophical discussions of identity to the present day. Harold Noonan has it dead to rights when he states: "It has been said that all subsequent philosophy consists merely of footnotes to Plato. On this topic [i.e. personal identity], at least, it can be truly said that all subsequent writing has consisted merely of footnotes to Locke" (1989, 24). The fourth, and philosophically important reason is Locke's rejection of "substance-based" theories of diachronic identity, which I will discuss later in this chapter.

In 2.27.3, when Locke boldly claims "'tis easy to discover, what is so much enquired after, the *principium Individuationis*," he *seems* to be joining a tradition whose origins go at least as far back as Boethius in the early sixth century. But is he? To consider this question, we should distinguish *individuation* (which includes the notions of unity, number, distinction) from *persistence* or identity over time. For the purposes of this paper, *persistence conditions* will be informative, non-circular conditions, necessary and jointly sufficient, for something at one time to be numerically identical to something at another time. They will be substitution instances for *C* in this general schema:

$x$  at  $t_1$  is numerically the same as  $y$  at  $t_2$  iff  $C$ .

And, of course, much of 2.27 is an attempt to provide persistence condition for various kinds of things. But when Locke announces the *principium individuationis*, he is not yet concerned with persistence over time; he is concerned with individuation and that in virtue of which a thing is *a* thing, an individual numerically distinct from other things.

The individuation conditions for any substance involve two aspects: one is the spatio-temporal origin of the substance and the other is the notion that a substance precludes spatio-temporal colocation of any distinct substances of the same *kind*.

Locke appears to give three “principles” of individuation, which I will refer to as “L1,” “L2,” and “L3.” But it is clear from the discussion that L1 is the “primary” principle, perhaps axiomatic, from which the other two follow.

- L1: It is impossible for two (or more things) of *the same kind* to be in place,  $p$ , at time,  $t$ .
- L2: It is impossible for one thing to have more than one beginning.
- L3: It is impossible for two or more things (of the same kind) to have one beginning.

These principles, it should be noted, do not rule out collocation altogether, but only collocation of things of the same kind. So, even at this early stage of the discussion, Locke calls attention to the crucial role played by *kinds* in individuation and identity. In fact, what has made the issue of identity problematic, according to Locke, is carelessness concerning the *kind* (“precise Notions” or “specifick Ideas”) of things in question. Both the *principium individuationis* and the persistence conditions of an individual or thing will be determined by what kind of thing it is. As Locke states:

we must consider what *Idea* [i.e. nominal essence] the Word it is applied to stands for: It being one thing to be the same *Substance*, another the same *Man*, and a third the same *Person*, if *Person*, *Man*, and *Substance* are three Names standing for three different *Ideas*; for such as is the *Idea* belonging to that Name, such must be the *Identity*.

E.J. Lowe nicely observes: “Locke’s great insight is that the identity conditions [i.e. persistence conditions] of things of one sort need not be the same as the identity conditions of things of another sort” (Lowe 2005, 88–9).

Unfortunately, a detailed discussion of Locke’s view of kinds is beyond the scope of this paper, but something, however brief, must be said here because of the central role kinds play in the theory of identity. Locke, as I read him, holds that there are no mind-independent natural kinds, nor are there species-delimiting entities such as substantial forms. The world gives us only observable objective-similarities between individuals. Kinds or species, however, are merely nominal essences or general abstract ideas which we form on the basis of the given objective similarities between individuals. Locke famously distinguishes between “real essences” and “nominal essences,” and it is the latter that are Lockean kinds, both in the context of 2.27 and elsewhere. The real essence of a material substance is that unknown arrangement of corpuscles responsible for *all* of that body’s observable qualities. Now suppose that there is a body  $B$ , and  $B$  has observable qualities  $Q_1–Q_{10}$ . The real essence of  $B$  is responsible for  $B$  having  $Q_1–Q_{10}$ . Suppose there is another body  $B^*$ , and  $B^*$  bears some objective similarity to  $B$ ; that is,  $B$  and  $B^*$  have some observable qualities in common. On the basis of the objective similarity between  $B$  and  $B^*$ , we abstract away just those – and not necessarily *all* of those – objective similarities. The abstract idea we form by including those qualities and excluding qualities that  $B$  and  $B^*$  do not have in common is what Locke calls a *nominal essence*. More precisely, it would be the nominal essence of

some kind *K*, and *K* then has *B* and *B*\* as members. Locke's attempt to be more precise about "notions" and 'Specifick ideas," when giving his theory of identity, consists in recognizing different kinds of things and not confusing the nominal essences.

## 12.2 Easy Cases: God, Finite Intelligences, Atoms, and Masses

After introducing L1–L3, Locke goes rather quickly through those things whose individuation and persistence conditions are unproblematic – at least, according to Locke. God's identity is unproblematic because He is immutable, and the issue of identity simply does not arise for a being that cannot change. (In fact, it would not be wrong to think of a theory of diachronic identity as a theory about how things persist *idem numero* despite their various intrinsic changes.) What Locke calls "Finite Intelligences" (i.e. thinking substances) also present an easy case, presumably because they are simple in the sense of lacking any proper parts or, at least, any "discerpi-ble" parts. Like most of his predecessors and contemporaries, Locke thinks that finite intelligences have location/place, as well as temporal location. So, they too cannot be colocated. The fact that finite intelligences have location plays a crucial role in Locke's individuation of persons, as we shall see.

Some bodies, despite having proper parts, are easy cases, too. These are atoms and masses of atoms. The former, not because they are mereological simples (i.e. lacking proper parts), but because they have an impenetrable and immutable surface which eliminates the (natural) possibility of a change of parts or a discerning of parts. With respect to masses of atoms, Locke is a mereological essentialist: that is to say, a mass of atoms has all of its parts essentially, so that a mass could not have been composed of a different sum of atoms, nor could a mass survive any change in its parts – though a mass can survive any amount of internal rearrangement.

## 12.3 A Less Easy but More Interesting Case: Organisms

The issues become stickier when Locke turns to the individuation and persistence of organisms (living things, *vivens*), precisely because organisms persist *by* changing parts, because the biological life, which will constitute the identity of an organism, includes the processes governing an organism's change of parts. It is simply part of the nominal essence of an organism that it in fact loses and gains parts. So, far from preventing the diachronic identity of an organism, changes in parts make possible the diachronic identity of an organism.

As we shall see later, Locke's attack on substance-based theories of identity focuses mainly on theories of personal identity in which sameness of a thinking substance is necessary and sufficient for personal identity. Locke's view of the identity of organisms, however, already rejects the notion that sameness of corporeal substance

is either necessary or sufficient for the sameness of the organism. Because Locke holds that organisms persist and yet it is undeniable that the mass of matter that constitutes the organism at a time does not persist, Locke must look elsewhere to account for the organism's persistence. According to Locke, the same organism "is the same continued Life communicated to different Particles of Matter, as they happen successively to be united to that organiz'd living Body" (2.27.8). It is in virtue of the life functions and biological functions of one organized body (i.e. a body with organs, properly disposed to fulfill the biological functions of an organism) being maintained by gradually changed but properly disposed bodies that an organism persists over time.

