

“Some issues in Locke’s Philosophy of Mind”

William Uzgalis (supplement to SEP article “John Locke”)

<https://plato.stanford.edu/entries/locke/supplement.html>

In this supplement, we consider some of the most interesting and controversial claims that Locke makes in the Philosophy of Mind. The two most important of these are Locke’s remarks in Book IV, Chapter 3 section 6 of *An Essay Concerning Human Understanding* that for all we know God could just as easily make matter fitly disposed to think as He could add thought to an immaterial substance; the second is the revolutionary theory of personal identity that Locke added in Chapter 27 of Book II of the second edition of the *Essay*. These claims and suggestions, along with a number of others, turn out to be connected in interesting ways and raise important questions about the degree to which Locke is willing to endorse substance dualism and the extent of his commitment to Boylean mechanism.

- Thinking Matter, Immateriality of the Soul and Immortality
- Superaddition
- The Upshot
- Personal Identity
- The Definition of Man
- Persons and Consciousness
- Man and Person
- Soul and Consciousness
- Caring for the Self and ‘person’ as a Forensic Term.

Thinking Matter, Immateriality of the Soul and Immortality

In giving us his estimate of the limits of human understanding, Locke made some claims which surprised his contemporaries. In IV.3.6 he suggests that given our ignorance of substances, it was possible that God could make matter fitly disposed think. He writes:

We have the *Ideas* of *Matter* and of *Thinking*, but possibly shall never be able to know, whether any material being thinks, or no; it being impossible for us, by the contemplation of our own *Ideas*, without revelation, whether Omnipotency has not given to some System of Matter fitly disposed, a power to perceive and think, or else joined to matter so disposed, a thinking immaterial substance.

In Book IV of the *Essay* Locke makes a distinction between what we can know and propositions that are only probable. In the passage quoted above he is telling us that we may never be able to know whether dualist or materialist theories of mind are true. If we can’t know which position is true, that would seem to be the end of it. But, in fact, even though we may never know

which is true, this leaves entirely open the question of which view is more plausible or probable.

In his recent book, *Locke’s Touchy Subjects* (2015) Nicholas Jolley argues that a variety of different passages in *An Essay Concerning Human Understanding* and other works, including his Correspondence with Bishop Stillingfleet suggest that Locke is trying to show that a weak form of materialism is a plausible candidate in the Philosophy of Mind (Jolley 2015: 8). What Jolley has in mind is form of materialism that is compatible with a dualism of properties (Jolley 2015: 11–12). Locke motivations are theological as much as metaphysical. Jolley claims that as a theologian Locke deals with issues in both natural, and perhaps more surprisingly, revealed religion. In Jolley’s view Locke’s weak form of materialism is an important piece in a theological reconciling project. Let us return to Locke’s remarks about thinking matter to see how this might be so.

In the passage from Book IV quoted above, Locke is telling us that both possibilities, thinking matter and immaterial thinking substance, are beyond our comprehension and requires in either case God to add the powers of perception and thought to matter or to the immaterial substance—which is then joined to the fitly disposed matter.

The thinking matter hypothesis disturbed a number of Locke’s early critics (and some later ones). In his proof for the existence of God in Book IV of the *Essay* Locke had adopted a conclusion of an argument substance dualists used to defend dualism, namely: “For it is impossible to conceive, that ever bare incognitive Matter, should produce a thinking intelligent Being, as that Nothing of itself produces Matter” (IV.10.9 10–13, N: 623). This is because, as the dualist claimed, from motion all you get is motion, from figure only other figures. (Let us call this the Principle of Homogeneity.) Since perception and thought are not motions or figures, they cannot be caused by matter. It appeared to some of his critics that this conclusion, which Locke uses to show that there must be some eternal immaterial cognitive Being (namely God), would also rule out Locke’s thinking matter hypothesis. Indeed, it would seem that his suggestion in IV.3.6 and his adoption of the principle in IV.10.9 amount to a contradiction. Locke, however, explicitly denies this. In another part of the discussion of thinking matter, he writes:

