

# The Metaphysical Fact of Consciousness in Locke's Theory of Personal Identity

Shelley Weinberg

University of Illinois at Urbana-Champaign

*Journal of the History of Philosophy (forthcoming)*

## 1. Introduction<sup>1</sup>

Locke's theory of personal identity was philosophically groundbreaking for its attempt to establish a non-substantial identity condition.<sup>2</sup> Locke states, "For the same consciousness being preserv'd, whether in the same or different Substances, the personal Identity is preserv'd" (II.xxvii.13). There are two good reasons why Locke would want a theory of this kind. First, relying on a psychological rather than a substantial condition allows Locke to remain agnostic on the nature of thinking substance. Indeed, having a non-substantial identity condition is consistent with Locke's claims to know very little at all about the nature of substance. Second, at least by Locke's lights, the theory is compatible with his theological concerns. A psychological condition allows for some flexibility with regard to the theoretical particularities involved in Locke's theological commitments to the immortality of the soul and the resurrection of the body.<sup>3</sup> Moreover, both of these reasons are consistent with Locke's overall task in the *Essay* of outlining the limits of human understanding.

---

<sup>1</sup> I am enormously grateful to Donald Ainslie. I would also like to thank Paul Franks, Don Garrett, Jennifer Nagel, Marleen Rozemond, Helga Varden, Gideon Yaffe, audiences at the 2008 Eastern Division Meeting of the American Philosophical Association and the 2009 Southwest Seminar in Early Modern Philosophy, and the anonymous referees for *JHP*.

<sup>2</sup> See Udo Thiel, "Personal Identity", in *The Cambridge History of Seventeenth Century Philosophy*, ed. D. Garber and M. Ayers, (Cambridge: Cambridge University Press, 1998), pp. 868-912, for a comprehensive discussion of his precursors and the revolutionary nature of Locke's theory.

<sup>3</sup> Thiel, in "Personal Identity", especially pp. 870-71, and Michael Ayers, *Locke*, (London and New York: Routledge, 1991), II: pp. 205, discuss these issues. See also Edwin McCann, "Locke on Identity: Matter, Life, and

But what exactly did Locke mean by having the same consciousness? Let's take a look at some of the interpretive history. Many have interpreted Locke to think that consciousness identifies a self both synchronically and diachronically by attributing thoughts and actions to a self.<sup>4</sup> So, as far as we are conscious of our past thoughts and actions we are the same person.<sup>5</sup>

Locke even seems to confirm this interpretation when he refers to consciousness as that

“whereby I am my *self* to my *self*,” and then continues by saying,

If there be any part of its Existence, which I cannot upon recollection join with that present consciousness, whereby I am now my *self*, it is in that part of its Existence no more my *self*, than any other immaterial Being. For whatever any Substance has thought or done, which I cannot recollect, and by my consciousness make my own Thought and Action, it will no more belong to me, whether a part of me thought or did it, than if it had been thought or done by any other immaterial being any where existing. (II.xxvii.24)

If I am not presently conscious of having done something or I am incapable of ever being able to be conscious of having done it, I am not the same person.

Unfortunately, seeing personal identity as the result of being conscious of past thoughts and actions has also led many to interpret Locke to mean by the same consciousness simply having

---

Consciousness”, *Archive für Geschichte der Philosophie*, 69 (1987), p. 68, who cites as another reason Locke's desire to provide a theory that is compatible with Christian doctrine but does not also appeal to substantial forms. Says McCann, “If Locke can account for the identity of persons without calling upon any immaterial principles of unity—souls, substantial forms—then he will have snatched the best case away from the Scholastics”.

<sup>4</sup> I will follow the bulk of Locke scholarship by treating ‘self’ and ‘person’ as synonymous. Note, however, that Locke does draw the following distinction: “*Person*, as I take it, is the name for this *self*. Wherever a Man finds, what he calls *himself*, there I think another may say is the same *Person*” (II.xxvii.26). I interpret the difference to be that ‘self’ denotes that which we perceive from the subjective (first personal) point of view, where ‘person’ denotes the same thing as referred to from the objective (third personal) point of view. For alternative views, see Thiel, “Personal Identity”, p. 891 who interprets ‘self’ to refer either to the person or to the human being, and John Yolton, *The Two Intellectual World's of John Locke*, (Ithaca & London: Cornell University Press), 2004, pp. 23-4, for the view that ‘person’ designates a ‘moral self’, while ‘self’ designates a ‘secular self’.