This view of the persistence of organisms entails some things about the mereological composition of an organism: an organism persists through a change of parts only if *either* the new parts are "negligible" (i.e. they replaced parts that were functionally irrelevant, e.g. an atom at the tip of a thumbnail), *or* the new parts are functionally isomorphic to the parts they replaced and allow for the same Life to be instantiated as before.

Although they are only mentioned briefly, Locke states two further conditions for the identity of organisms and artifacts. First, there is what I will call the *Continuity Condition*, according to which  $x$  at  $t_1$  is the same organism as  $y$  at  $t_2$  only if that organism existed at every time between  $t_1$  and  $t_2$  (2.27.4). Second, there is the *Gradual Replacement Condition*, according to which the change in an organism's parts must be sufficiently gradual, or at the very least "not shifted all at once" (2.27.8). Both of these conditions reflect what Locke takes the nominal essence of *organism* to entail: something that could cease to exist altogether for a period of time and then come back into existence just wouldn't fall under the nominal essence of *organism*; organisms are simply not like that. Also, because organisms grow, diminish, metabolize, and slough off parts during their careers, anything that didn't do these things just wouldn't be an organism. That is, although organisms can persist through a change in parts, they cannot persist if their parts are "shifted all at once" (2.27.8).

## 12.4 A Problem about Coincidence

This problem, which finds its most forceful statement in Chappell (1989, 1990), is as follows: grant L1, L2 and Locke's account of the persistence conditions for masses and organisms. First, it follows that a mass of matter at  $t$  and the organism it "constitutes" at  $t$  are not numerically identical to one another. A mass of matter, in Locke's view, is simply a "conjoined" aggregate of atoms, and a mass persists solely in virtue of having all and only the same atoms conjoined. So, suppose we have a mass  $M_1$  that constitutes an elm tree  $E_1$  at  $t_1$ . Now,  $M_1$  ceases to exist at  $t_2$  due to loss of an atom. Thus,  $M_2$ , a numerically distinct mass, now constitutes the persisting  $E_1$ . According to L2, one thing cannot have two beginnings.  $M_2$  begins to exist at  $t_2$ , but  $E_1$  began to exist earlier than  $t_2$ . Therefore,  $M_2 \neq E_1$ . Generalizing from this, we can arrive at

the conclusion that all organisms and artifacts are not numerically identical to the masses that constitute them at any given time. But this, by itself, is not problematic for Locke as long as the two things are “of different kinds.” And a natural response is: “yes, they are. One is a mass, and the other is a tree/organism. Those are different kinds. So there is no violation of L1.” However, at least in 2.27.2, Locke seems to commit himself to the view that there are only *three* kinds of substances: God, finite intelligences, and bodies. So, there seem to be two bodies in the same place at the same time, which would violate L1, *or* there are more than three relevant kinds of substances, *or* the organism is simply not a substance (see Kaufman 2007, 2008).

Although I cannot do justice to the richness of the scholarship on this problem here, I will mention a few of the interpretive strategies found in the literature.

#### 12.4.1 *The Coinciding Substances Interpretation*

According to Vere Chappell and Ken Winkler, Locke’s position is that it is possible (and, in the case of the mass and elm, actual) that numerically distinct bodies/corporeal substances occupy the same place at the same time. This, of course, requires a few things: that Locke has a more fine-grained view of kinds; that he holds that *mass* and *organism* must be of different kinds; and that these different kinds allow for colocation.

#### 12.4.2 *Relative Identity Interpretation*

According to this interpretation, defended most recently by Matthew Stuart, identity is not “absolute” but *relative to a sortal* (in fact, it was Locke who coined the term “sortal”). So, the proper form of a statement of diachronic identity is not “ $x$  at  $t_1$  is the same as  $y$  at  $t_2$ ”, but rather “ $x$  at  $t_1$  is the same  $F$  as  $y$  at  $t_2$ ,” where the sortal “ $F$ ” refers to a particular nominal essence. Locke, then, does not have a problem with coincidence because he may simply say that the thing at  $t_2$  is *the same elm tree* as the one existing at  $t_1$ , but it is *not the same mass*. It is true that Locke thinks that we must consider *kinds* when giving persistence conditions, but whether this commits him to any stronger claim about the relativity of identity is a matter of significant disagreement.

#### 12.4.3 *The Mode Interpretation*

Perhaps Locke avoids the coincidence problem by holding that organisms (and persons) are simply not substances at all. Instead, these things are complex *modes*. In our puzzle case, the mass is a substance, but the elm tree merely a mode of that substance. Not only are the kinds *substance* and *mode* sufficiently different to allow colocation, generally speaking, there is not much objectionable to the view that a substance and a mode of that substance are collocated. In fact, it would be objectionable to say that

a substance and its mode are *not* colocated. Proponents of this view include Edmund Law (1823), Uzgalis (1990), Bolton (1994), LoLordo (2012), and Lowe (2005).

#### 12.4.4 Lockean Four-Dimensionalism

Christopher Conn, the most prominent advocate of this interpretation, argues that a four-dimensionalist ontology best captures what might be going on in Locke's thinking on persistence and the relationship between persons (or organisms) and finite intelligences and bodies. Organisms, according to the four-dimensionalist reading, are perduring, rather than enduring, entities – a distinction whose *locus classicus* is David Lewis' *On the Plurality of Worlds* (1996). A tree, for instance, is a sum of successive temporal parts (masses) linked together in the appropriate fashion – in the case of a tree, by each successive mass “taking on” the previous mass’ “Life.” In the case of personal identity, persons are collections of temporal parts (in this case, finite intelligence/thinking substances). If Locke were a four-dimensionalist, then coincidence is not a problem: the mass of matter that constitutes a tree at  $t$  is merely a proper temporal part of the perduring tree, and surely there is no problem with the partial overlap of part and whole; in fact, partial overlap *just is* parthood. Moreover, in the case of persons, Locke seems to think that consciousness of my thumb is what makes my thumb a *part* of me. But consciousness doesn't merely unite my spatial parts; it unites my existence at different times. The four-dimensionalist could appeal to the former case, in which consciousness unites spatial parts of me, to draw the conclusion that Locke thinks that consciousness unites temporal parts of me.

#### 12.4.5 Defeatism

According to this view, which I have argued for elsewhere, there is no interpretation of Locke in which his view concerning the relationship between masses and organisms is consistent. Even granting that Locke accepts a more fine-grained ontology of kinds (which includes more than the three kinds of substances listed in 2.27.2), a consistent reading of Locke's view of the relationship between the mass and organism requires, at the very least, that they differ in kind. In Lockean terms, it would need to be possible that  $x$  and  $y$  have numerically the same real essence at  $t$  but have different nominal essences at  $t$ . This is impossible, according to Locke, or so I argue elsewhere (Kaufman 2007).