For I see no contradiction in it, that the first Eternal Thinking Being, should, if he pleased, give to certain systems of created, senseless matter, put together as he thinks fit, some degree of sense, perception and thought: Though I think as I have prove, Lib IV c.10th it is no less a contradiction to suppose matter (which is in its own nature void of sense and thought) should be that eternal first thinking Being. What Certainty of knowledge can anyone have that some perceptions, such as v.g. pleasure and pain, should not be in some bodies themselves, after a certain manner modified and moved, as well as they should be in an immaterial substance, upon the motion of the parts of the Body: Body as far as we can conceive being only able to strike and affect body; and Motion according to the utmost reach of our *Ideas*, being able to produce nothing but Motion, so that when we allow it to produce pleasure or pain, or the *Ideas* of *Colour*, or Sound, we are fain to quit our Reason, go beyond our *Ideas* and attribute it wholly to the good pleasure of our Maker. For since we must

allow that he has annexed Effects to Motion, which we can no way conceive Motion able to produce, what reason have we to conclude, that he could not order them as well as to be produced in a Subject we cannot conceive capable of them, as well as in a Subject we cannot conceive the motion of Matter can any way operate upon? (IV.3.6, N: 18–30)

Locke claims that it was no farther beyond our comprehension that motions of the body could give rise to pleasure and pain, colors and sounds, than that an immaterial soul could feel pain or see colors after the occurrence of some motions in the body. Plainly, he is using the mind body problem to suggest that there are features of substance dualism that are just as puzzling as the fact that the standard corpuscularian mechanism cannot explain thinking matter. Locke is putting the dualist and materialist positions on the same footing. He immediately draws the theological conclusion that follows from the two hypotheses being equally plausible:

All the great ends of Morality and Religion, are well enough secured without the philosophical Proofs of the Soul's Immateriality; since it is evident that he who, at first made us beings to subsist here, sensible intelligent Beings, and for several years continued us in such a state, can and will restore us to a like state of Sensibility in another World, and make us there capable to receive the Retribution he has designed to men, according to the doings in this life. And therefore tis not a mighty necessity to determine one way or t'other, as some overzealous for or against the Immateriality of the Soul, have been foreward to make the World believe.

In this passage about the great ends of morality and religion Locke is claiming that the issue of whether the soul is material or immaterial is of no great importance for morality or the ends of religion. There will be a resurrection of the dead, judgment and just rewards and punishments will be meted out, whether the thing that thinks in us is material or immaterial. By showing that on either account the ends and purposes of morality and religion will be served, Locke is reconciling the contestants on either side by showing that what is at stake will be gained either way. So the fierce contention over the immateriality of the soul can be set aside because it doesn't matter. God's Omnipotence resolves the need for this debate. This is, I take it what Jolley has in mind by Locke's "reconciling strategy".

Superaddition

What is it that God adds (or in Locke's technical language) superadds to matter fitly disposed to make it perceive and think? In a passage from Locke's Correspondence with Bishop Stillingfleet Locke says this:

The idea of matter is an extended solid substance; wherever there is such a substance, there is matter; and the essence of matter, whatever other qualities, not contained in that essence, it shall please God to superadd to it. For example, God creates an extended solid substance, without the superadding of anything else to it, and so we may consider it at rest: to some parts of it he superadds motion, but it still has the essence of matter: other parts of it he frames into plants, with all the excellencies of vegetation, life and beauty, which are to be found in a rose or a peach-tree, &c. above the essence of matter in general, but it is still but matter:

to other parts he adds sense and spontaneous motion, and those other properties that are to be found in an elephant. Hitherto it is not doubted but the power of God may go, and that the properties of a rose, a peach, or an elephant, superadded to matter, change not the properties of matter; but matter is in these things matter still. (Locke, 1823, Vol. III: 460; quoted in Jolley 2015: 77)

It would appear from this account that, in respect to material bodies, superadded properties are those God adds to matter to create certain kinds of things beyond the essence of matter, that yet leaves the essence of matter unaffected. Locke's God is a clockmaker and Locke seems to be claiming that any property and organization of matter beyond its essence is a divine superaddition. So, the addition of motion to some parts of matter previously at rest is a superaddition. So too is the organization of matter into vegetative life such as roses and peach trees. The organization which produces the sense and spontaneous motion of animals, such as the elephant is also superaddition. It is but a further step to perception and thought. So, why object to the last step, Locke says, when everything leading up to it is unproblematic?