### 12.5 The Least Easy but Most Interesting Case: Persons

Personal identity receives the lion's share of attention in 2.27. This fact is perfectly understandable because what a person is and how a person persists *idem numero* has importance even outside the metaphysics of identity. The general resurrection, one's

eternal future, as well as the reward or punishment and pleasure or pain that comes with it, is grounded completely in one's personal identity.

Although personal identity has a significance that is not present in the earlier cases of atoms, masses, and organisms, Locke's discussion of personal identity begins in exactly same manner as the earlier cases, namely: "we must consider what *Person* stands for." According to Locke, a person is:

[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; which it does only by that consciousness, which is inseparable from thinking, and as it seems to me essential to it. (2.27.9)

Locke's account of persons, among other things, clearly allows that there may be human beings ("men") who are not persons, as well as persons who are not human beings. The *severely* mentally disabled, the "dull irrational man" in 2.27.8, young children, babies, and perhaps even fetuses are certainly human beings in Locke's sense. But, lacking the relevant features above, they are not Lockean persons. On the other hand, it is also possible that all sorts of non-human entities are Lockean persons – intelligent extraterrestrials, the homicidal computer HAL from *2001: A Space Odyssey*, time-traveling dolphins from the not too distant future, angels, etc.

Readers of Locke tend to focus on the issue of personal identity *qua* problem about persistence over time. This is understandable, given that this is what Locke spends most of the time examining. However, in line with his general approach, Locke first shows how his "principles of individuation," in particular L1, apply also to persons.

It being impossible for any one to perceive, without perceiving, that he does perceive ... Thus it is always as to our present Sensations and Perceptions: And by this every one is to himself, that which he calls *self* ... For since consciousness always accompanies thinking, and 'tis that, that makes every one to be, what he calls *self*; and **thereby distinguishes himself from all other thinking things.** (2.27.9, emphasis mine)

For instance, *I* am conscious of *my* thinking at 1:22 pm in my office chair, and according to L1, there cannot be more than one thing of the same kind in the same place at the same time. Thus there cannot be more than one conscious, thinking being in my office chair at 1:22 pm. For any time in the career of a Lockean person, that person is thinking and conscious of her thinking. According to Locke, this makes the person herself *to her-self*; whenever there is a case in which someone is herself *to her-self*, there is a person then and there. Therefore, *I* am the only possible thinking thing in this chair at 1:22. I am individuated and distinctified "from all other thinking things" by L1. And anything, at any other time, bearing the appropriate sort of relation to the thinking thing in my office chair at 1:22 pm is *me*.

What is that relation? A common response is that Locke holds that the latter thing must *remember* sitting in the chair at 1:22 pm, appealing to Locke's words that "as far as this consciousness can be extended backwards to any past Action or Thought,

so far reaches the Identity of that *Person*" (2.27.9). According to this *Simple Memory Interpretation*, personal identity over time consists in a relation of memory between a later person and an earlier person. However, the Simple Memory Interpretation is now largely believed to be incorrect because, among other reasons, it takes consciousness simply to be memory. But this is far from Locke's actual view. Locke thinks that a person includes everything to which their consciousness "extends," and a (human) person's consciousness extends to include parts of her human body. Locke has us consider:

our very Bodies, all whose Particles, whilst vitally united to this same conscious self, so that we feel when we are touch'd, and are affected by, and conscious of good or harm that happens to them *are a part of our selves: i.e.* of our thinking conscious self. Thus the Limbs of his Body is to every one a part of *himself*: He sympathizes and is concerned for them. (2.27.11)

And: "Thus everyone finds, that whilst comprehended under that consciousness, the little Finger is as much a part of it *self*, as what is most so" (2.27.17). Moreover, Locke is clear that the consciousness most relevant to personal identity is my *present* consciousness. Although memory may be the kind of present consciousness that can link my present self to an earlier self (by "extending backwards"), memory is *merely that kind of consciousness*.

Locke clearly has a "consciousness-based theory" of personal identity and is sometimes credited with originating the "psychological continuity theory" of personal identity. There is, however, an obvious sense in which Locke does not hold a *continuous* consciousness theory. For instance, as he states, consciousness is often "interrupt'd by forgetfulness." Also, Locke's thought experiments in 2.27 (e.g. Dayman/Nightman, Socrates-waking/Socrates-sleeping) do not make any sense unless consciousness can be non-continuous or gappy. A better label for Locke's theory is a "Connected Consciousness Theory."

Despite the gappyness of consciousness, Ed McCann (1987) is completely right when he says: "Consciousness is the Life of Persons." In an analogy that is "at the heart of Locke's theory of personal identity" (Noonan 1989), Locke states:

Different Substances, by the same consciousness (where they do partake in it) being united into one Person; as well as different Bodies, by the same Life are united into one Animal, whose *Identity* is preserved, in that change of Substances, by the unity of one continued Life. (2.27.10)

Consider the two roles that Life plays in the identity of an organism. There is the synchronic-unifier role: an organism's Life makes any material part that is included in the functional organization of the living body *a part of the organism*. Life also plays a role in the diachronic identity of the organism: the same Life inherited by similarly arranged but numerically distinct, successive masses is both necessary

and sufficient for the same organism. In personal identity, consciousness plays these two roles: synchronically, consciousness unites all of a person's parts and concurrent psychological states; diachronically, consciousness links numerically distinct thinking substances into one persisting person.

Now, even if Locke were never to have composed 2.27, his views about the connection between thinking and consciousness would be well known from his attack on Cartesianism in the First Edition (1689) of the *Essay*. For instance:

I do not say there is no Soul in a Man, because he is not sensible of it in his sleep; But I do say, he cannot think at any time waking or sleeping, without being sensible of it. Our being sensible of it is not necessary to any thing, but to our thoughts; and to them it is; and to them it will always be necessary, till we can think without being conscious of it. (2.1.10)

A thinker “*must necessarily be conscious of its own Perceptions*” (2.1.12). Thus, it is natural for Locke to hold that, if persons are thinking intelligent beings, then the consciousness that is essential to thinking will be that which unites a person diachronically, “as far as ever it can be extended, should it be to Ages past, unites Existences, and Actions, very remote in time, into the same Person” (2.27.16). And:

For since consciousness always accompanies thinking, and 'tis that, that makes every one to be, what he calls *self*; and thereby distinguishes himself from all other thinking things, in this alone consists *personal Identity*. (2.27.9)

Another point about consciousness: a person's consciousness is always consciousness of a particular thing, action, or thought. I am conscious of my current sensation of warmth, the array of colors before me, and the song I am listening to (see also Strawson 2011). I am also conscious of earlier events, such as riding my bike yesterday, eating a burrito for dinner last night, and so on. These facts entail that *I* am the person who is warm, seeing colors, hearing music, and who rode his bike and ate a burrito. All of the possible “person-preserving links” of consciousness are links to particular thoughts or acts (see 2.27.9, 2.27.10, 2.27.14).

An additional feature of Locke's Connected Consciousness theory is that because persons (contingently?) lack foresight or consciousness of their future thoughts and actions, they are not presently identical to any future person. Locke's view, however, would allow the possibility of a present person being the same person as a future person if only she were able to be conscious of that future action in the same way that she is conscious of her present and past actions and thoughts. Locke cryptically mentions this in passing:

For it is by the consciousness it has of present Thoughts and Actions, that it is *self* to it *self* right now, and so will be the same *self* as far as the same consciousness can extend to Actions past *or to come*. (2.27.10, emphasis mine)

It is only because a person, for whatever contingent reason (not some conceptual barrier or metaphysical impossibility), can extend consciousness to past actions but not to future actions that prevents her identity with a future person, contrary to a recent suggestion by LoLordo (2012).

A person's consciousness brings with it a special kind of "attitude," which Locke calls "concernment" or "concern." Persons are concerned about their parts, thoughts, and actions in a way unlike their attitude toward other things. Consider the psychological relationship I have with my currently attached left thumb. The thumb is a part of my *self* in virtue of my being conscious of it insofar as it is a source of sensations and a cause of (potential) pleasure or pain. So, when I see that my thumb is about to be crushed by a hydraulic press, I am concerned for it; I very quickly move my thumb out of harm's way, anticipating the exquisite pain that this part of my self was about to cause in me. Contrast that concernment with my psychological relationship to Deaton's thumb. Of course, if I see that my friend's thumb is about to be crushed by the press, I have all sorts of psychological reactions and attitudes. But, because what happens to Deaton's thumb does not cause sensations like pain in me, I am not *concerned*, in Locke's sense, about the fate of Deaton's thumb. Locke goes even farther: if my thumb were cut off, resulting in my no longer being conscious of it, I wouldn't have concernment for it "any more than the remotest part of Matter." If being separated from my consciousness renders a body part that was once mine no longer a part of my self, then presumably the same is true of any part that *will be* mine but which is currently not something of which I am conscious.

Persons also have this special concernment for their actions and thoughts, both past and present. This is because their actions and thoughts, too, are potential sources of pleasure and pain, and reward and punishment (see 2.27.17).

Locke's theory of personal identity, though grounded completely in consciousness, stretches out to other things that bear certain quasi-logical relations to consciousness, e.g. Appropriation, Concernment. Despite the claims of some scholars, the notion of Appropriation is not the same as or coextensive with Consciousness. Of course, *all* of the objects of my consciousness are objects that are Appropriated by my self, but this is the case because my Consciousness is not just my awareness of an object; it is awareness that *I* am aware of an object. The first-person perspective is an essential feature of Consciousness. However, it is possible for a person to appropriate an action or thought without having any Consciousness of that action or thought.

We have already seen that a person's Consciousness brings with it Concernment. More precisely, Consciousness of potential or actual pleasure or pain *entails* Concernment. My Consciousness extends to my thumb, and I am conscious that my thumb is a source of potential pleasure or pain for my self. I am also Conscious of my past thoughts and actions, and this leads to my Concernment about those thoughts and actions being a source of pleasure or pain (2.27.17), including the pleasure or pain that is involved with reward and punishment. My past acts, insofar and only insofar as

I am conscious of them, are potential sources of Reward or Punishment. As a person, I had better be concerned about this! After all, my happiness and misery (“consisting only in pleasure and pain”) are at stake.

But Locke makes the relationship between Consciousness and Concernment stronger.

In 2.27.26, he introduces the famous notion that “*Person is a Forensick Term*” and proceeds to explain what that means:

It is a Forensick Term appropriating Actions and their Merit; and so belongs only to intelligent Agents capable of a Law, and Happiness and Misery. This personality extends it *self* beyond present Existence to what is past, only by consciousness, whereby it becomes concerned and accountable, owns and imputes to it *self* past Actions, just upon the same ground, and for the same reason, that it does the present. All of which is founded in a concern for Happiness *the unavoidable concomitant of consciousness* ... And therefore whatever past Actions it cannot reconcile or appropriate to that present *self* by consciousness, it can be no more concerned in, than if they had never been done: And to receive Pleasure or Pain; *i.e.* Reward or Punishment, on the account of any such Action, is all one, as to be made happy or miserable in its first being, without any demerit at all. For supposing 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? (2.27.26)

A person, as a moral agent, is responsible for, and is concerned about, all and only those actions, past or present, of which it is conscious. Also, because punishment requires that the person be conscious of the reason for the pain inflicted upon herself; and lacking that consciousness, the pain of punishment would be just pain, gratuitous pain.

## 12.6 Locke against Substance-Based Theories of Personal Identity

One of the more radical aspects of Locke's theory of personal identity – and the one that his contemporaries had the most difficulty swallowing – is his rejection of what I will call “substance-based theories of identity over time,” namely a theory in which sameness of a particular substance – mind/soul/spirit/form or corporeal substance – is necessary and/or sufficient for the persistence of some thing. We have already seen Locke's rejection of a substance-based theory of the persistence of an organism. However, Locke's rejection of a substance-based theory of *personal identity* is noticeably more radical because it amounts to a denial that the same finite intelligence (*i.e.* mind/soul/spirit) is necessary and/or sufficient for sameness of person, a view held by, among others, Descartes (see especially *AT*, IV 166–7; *CSMK*, 243).

### 12.6.1 *Sameness of Thinking Substance is not Necessary for Personal Identity (For All We Know)*

Locke's argument against the necessity of the same soul begins with a statement of what Lisa Downing (1998) calls Locke's "epistemic modesty." Locke *may* be willing to entertain the necessity claim, but not "till we have clearer views of the Nature of thinking Substances" (2.27.13). Unfortunately, Locke's position about our ideas of substances in 2.23 is not confidence inspiring: "*We have as clear a Notion of the Substance of Spirit, as we have of Body*" (2.23.5). That may initially sound promising, but here is the rub: Locke claims that "the Substance of each being equally unknown to us" (2.23.28). Our idea of body is utterly unclear; so too is our idea of thinking substance. Our ideas of both body and thinking substance are "equally obscure," a "*je ne sais quoi*" in which features inhere. Contrary to Descartes and the Cartesians, Locke argues that we not only lack an utterly clear and distinct idea of thinking substance, "we have no clear and distinct *Idea* of the *Substance* of a Spirit" (2.23.5).