The language of 'addition' or 'superaddition' seems to suggest a different model of superaddition than the one offered in the last paragraph. It suggests that God creates matter fitly disposed, and then *adds* something to it to make it perceive and think. This model of superaddition makes one wonder what exactly it is that God is adding to the matter fitly disposed. It seems that there are no very good answers. This should lead us to try to hang on to the model that says God creates the organization of a rose, and elephant and a man and the creation of these *are* the superadditions with no further addition necessary. But if something in addition is required we might think of putting the machine in motion. The simplest case is, perhaps, motion. Here, presumably, all God has to do is give certain parts of matter a push and they move. With peach trees and elephants, there is more of a puzzle. Locke holds that all of vegetative life can be explained mechanically. Still, the starting of the plant machine comes from within. Perhaps whatever it is that sets the peach tree machine going is that thing superadded. Perhaps the organization and the push are both superadditions. Perhaps one can say the same of the elephant and man

Given all this, how does Locke avoid the contradiction between his adoption of the Principle of Homogeneity and his assertion that thinking matter is possible? There is a clue in the passage from his correspondence with Stillingfleet quoted above. Locke's account flies in the face of ancient materialism that claimed that matter has eternally been in motion. It is this claim that the dualist Principle of Homogeneity is aimed to counter. Let matter be eternally in motion, it will still never produce anything other than motion. So, Locke uses the Principle of Homogeneity in its strong dualist form to block the creation of a material God. But once God is recognized to be necessarily immaterial, Locke has a different account of the essence of matter than the ancient materialists. As we have seen in the passage of the correspondence with Stillingfleet, motion is not an essential feature of matter, it is superadded. So, Locke can adopt a much weaker version of the Principle of Homogeneity that puts no constraints on God's Omnipotence. God can and does make things which are quite different from one another cause one

another. Motions in the body cause pleasure and pain and the ideas of secondary qualities such as color and sound. So motion can cause perception and thought.

It is certainly the case as Margaret Wilson pointed out that Locke sees limits to Boylean mechanism in Locke's discussion of thinking matter (Wilson 1999: 199). The problem is just that the banging together of pieces of matter cannot explain perception and thought. So, God must employ some other means which go beyond our ideas to make this happen—and thus it is incomprehensible to us. Locke also claims that the God and perhaps the angels know how the apparent qualities of man arise from atoms (III.6.3, N: 440). Locke's suggestion that corpuscularian mechanism cannot explain our ideas of secondary qualities, pleasure and pain seems to mean that he has given up this claim of a divine demonstrative science of nature (Wilson 1999: 197, 199–204). This is because on a corpuscularian mechanist basis, the connections between motions and colors, for example, are arbitrary. But, given that God is using some other way to connect motions and colors, it does not follow that the connections are arbitrary, though they seem so to us. So, a divine understanding of nature might still be possible.

The Upshot

Locke's suggestions about thinking matter were often taken as stronger than intended. Many of Locke's critics were suspicious that Locke had materialist tendencies. Instead of the reconciling conclusions about immaterial versus material substance that Locke is arguing for, his remarks were sometimes treated as proposing that matter can and does think. It hardly matters however. Samuel Clarke, for example, a member of Newton's inner circle and an orthodox Anglican theologian, engaged in a debate by correspondence or rather public pamphlet with Anthony Collins over this issue between 1707 and 1708 (Uzgalis 2007). Clarke sought to show that from our ideas alone it would be possible to show that matter thinking would involve a contradiction. If Clarke is right, Locke (even on the weaker interpretation explored here) would be wrong. There was an explosion of refutations of the claim that for all we know matter can think and the discussion of this issue lasted at least three quarters of the way through the eighteenth century. For accounts of this debate see Yolton (1983) and Martin and Baresi (2000)

Personal Identity

Locke added his Chapter "Of Identity and Diversity" (II.27) which gives his account of identity and personal identity to the second edition of the *Essay*. His account of personal identity is embedded in a general account of identity. Edwin McCann has argued that Locke's general account of identity is intended to supply a corpuscularian alternative to the Aristotelian account of identity with its 'substantial forms.' Hence we find Locke talking of the identity conditions for single atoms or masses of atoms, which are plainly not observable by the senses. (McCann 1987: 54–7) But the general account is linked in important ways with the main topic of the chapter, which is personal identity. Udo Thiel in *The Early Modern Subject* (2011) provides a massive study of the possible sources and influences on Locke's account of personal identity, as well as a thoughtful, thorough and sometimes quite origi-

nal treatment of the host of issues raised by this remarkable chapter of the *Essay*.