To see what Locke is driving at, consider *how* someone *might* acquire the idea of a thinking substance. It seems that "reflection" or introspection is the only plausible candidate for the origin of an idea of thinking substance. So, let us close our eyes (optional), and introspect. As Hume would later point out, introspection may reveal all sorts of things; for instance, that I'm thinking about Hume, feeling very tired and a bit hungry, seeing various colors, etc. All of this, Locke grants. But none of the perceptions I have when introspecting are perceptions of the thinking substance having those perceptions. To be sure, Locke thinks that I (the person I am) have all of my thoughts attended by consciousness, which means that when I have the perception of hunger or of seeing blue, I recognize that *I* am hungry and *I* see blue. However, I never have a perception of a particular thinking substance which is hungry and seeing blue. My ideas acquired through introspection are consistent with any number of hypotheses not only about the *nature* of thinking substance, but also consistent with any *number* of thinking substances a person may have throughout its duration. Thus, when we consider whether sameness of thinking substance is a necessary condition for personal identity, we must simply shrug our shoulders, embrace epistemic modesty, and say, as Locke does, that "that cannot be resolv'd, but by those, who know what kind of Substances they [i.e. thinking substances] are, that do think; and whether the consciousness of past Actions can be transferr'd from one thinking Substance to another" (2.27.13. cf. 2.1.10), shifting the burden to the advocate of substance-based theories to explain why the same consciousness could not be "instantiated" by successive, numerically distinct thinking substances.

All I can be sure of, as an epistemically modest person, however, is my identity at different times. So, Locke thinks that it must be the case that "that *self* is not determined by Identity or Diversity of Substance, *which it cannot be sure of*, but only by Identity of consciousness" (2.27.23, emphasis mine).

### 12.6.2 *Sameness of Thinking Substance is not Sufficient for Personal Identity (For All We Know)*

A similar approach is taken to the sufficiency claim, namely epistemic modesty about thinking substances and our concern for our *selves*. Locke has us picture a thinking substance “strip’d” of its previous consciousness. In the absence of some compelling evidence, there is no reason to suppose that a “strip’d thinking substance” would be the same person as any previous person. There is nothing in virtue of which that thinking substance would “latch up” to a particular person. Now, if it is replied that a thinking substance, even one that is “strip’d,” just *is* the person, Locke would be curious about how the author of this reply came to have such insight into the connection between strip’d thinking substance and a person.

It is also due to considerations of persons in the “forensick” sense that Locke rejects the sufficiency claim. Suppose, for instance, that a current person in fact has what was once Nestor’s soul or mind. By itself, this:

would no more make him the same Person with *Nestor*, than if some of the Particles of Matter, that were once a part of *Nestor*, were now a part of this Man, the same immaterial Substance without the same consciousness, no more making the same Person by being united to any Body, than the same Particle of Matter without consciousness united to any Body, makes the same Person. But let him once find himself conscious of any of the Actions of *Nestor*, he then finds himself the same Person with *Nestor*. (2.27.14)

Locke’s Dayman and Nightman case (2.27.23) is also an example of two persons with the same thinking substance. But having “distinct, incommunicable consciousnesses,” the two persons have no concernment for the thoughts, actions, or parts of the other. Even in the First Edition of the *Essay*:

I ask then, Whether *Castor* and *Pollux*, thus, with only one Soul between them, which thinks and perceives in one, what the other is never conscious of, nor is concerned for, are not two as distinct Persons, as *Castor* and *Hercules*; or, as *Socrates* and *Plato* were? (2.1.12)

Locke’s answer: “yes.”

Finally, Locke offers a big “selling point” for his theory and against the substance-based view: it makes the issue of the general resurrection less problematic (Locke thinks). To be sure, Locke holds the Christian doctrine that “the same persons shall be raised, and appear before the judgment-seat of Christ, to answer for what they have done in their bodies” (*Works* 3, 330). But sameness of substance is irrelevant. A resurrected person need not have the same body (conceived either as a mass of matter or as an organism) or the same thinking substance as she did premortem. As long as it is the same *person*, both pre- and postmortem, it is a resurrection. Moreover, by

making sameness of body or thinking substance irrelevant to personal identity, Locke thinks that:

we may be able without any difficulty to conceive, the same Person at the Resurrection, though in a Body not exactly in make or parts the same which he had here, the same consciousness going along with the Soul that inhabits it. (2.27.15)

Locke need not worry about various puzzles concerning the resurrection of the same body, despite the persistent efforts of Stillingfleet.

### 12.6.3 Reactions

Locke's theory was almost immediately attacked by numerous philosophers, theologians, and churchmen. It was attacked both as being dangerous to the Christian faith and as being philosophically problematic.

### 12.6.4 "Amnesia"

One of the first worries that strikes readers of Locke concerns alleged cases of persons with amnesia. Surely, it is thought, persons with amnesia do not cease to be the same persons they were before. But this thought is wrong. "Amnesia" is a loaded term. If amnesia is supposed to be a condition in which there is sameness of person without sameness of consciousness, then alleged cases of amnesia beg the question against Locke. In fact, Locke takes alleged cases of amnesia to be cases in which there is the same *man*, but different consciousnesses – and hence different persons. The same man is merely "home" to distinct persons. When addressing this issue, Locke states: "we must here take notice what the Word *I* is applied to, which in this case is the Man only" (2.27.20).

## 12.7 Getting Away with Murder?

Does Locke's theory allow persons to get away with murder? Suppose that a human being is "home" to two distinct persons, Soberman and Drunkman. Drunkman commits all kinds of heinous acts, but Soberman and Drunkman have "distinct, incommunicable consciousnesses." In Locke's view, it seems that Soberman has gotten away clean. If Soberman is not conscious of any of Drunkman's acts, then Soberman is very literally not the same person as Drunkman. To be sure, Locke thinks that Soberman and Drunkman may be the same human being. But when looking for someone to hold culpable for the crimes of last night, Locke's view entails that no person will be found because there simply is not a person today who is identical to the perpetrator of last night's crimes. All there is is a person (Soberman) "associated with" a human being, and the human being looks a lot like the Drunkman.

In what strikes me as a strange turn, Locke does think it is permissible for a court of law to punish Soberman for Drunkman's acts, despite the fact that they are not the same person; they are only the same human being/man. Perhaps this is why Molyneux suggested to Locke that that kind of punishment is permissible because drunkenness itself is a crime, and "one crime cannot be alledged in excuse for another" (*Corr.* 1693). Locke, however, does not accept Molyneux's assessment of the situation. As Locke asks rhetorically: "what has this to do with consciousness?" (*Corr.* 1693). According to Locke, this is permissible because laws (and actual systems of law) punish "according to their degree of knowledge." The judge *knows* that the defendant is the same human being/man who committed the crimes, but the judge cannot know that the person-constituting consciousness is lacking. God, however, who has perfect knowledge, will not hold Soberman responsible and punish him for Drunkman's acts precisely because, lacking a link of consciousness, they are simply not the same person.

It is here, in Letter 1693, that Locke provides perhaps the best question to be asked when making judgments about personal identity and the criteria for personal identity: *What has this to do with consciousness?*

## 12.8 Two Famous Objections: The Circularity Objection and the Transitivity Objection

Locke's theory of individuation and identity, especially personal identity, was met with a swarm of objections in the decades following his death. I only have space to bring up what are commonly considered two of the most important objections (see Thiel 2011, ch. 6).