In this general account of identity Locke begins by noting:

When we see anything in any place in one instant of time, we are sure, (be it what it will) that it is that very thing and not another, which at that time exists in another place, how like an undistinguishable it may be in other respects. And in this consists *identity* when the ideas it is attributed to vary not at all from what they were that moment, wherein we consider their former existence, to which we compare the present. (II. 27, 1, N. 328)

Locke then goes on to state that two things of the same kind cannot be in the same place at the same time and to draw a corollary from this principle:

For we never finding nor conceiving it possible, that two things of the same kind should exist at the same place at the same time, we rightly conclude that anything that exists anywhere at any time, excludes all of the same kind, and is there itself alone. When therefore we demand whether anything be the same or no, it refers always to something that existed such a time in such a place, which 'twas certain, at that time was the same as it self, and no other. From whence it follows that one thing cannot have two beginnings of Existence, nor two things one beginning, it being impossible for two things of the same kind to be or exist in the same instant, in the very same place; or one and the same in different places. That therefore that had one beginning is the same thing, and that which had a different beginning in time and place is divers. (II, 27, 1 N. 328)

That two things can't be in the same place at the same time seems like common sense. But what does Locke mean by 'kind' and why does he add this condition? Locke's metaphysical machinery includes substances, modes and relations. Substances are independent existences, while modes and relations depend on substances for their existence. Locke claims that we have the ideas of but three sorts of substances: God, finite intelligences and bodies. He goes on to distinguish and explain the difference between the identity of a single atom, masses of atoms and living things. Each individual atom is the same at a time, and stays the same over time. So, there is no problem about the identity of atoms. Masses of atoms are individuated by their constituent atoms without regard to the way in which they are organized. If such a mass gains or loses a single atom, it has a new set of constituents and so is a new mass, even if the organization of that mass stays the same. Living things, by contrast, are individuated by their functional organization. This organization is instantiated at any time by a collection of atoms. But the organization can persist through changes in the particles which make it up—at least gradual change which continues the functions which the organization performs. Clearly the most important of these functions is the continuation of the same life. It is the continuation of the same functional organization and thus the same life which is the criterion of identity for sameness of living thing, be it an oak or a horse.

Locke's principle allows God, a soul and a body all to be in one place at the same time for each of them belongs to one of the three different kinds of

substances that Locke has explicitly identified. But it turns out that there are problems with Locke's apparently obvious principle that two things of the same kind cannot be in the same place at the same time. Some commentators would like to take 'kind' to mean the nominal essences of Book III of the *Essay* to which particular things belong. But were that the case a horse and an elephant, which have different nominal essences, could be in the same place at the same time! If to avoid this problem, we ascend the ladder of abstraction to 'animal' similar results can be produced. Instead of a horse and an elephant, one could have a horse and an oak tree in the same place at the same time. This suggests that Locke was perceptive in making bodies one kind.

Locke himself, however, seems to create a problem when he distinguishes a mass of atoms (which he calls a body) and the living thing which depends on that body for existence at a time. If living things and masses (which have a quite different set of individuating criteria from living things) are both bodies, then we have two things of the same kind (bodies) in the same place at the same time! For a discussion of the various ways that have been proposed to solve this problem see Stuart (2013) and Gordon-Roth (2015).

An issue closely related to the problem of kinds is whether we should take Locke to be an absolute identity theorist or a relative identity theorist. Absolute identity theorists hold that individuating criteria for different kinds are the unity conditions for that kind; they tell us under what conditions you have an individual of that kind. Individuating criteria thus vary from one kind to another. Identity, however, determines that we have the same individual over time and that relation is the same for all different kinds. Relative identity theorists lump unity and sameness together. So, they hold that there are different identity relations for each sort of thing. It follows from the most extreme form of relative identity theory, that it is possible for x to be the same F as y (where x and y stand for individuals and F , G , etc. for kinds) but not the same G as y , even though x and y are both F s and G s. Absolute identity theorists deny that these kind of true relative identity statements are possible. They hold that relative identity is, in fact, an incoherent doctrine.

Peter Geach (1967: 11) first suggested that Locke's account of identity should be understood in terms of relative identity. In Locke's chapter the debate ranges over a series of pairs of kinds of individuals that the absolute identity theorist sees as distinct, while the relative identity theorist sees as referring to the same thing. Are the oak and the mass of matter which compose it at an instant two distinct things or are they one thing to which two different sortals with different identity criteria apply? Can one say that this individual horse (x) is the same horse as that (y) but not the same mass of matter, even though both x and y are horses and masses of matter? The same issue applies to man and mass of matter, and soul and person. The most recent defender of the relative identity position is Stuart (2013).

The Definition of Man

Locke holds that man is an animal and is thus individuated just like other living things. So 'man' refers to a living body of a particular shape. Locke is perfectly aware that the definition of man is not really settled, and that there are a variety of competing definitions. He argues for his own definition,

which involves distinguishing between 'man' and 'person' by using a variety of thought experiments and deducing unacceptable consequences from competing definitions. He points out, for example, that while those who individuate man solely in terms of the possession of a soul can explain the sameness of man from infancy to old age, if they accept some doctrine of reincarnation, their definition requires that the same soul in different bodies be the same man as much as infant and old man. If the doctrine of reincarnation allows the soul of a man to be reborn in the body of an animal, such as a hog, if we knew that the soul of a man was in one of our hogs, it would require us to call the hog a man. Locke pairs the examples of a rational talking parrot with a creature that has the shape of a man but cannot engage in rational discourse as a thought experiment which demonstrates that rational discourse is neither a necessary or sufficient condition for being a man. So, Locke rejects the Aristotelian definition of man. On Locke's account, then, man is defined as a living body, an animal of a certain shape. Animals, Locke tells us, are like machines, in that they are an organization or construction of parts to a certain end. Locke makes this simile explicit. Having described a watch he goes on to note the similarities and differences between machines and animals:

If we should suppose this machine one continued Body, all whose organized Parts were repair'd, increas'd, or diminish'd by a constant addition or separation of insensible parts, with one Common Life, we should have something very much like the body of an Animal, with this difference, that in an animal the fitness of the organization and the motion wherein Life consists, begin together, the motion coming from within; but in machines the force coming sensibly from without, is often away when the Organ is in order, and well fitted to receive it. (II.27.5, N: 331)

Earlier in the *Essay* Locke, in distinguishing between plants, animals and humans had noted that while plants can be treated mechanistically, mechanism begins to fail in animals. Bird song is learned and so Cartesian mechanistic accounts fail, because mechanism cannot account for learned behavior. As noted earlier, it also fails for the perception of secondary qualities and thought. Whether Locke has retreated from that position here in II xxvii, and returned to a more Cartesian view is an interesting question. But assuming that this problem can be solved, what is a person?

Persons and Consciousness

In section 9 of II.27 Locke tells us that 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 is done only by that consciousness, which is inseparable from thinking, and as it seems to me essential to it; It being impossible for anyone to perceive, without perceiving that he does perceive.

'Consciousness' was quite a new term in English philosophy at the end of the seventeenth century. Ralph Cudworth had coined the term in his *The True Intellectual System of the Universe* as a translation of a Greek word used by Plotinus. Cudworth used the term in attacking Hobbes' materialism and Cudworth took consciousness to be essentially immaterial. The term had

also been used in a public controversy over the Trinity between two English divines, Robert South and William Sherlock. Thiel holds that Locke had both Cudworth's use of the term and the Trinity controversy on his mind when he wrote his account of personal identity.

Locke's account of consciousness, unlike Cudworth's is neutral in respect to the materiality or immateriality of the soul. Locke writes:

Self is that conscious thinking thing, (whatever Substance made up of whether Spiritual or Material, Simple or Compounded, it matters not) which is sensible or conscious of Pleasure and Pain, capable of Happiness or Misery, and so is concerned for its *self* as far as that consciousness extends. (II.27.17, N: 341)

Here we get an even more detailed list of possibilities as in the thinking matter passage, we have not only material or immaterial, but simple and compounded. Locke is saying that the substance that thinks in us could be any one of the eight combinations of possibilities, that is either simple and immaterial, simple and material and so on, and still produce the full range of conscious phenomena that Locke enumerates.