The *Circularity Objection*, associated with Joseph Butler (1736), and the *Transitivity Objection*, associated with Thomas Reid (1785) and Bishop George Berkeley (*Alciphron* 7.8 [1732]), were made long after Locke's death; so we do not know how Locke would have responded to them. And the pitfalls of asking what one of the Mighty Dead Philosophers *would* have said are well known. Nevertheless, scholars have attempted to interpret Locke's theory of personal identity in such a way that it is *already* immune to the objections. In his recent book, Galen Strawson appeals to early evidence (e.g. parts of the First Edition) of Locke's theory of personal identity to suggest that he would not have given a theory *so obviously* subject to counterexamples (Strawson 2011). And E.J. Lowe (2005, 96), on the other hand, suspects that Locke's response "would have been to 'bite the bullet' – that is, to stick to his original account and deny the force of the objections." In any case, it is interesting and instructive to look at these two objections.

Although a type of circularity objection continues to be raised against Locke's theory, as well as contemporary Lockean theories, there is actually very little in common between the contemporary circularity objection and the circularity objection made

by early-modern readers of Locke – as in the case of Joseph Butler, whose writing is considered the *locus classicus* of the circularity objection. In fact, Butler’s “objection” to Locke is actually at least two objections, neither of which packs a punch. The first accuses Locke of confusing the metaphysics of personal identity with the *epistemic* issue of how a person can know “from the inside” that she is the same person as a past person. Butler allows that “consciousness of what is past does thus ascertain our personal identity to ourselves” (Butler 1736, 302), but consciousness cannot constitute personal identity. Of course, this mere assertion simply begs the question against Locke’s theory. Second, due to Butler’s particular (peculiar?) view about the relationship between consciousness (which he understands differently than Locke!), memory, and knowledge, he makes the accusation of the circularity in Locke as follows (1736):

And one should really think it Self-evident, that consciousness of personal Identity presupposes, and therefore cannot constitute personal Identity, any more than knowledge, in any other case, can constitute truth, which it presupposes. (264)

If *this* is Butler’s objection to Locke, Locke has little to worry about. I can put it in no better terms than Harold Noonan does:

[Butler’s] thought is simply that, in general, one cannot define *what it is for it to be the case that P* in terms of *what it is for it to be known the P*, and that, as a special case of this, one cannot define *what it is for personal identity to obtain* in terms of *what it is for it to be known – or to be an object of consciousness – that personal identity obtains*. (2003, 55–6)

Butler mistakenly thinks that on Locke’s view what constitutes personal identity is what persons are conscious of, *and* what persons are conscious of is *their own personal identity!* But, at this point, it should go without saying that this is far from Locke’s view. Persons are conscious of their own actions and thoughts; and it is by a person *x* being conscious of a thought or action belonging to *y* that *x* and *y* are the same person. Frankly, Butler does not appear to have understood Locke’s theory of personal identity very well.

There is, however, a different type of circularity objection in more recent discussions of Locke and Lockean theories of personal identity. According to the contemporary circularity objection, Locke’s theory, as grounded in consciousness (which, unfortunately, many contemporary philosophers still take to be the same as memory), requires, in its analysis, the presupposition that the things in question are identical. It is claimed that, if Locke’s persistence conditions for persons have a chance of being true, there must be some way to distinguish *real veridical memories* from *false memories*. We know all too well that mental illness, as well as certain questionable practices from “regression therapy” to downright abusive types of psychological manipulation such as “gas-lighting” and “psychic driving,” frequently results in a person *seeming* to remember, to be conscious of, and to believe various things

about herself and her past which are not true. From the person's own perspective, however, the non-veridical "memories" *seem* just as accurate and "real" as her real veridical memories. So, it is objected that what distinguishes a person's false memory of a prior event from a veridical memory of that prior event is that, in the latter case, the person remembering the event is *identical* to the person who was present at the earlier event. Thus, according to this objection, Locke's analysis of personal identity is circular because the *analysans* explicitly includes the very *analysandum* at hand. That is:

$x$  at  $t_1$  is the same person as  $y$  at  $t_2$  if and only if ..., and  $x$  at  $t_1$  is the same person as  $y$  at  $t_2$ .

Most Locke scholars now agree that the contemporary version of the circularity objection is no genuine threat to Locke. Aside from the fact that it rests on conflating consciousness and memory, the objection, at most, points out a *prima facie* strangeness in Locke's theory, but it doesn't show that Locke's theory is circular. On Locke's theory, the person herself is the authority when it comes to what she is conscious of, what she appropriates, and hence what she is concerned for. Does this mean that if I appropriate the acts of serial killer, Ed Gein, then I am to be punished for those acts? Any person identical to Gein, after all, is surely to be punished justly and swiftly by God. Locke's position on this is not entirely clear, but there is at least one good Lockean reason to think that my divine punishment is, in fact, deserved. *If* I am conscious of those acts, and *if* I appropriate them, then *I take myself* to be the person who committed those unspeakable acts. However, because I am conscious of these acts, I have every opportunity to repent and ask for God's forgiveness for those acts. Now, *if* I take *myself* to have committed those horrible acts, and I have intentionally not repented for them, then surely I have no cause for quarrel if God punishes me for those acts I have appropriated as my own. This is, perhaps, part of what Locke means by "that consciousness, which draws Reward or Punishment with it" (2.27.13).

## 12.9 The Transitivity Objection

There are certain *desiderata* we generally take to be necessary for any theory of identity. Identity, for instance, must be transitive; that is, if  $a = b$ , and  $b = c$ , then  $a = c$ . It seems that any relation that is not transitive simply could not be identity. Perhaps the most famous objection to Locke's theory of personal identity is that it not only fails to preserve transitivity, but in many cases is positively committed to a denial of transitivity. One might think that this is a natural consequence of Locke's theory of personal identity because the consciousness that constitutes personal identity is pretty clearly not transitive.

Be that as it may, it still seems like a bad consequence, perhaps even a *reductio* of Locke's theory. Consider Reid's famous "brave officer" example, a kind of example which, Reid admits, probably never occurred to Locke. (According to M.A. Stewart,

Reid apparently borrowed the brave officer example from his colleague, George Campbell (Stewart 1997.)