There have been a number of attempts to explain what Locke means by 'consciousness.' Some have treated it as a species of reflection, others as a Higher Order of Perception (HOP) type of theory—that is when one perceives that one perceives one is conscious. Udo Thiel in *The Early Modern Subject* treats reflection as an HOP theory, but holds that consciousness is neither. Thiel writes:

Locke says that "consciousness is *inseparable* from thinking, and as it seems to me essential to it...When we see, hear, smell, taste, feel, meditate or will anything, we know that we do so. Thus it is always as to our present Sensation and Perception." For Locke, being conscious denotes an immediate awareness that is an integral part of all acts of thinking as such...According to Locke, the mind relates to itself in the sense that it observes its own operations and produces ideas of them. In other words, consciousness is a presence of the mind to itself, that unlike reflection, consciousness is not a HOP. Without consciousness, reflection would not have any objects on which to reflect. (Thiel 2011: 114–115)

Man and Person

Why does Locke make this distinction between 'man' and 'person?' There are several answers. One is that having denied that man is a rational animal, one wonders what role rationality is to play, if any. What Locke does is to transfer it to the definition of 'person.' Another answer is that the distinction solves a problem associated with the resurrection of the dead. What is this problem? The problem begins with Biblical texts asserting that we will have the same body at the Resurrection as we did in this life. The issue is in what sense this is true (and whether the Bible actually says this). Clearly there are problems with the supposition that one will. Robert Boyle, in his essay, "Some Physico-Theological Considerations About the Possibility of the Resurrection" had raised some of these puzzles. Boyle writes:

When a man is once really dead, divers of the parts of his body will, according to the course of nature, resolve themselves into multitudes of steams that wander to and fro in the air; and the remaining parts, that are either liquid or soft, undergo so great a corruption and change, that it is not possible so many scattered parts should be again brought together, and reunited after the same manner, wherein they existed in a human body whilst it was yet alive. And much more impossible it is to effect this reunion, if the body have been, as it often happens, devoured by wild beasts or fishes; since in this case, though the scattered parts of the cadaver might be recovered as particles of matter, yet already having passed into the substance of other animals, they are quite transmuted, as being informed by the new form of the beast or fish that devoured them and of which they now make a substantial part. (Boyle 1675 [1979: 198])

These difficulties with putting bodies back together are obviously considerable, though not perhaps beyond the powers of Omnipotence. The culminating problem, however, is what happens to the man whose body is eaten by cannibals? Boyle continues:

And yet far more impossible will this reintegration be, if we put the case that the dead man was devoured by cannibals; for then, the same flesh belonging successively to two different persons, it is impossible that both should have it restored to them at once, or that any footsteps should remain of the relation it had to the first possessor.

These problems I suspect represent the kinds of difficulties which faced the scientists of the Royal Society, and with which Boyle was particularly concerned, in integrating the kinds of explanations of natural phenomena in terms of particles and matter in motion, with the truths of religion.

Locke explicitly tells us that the prince and the cobbler thought experiment in section 15 shows us the resolution of the problem of the same body at the resurrection. The case is one in which the soul of the prince with all of its princely thoughts is transferred from the body of the prince to the body of the cobbler, the cobbler's soul having departed. The result of this exchange, is that the prince still consider himself the prince, even though he finds himself in an altogether new body. Locke's distinction between man and person makes it possible for the same person to show up in a different body at the resurrection and yet still be the same person. Thus his answer is the same as the one Boyle gives in the work quoted above. In his correspondence with Bishop Stillingfleet he all but says that the idea of being resurrected in the same body is incoherent. Locke focuses on the prince with all his princely thoughts because, on his view, it is consciousness which is crucial to the reward and punishment which is to be meted out at the Last Judgment.

In this chapter on identity, Locke is also making a claim that consciousness need not be connected to a single soul. Let us turn then, to the distinction between soul and consciousness.

Soul and Consciousness

Though the distinction between man and person is controversial, Locke's severing the connection over time between the soul or the thing which

thinks in us and consciousness is even more radical. Locke holds that consciousness can be transferred from one soul to another, and that personal identity goes with consciousness. In section 12 of the Chapter of Identity and Diversity (27) he raises two questions: "...if the same Substance which thinks be changed, it can be the same person, or remaining the same, it can be a different person". Locke's answer to both of these questions is affirmative. Consciousness can be transferred from one substance to another and thus while the soul is changed, consciousness remains the same and thus personal identity is preserved through the change. And on the other hand, consciousness can be lost as in utter forgetfulness while the soul or thinking substance remains the same. Under these conditions there is the same soul but a different person. These affirmations amount to the claim that the same soul or thinking substance is neither necessary nor sufficient for personal identity over time. The arguments are developed by analogy with the functional organization of animals which is preserved through the gradual changes in the atoms which instantiate that organization at any given time. So, at any given time there must be a soul or thinking substance, but over time there is no necessity that one have the same soul to preserve personal identity.

The rejection of having the same soul as a sufficient condition for personal identity does a fair amount of work for Locke. It underlies the Castor and Pollux thought experiment in II.27.9–20 and many of the thought experiments in II.27. including the Day Man Night Man, and most notably the Prince and the Cobbler. The Castor Pollux thought experiment makes it plain that not only does Locke reject Descartes' claim that thinking is his essence, he also rejects the claim that as a person he is essentially a substance. Why Locke rejects the claim that having the same soul is a necessary condition for personal identity is not as clear. This is one of the most significant changes from Locke's discussion of personal identity in the first edition passage (II.1.9–20) to the chapter "Of Identity and Diversity" added to the second edition.

One answer is epistemological. Locke is skeptical about our ability to reidentify the same soul over time. He claims that if we were always awake, we could be certain that we had the same soul. But consciousness has natural gaps in it, such as periods during which we are asleep. Locke claims that there is no way of knowing that one soul has not been substituted for another during this period of absence of consciousness. Thus if having the same soul is necessary for personal identity we could never be sure that we were the same person as the day before. Perhaps another reason is that the preservation of personal identity through changes of substance is an important similarity in the analogy between living things and persons that Locke announces in II.27.11–12 and is part of an anti-Cartesian argument. Still, the rejection of the soul as a necessary condition for personal identity causes Locke at least one serious difficulty, which he discusses in II.27.13. If consciousness can be transferred from one soul to another it is information that is transferred and not the act of consciousness. But information, as in dreams, can be false. So, it seems possible that a person might come to believe that they had done something that, in fact, some other person had done or that no one had done. (For a discussion of this problem see Garrett 2003 and Strawson 2011) The rejection of the same soul as either a neces-

sary or a sufficient condition for personal identity leads to the question of whether, on Locke view, the soul is a substance or a mode.

In recent years there has developed a debate about whether persons are substances or modes. Ruth Mattern in her article "Moral Science and the Concept of a Person in Locke" (1980), Antonia LaLorde in "Person, Substance, Mode and the Moral Man in Locke's Philosophy" (2010) and Galen Strawson in *Locke on Personal Identity* (2011) have developed the evidence for mode interpretations. Strawson is the closest to Edmund Law's eighteenth century mode interpretation, making persons the forensic aspect of human beings. Uzgalis (2000) has a different version that makes both living things and person modes. For the view that Locke thinks persons are substances see Alston and Bennett 1988.

Caring for the Self and 'person' as a Forensic Term.

A number of features of Locke's definition of person as a thinking intelligent being that has reason and reflection and can know itself as itself, the same thinking thing in different times and places is designed to account for the fact that we are creatures who are capable of operating the machinery of the law. It is for this reason that 'person' is a forensic term. When contemplating an action we can think that in the future we will be the same being who will be punished or rewarded for the course of action which we choose. When being punished we can look back and see that we are the same being who committed the act for which we are being punished. Locke holds that consciousness is essential for justice to be done. If one is punished for doing something which one does not remember doing, it is equivalent to being created miserable. So, since consciousness plays the most important role in our being punished or rewarded at the last judgment for our actions, and consciousness can be transferred from one soul to another, and we have no mechanism to reidentify souls over time, it becomes clear why consciousness is Locke's choice for the bearer of personal identity, and why he makes the distinction between the substance which thinks in us and consciousness. I think this account explains a variety of oddities and difficulties in Locke's account. On his account, for example, memory must be completely accurate—at least in the respects relevant for divine judicial purposes. Evidence which others might produce about one's identity has no role to play in divine justice and so forth. Locke's account of freedom of action is also connected with his view of the forensic nature of personal identity. Freedom to review the decisions one has made about how to act are clearly of great importance in being able to operate the law. If one could not pause to consider, and change one's mind about what one was going to do, it might well be said that one could not do otherwise.

Rationality and caring for the self are also important components of the forensic character of personal identity. If one did not care about oneself, did not care whether one was happy or miserable, then there would be little reason to pursue the one and avoid the other. Reason presumably should play an important role in figuring out what decisions to make and what road to take to pursue happiness rather than misery. Obeying the laws both human and divine might well be the road to happiness, while violating them might lead in the direction of misery.