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. These things being supposed, it follows, from Mr. Locke's doctrine, that he who was flogged at school is the same person who took the standard, and that he who took the standard is the same person who was made a general. Whence it follows, if there be any truth in logic, that the general is the same person with him who was flogged at school. But the general's consciousness does not reach so far back as his flogging – therefore, according to Mr. Locke's doctrine, he is not the person who was flogged. Therefore, the general is, and at the same time is not the same person with him who was flogged at school. (*Inquiry* 3.6)

Take three times,  $t_1$ ,  $t_2$ , and  $t_3$ . At  $t_2$  a Brave Officer is conscious of having been flogged as a Young Boy at  $t_1$ . And at  $t_3$ , an Old General is conscious of having taken the standard as the Brave Officer at  $t_2$ . However, the Old General at  $t_3$  is not (and cannot be made to be) conscious of having been flogged at  $t_1$ . According to Reid, Locke's theory entails that the Young Boy at  $t_1$  is the same person as the Brave Officer at  $t_2$ , and the Brave Officer at  $t_2$  is the same person as the Old General at  $t_3$ . However, because there is no consciousness extended backward from  $t_3$  to the flogging at  $t_1$ , the Old General is not the same person as the Young Boy. Therefore, Locke's theory of personal identity violates the transitivity of identity.

I must admit that I find this objection decisive: it does show that personal identity is not transitive in Locke. I find myself wanting it to be otherwise. Why? Aside from the obvious, because Locke's discussion of personal identity in 2.27 appears to be part of a *general* line of thought and discussion of identity and individuation – a line that begins with atoms and masses of atoms, followed by organisms and artifacts, and which culminates in the identity of persons. It would seem, then, that the most general features of identity, among which are reflexivity, symmetry, and transitivity, would be found at each stop along the discussion. Locke gives us no reason to suspect that these are absent at the other stops. In fact, violating transitivity at *any* of the stops before persons would certainly be thought to be a *reductio* of Locke's theory. So, at the very least, persons do not fit nicely and completely in the overall examination of identity.

While I lament the violation of transitivity, Galen Strawson embraces the violation as “an illustration of [the theory's] fundamental and forensic point, the commonsense point (it's commonsense relative to the story of the Day of Judgment) that human beings [*sic*] won't on the Day of Judgment be responsible for all the things they have done in their lives, but only for those that they're still Conscious of and

so still Concerned in" (2011, 59). I agree with Strawson that the forensic nature of persons and their Judgment by God ("accused by their own consciousness") is crucial to Locke's theory of personal identity. That is, Locke's theory of personal identity gets the *forensick-responsibility-divine judgment-and-allocation of punishment/reward* issue exactly right, but it does so at the expense of a plausible theory of the numerical identity of a person over time.

## 12.10 Attempts to Save Locke from the Transitivity Problem

### 12.10.1 Appropriation Interpretations

An influential attempt to save Locke is found in a classic paper by Ken Winkler (1991). According to Winkler, we must start by recognizing two important things: first, that identity is transitive, so the General is the same person as the Boy. Second, that it is precisely this transitivity that explains why the General and the Boy are the same person. They are the same person, according to Winkler, in virtue of the fact that the General is linked by consciousness to the Officer – which Winkler takes to be the same as saying that the General *appropriates* the actions of the Officer – and the General appropriates the actions appropriated by the Officer. But because the Officer appropriates the actions of the Boy, Winkler thinks, “the general commits himself to the actions the officer appropriated” (222). A similar move is made by Don Garrett. According to Garrett, if a person is conscious of the acts and/or perceptions of an earlier person, then this “extends the history to the person to whatever *other* perceptions or actions are implicated in *sameness of consciousness* with that earlier perception or action, regardless of present ability to remember them” (2003, 109). And in her recent defense of an Appropriation Interpretation of Locke, Antonia LoLordo claims that “to extend your consciousness backward to an action is simply to appropriate it as your own or to impute it to yourself” (2012, 70). Now, if *that* is all that is required – to impute something to one's self – for being a person-preserving consciousness, then the Transitivity Problem is not really a problem.

But consider the following: my dad has a photograph of me when I was a toddler, in which I am wearing a pair of adult-sized cowboy boots and looking adorable. I have seen this photo many times, and I believe, at least when I am not thinking about Locke, that I am the same person as that toddler. But am I? There is certainly no link of consciousness from my present self to that event or thought. Perhaps I appropriate this event, as part of my self, on the basis of something other than consciousness of the event; for instance, by being told by my parents that the kid in the photo is me, the kid in the photo looks like a very young me, etc. But we should approach this first by asking of this sort of appropriation, Locke's all-important question: *What has this to do with consciousness?* If there is no link of consciousness between my present self and the kid in the photo, then Locke's theory entails that my present self is not the

same person as the kid in the photo. It is just that simple. Second, because there is no link of consciousness in this case, the most that can be said, concerning the basis for thinking it is me (what my parent said, and the resemblance), is that I am the same human being ("man") as the kid in the photo. Third, because a link of consciousness is missing, what could connect my present self to the kid in the photo and which would make me certain that I am the same person as the kid? From my first-person perspective, even my being the same human being as the kid is not certain. Now, if there *were* a link of consciousness, I would be certain of my personal identity, and all the stuff my parents told me and my resemblance to the kid would be eliminated or ignored as totally irrelevant: it is the consciousness, of which I am certain, that does the work. Unfortunately, in the present case, I am not the same person as that adorable little kid in the cowboy boots.

### 12.10.2 *Yaffe's Proper Subject of Punishment Interpretation*

According to a recent interpretation by Gideon Yaffe, Locke can be understood only if we focus on the "forensick" aspect of personhood. Persons are moral agents for whom moral appraisal for their actions is appropriate and, Yaffe thinks, the persistence conditions for persons should reflect this forensic aspect. Accordingly, he reads Locke as saying that  $x$  at  $t_1$  is the same person as  $y$  at  $t_2$  if it is proper to punish or reward  $y$  at  $t_2$  because of an act performed by  $x$  at  $t_1$ . Yaffe's view is actually even more radical: Locke should be understood so that the moral aspects of personhood determine and ground the metaphysical fact of personal identity. As Yaffe states: "Who is identical to whom depends on who is rightly rewarded or punished, rather than the reverse" (2007, 226. Cf. Mackie 1976, 83).

There is certainly something right about this view: punishment and reward are just only if the person punished/rewarded is the same person who performed the act for which the punishment is given. Contrary to Yaffe, one of the more explicit and transparent claims Locke makes is that the appropriateness of punishing a person is grounded in the identity of that person with the perpetrator, and not vice versa (see Thiel 2011, 214–15). As Locke claims: "In this *personal Identity* is founded all the Right and Justice of Reward and Punishment" (2.27.18). Additionally, in a letter to Molyneux from January 1694, Locke unequivocally reiterates this point: "I am shewing that Punishment is annexed to personality, and personality to consciousness: How then can a drunkard be punish'd for what he did [*sic*], whereof he is not conscious?" (Letter 1693).

When all is said and done, it remains unclear how Yaffe's interpretation could preserve the transitivity of identity. If anything, Yaffe's interpretation would even more strongly entail that the General is not the same person as the Boy because punishing the General for what the Boy did is completely inappropriate. It is just like a case stated by Locke:

For supposing 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? (2.27.26)

### 12.11 Conclusion: The Silver Lining

When we consider the important relationship between personal identity, accountability, reward, and punishment, the non-identity between the Old General and the Young Boy strangely appears to be a welcome result of Locke's theory. Indeed, more than once in the *Essay*, Locke appeals to a person's accountability for her actions – *accountability to God and His Law*. According to Locke, we need not worry about each person getting their just deserts in this life; eventually, God will sort things out in accordance with what a person is known by God to be conscious of. Locke's God does not play an active role in the consciousness of a person. God does not, on the Last Day, place the consciousness of being flogged in the Old General. Rather, Locke's God simply "reads off" a person's evaluable acts from her own consciousness. As Locke states:

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. (2.27.22) [T]he Apostle tells us, that at the Great Day, when every one shall *receive according to his doings, the secrets of all Hearts shall be laid open*. The Sentence shall be justified by the consciousness all Persons shall have, that they *themselves* in what Bodies soever they appear, or what Substances soever that consciousness adheres to, are the *same*, that committed those Actions, and deserved that Punishment for them. (2.27.26)

In *The Reasonableness of Christianity*, published in 1695 (very close to the time he had written 2.27), Locke delves into the issue of redemption and its necessity for personal salvation: "Repentance is as absolute a Condition of the Covenant of Grace, as Faith; and as necessary to be performed as that" (*Works*, vol. 7, 103). And "Repentance is an hearty sorrow for our past misdeeds, and a sincere Resolution and Endeavor, to the utmost of our power, to conform all our Actions to the Law of God" (*Works*, vol. 7, 105). Moreover, the "misery without demerit" of punishing a person for an action of which she cannot be conscious is not reasonable or just. So, God would not punish a person for an action that, due to no consciousness of the action, the person simply *could not* ask for forgiveness. After all, no one should be expected to repent for the actions of another person, and without consciousness of an action, that action *is* the action of another person. Locke's theory, by my lights, gets things right with respect to the relationship between persons and their (ultimate) accountability to an eminently reasonable God.

### *Acknowledgments*

Thanks to Matt Stuart for asking me to write this chapter, and his nearly infinite patience while waiting for me to finish. Thanks also to James Baillie for generously answering some of my questions about the state of the art in the philosophy of personal identity. I also wish to thank Joseph “Joey” Stenberg and the students in my Seventeenth-Century Metaphysics course at CU-Boulder and my course on Individuation and Identity in Early Modern Philosophy at the University of Michigan, Ann Arbor, for comments, criticism, and discussion of the issues in this chapter.

### References

- Berkeley, G. (1732) *Alciphron, or the Minute Philosopher*. In A. A. Luce and T. E. Jessop (Eds.), *The Works of George Berkeley, Bishop of Cloyne*. 9 vols. London: Thomas Nelson, 1948–1957, volume 3.
- Bolton, M. B. (1994) Locke on identity: the scheme of simple and compounded things. In K. F. Barber and J. J. E. Gracia (Eds.), *Individuation and Identity in Early Modern Philosophy: Descartes to Kant* (pp. 103–131). New York: SUNY Press.
- Butler, J. (1736) *The Analogy of Religion*. London: printed for James, John and Paul Knapton, at The Crown in Ludgate Street.
- Chappell, V. (1989) Locke and relative identity. *History of Philosophy Quarterly*, 6, 69–83.
- Chappell, V. (1990) Locke on the ontology of matter, living things, and persons. *Philosophical Studies*, 60, 19–32.
- Downing, L. (1998) The status of mechanism in Locke’s Essay. *Philosophical Review*, 107, 381–414.
- Garrett, Don. 2003. “Locke on Personal Identity, Consciousness, and “Fatal Errors””. *Philosophical Topics* 3: 95–125.
- Kaufman, D. (2007) Locke on individuation and the corpuscular basis of kinds. *Philosophy and Phenomenological Research*, 75 (3), 499–534.
- Kaufman, D. (2008) The resurrection of the same body and the ontological status of organisms: what Locke should have (and could have) told Stillingfleet. In Hoffman, P., Owen, D. and Yaffe, G. (Eds.), *Contemporary Perspectives on Early Modern Philosophy*. Peterborough, ON: Broadview.
- Law, E. (1823) A Defense of Mr. Locke’s Opinion concerning Personal Identity. In *The Works of John Locke*, vol. 3, pp. 177–199. London: printed for Thomas Tegg.
- Lewis, D. (1986) *On the Plurality of Worlds*. Oxford: Blackwell.
- LoLordo, A. (2012) *Locke’s Moral Man*. Oxford: Oxford University Press.
- Lowe, E. J. (2005) *Locke*. London: Routledge.
- Mackie, J. L. (1976) *Problems from Locke*. Oxford: Oxford University Press.
- McCann, E. (1987) Locke on identity: matter, life and consciousness. *Archiv fuer Geschichte der Philosophie*, 69, 68.
- Noonan, H. (1989) *Personal Identity*. London: Routledge.

- Noonan, H. (2003) *Personal Identity*, 2nd edn. London: Routledge.
- Reid, T. (1785) *Essays on the Intellectual Powers of Man*. Edinburgh: John Bell and G.G.J and J. Robinson.
- Stewart, M. (1997) Reid on Locke and personal identity : some lost sources. *Locke Newsletter*, 28, 105–116.
- Strawson, G. (2011) *Locke on Personal Identity*. Princeton, NJ: Princeton University Press.
- Thiel, U. (2011) *The Early Modern Subject: Self-Consciousness and Personal Identity from Descartes to Hume*. Oxford: Oxford University Press.
- Uzgalis, W. (1990) Relative identity and Locke's principle of individuation. *History of Philosophy Quarterly*, 7, 283–297.
- Winkler, K. (1991) Locke on personal identity. *Journal of the History of Philosophy*, 29, 149–174.
- Yaffe, G. (2007) Locke on ideas of identity and diversity. In L. Newman (Ed.), *Cambridge Companion to Locke's "Essay Concerning Human Understanding"*. Cambridge: Cambridge University Press.

### Further Reading

- Ayers, M. (1991) *Locke: Epistemology and Ontology*. London: Routledge.
- Conn, C. (2003) *Locke on Essence and Identity*. Dordrecht: Kluwer.
- Fine, K. (2000) A counter-example to Locke's thesis. *The Monist*, 83, 357–361.
- McCann, E. (2008) Identity, essentialism, and the substance of body in Locke." In P. Hoffman, D. Owen and G. Yaffe (Eds), *Contemporary Perspectives on Early Modern Philosophy*. Peterborough, ON: Broadview.
- Strawson, G. (2011) *Locke on Personal Identity*. Princeton, NJ: Princeton University Press.
- Stuart, M. (2013) *Locke's Metaphysics*. Oxford: Oxford University Press.
- Thiel, U. (2000) The Trinity and human personal identity. In M.A. Stewart (Ed.), *English Philosophy in the Age of Locke* (pp. 217–243). Oxford: Oxford University Press.
- Thiel, U. (2011) *The Early Modern Subject: Self-Consciousness and Personal Identity from Descartes to Hume*. Oxford: Oxford University Press.