<sup>5</sup> Those attributing to Locke the view that the central issue concerns the subjective constitution of the self by means of consciousness of past thoughts and actions include Mackie, *Problems From Locke*, p. 183; David Behan, “Locke on Persons and Personal Identity,” *Canadian Journal of Philosophy*, vol. ix, no. 1, March 1979, pp. 53-75; Winkler, “Locke on Personal Identity”, pp. 153f; Edwin Curley, “Leibniz and Locke on Personal Identity” in ed. M. Hooker, *Leibniz: Critical and Interpretive Essays*, (Minneapolis: University of Minnesota Press, 1982), p. 310.

the same memories.<sup>6</sup> But to interpret Locke to have a memory theory of personal identity incurs all kinds of problems as ubiquitously pointed out in Locke scholarship. As early as Joseph Butler, Locke's theory was accused of circularity, since knowing that one is identical to a past self, namely remembering past thoughts and actions, presupposes that the criterion has already been met; there is already an identical self to know.<sup>7</sup> And Reid and Berkeley are famous for having accused Locke's theory of personal identity of failing transitivity.<sup>8</sup>

Although attributing an ancestral memory theory to Locke preserves transitivity, it does not solve the problem of circularity. One way to address this problem is to attribute to Locke an "appropriation theory" of personal identity. Passages thought key to supporting this interpretation include Locke's use of the terms 'appropriation', 'self-attribution', or the 'imputing' of actions to oneself. Kenneth Winkler, whose appropriation theory interpretation of Locke has been highly influential, considers the following to be the most "powerful evidence" supporting his proposal:<sup>9</sup>

*Person*, as I take it, is the name for this *self*. Where-ever a man finds, what he calls *himself*, there I think another may say is the same *Person*... 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. ... And therefore whatever past Actions it cannot reconcile or appropriate to that present *self* by consciousness, it can no more be concerned in, than if they had never been done. (II.xxvii.26)

---

<sup>6</sup> See Anthony Flew, "Locke on Personal Identity" in ed. C. B. Martin and D. M. Armstrong, (New York: Anchor Books, 1968), pp. 53-68, reprinted in John Perry ed., *Personal Identity*, (Berkeley, Los Angeles, and London: University of California Press, 1975).

<sup>7</sup> Joseph Butler, *The Works of Joseph Butler*, ed. W. E. Gladstone, (Oxford: Oxford University Press, 1896), I: p. 385.

<sup>8</sup> George Berkeley, *Alciphron* in *The Works of George Berkeley*, ed. T. E. Jessop and A. A. Luce, 9 Vols., (London: Thomas Nelson and Sons, 1948-57), 3: p. 299, was the first to lodge the objection, but Thomas Reid's formulation in *Essays on the Intellectual Powers of Man*, ed. Derek Brookes, (University Park: Pennsylvania State University Press, 2002), p. 276, version is the one most often mentioned.

<sup>9</sup> Winkler, "Locke on Personal Identity", p. 154.

The focus is on the way in which the appropriation or self-imputation of thoughts and actions constitutes the extension of “personality” through time. Winkler interprets Locke as saying that the self has a “certain authority over its own constitution”, so that whatever I take myself to include is what my consciousness “reveals to me” as mine in any particular moment.<sup>10</sup>

But the mere fact that we all experience gaps in memory, namely gaps in the consciousness of our past actions, during which time we would not be the same person, lends a distinct impracticability to the theory, especially since personal identity is essential to Locke’s theory of moral responsibility and divine rectification. Divine rectification is the theological doctrine that we will be judged and then rewarded or punished by God in the next life for what we have done in this life. Such reward and punishment is just only if on the day of judgment the one standing before God is the same person who committed the past act. But if the sole criterion for diachronic identity is a person’s own consciousness of past and present mental states, namely what is revealed to oneself in a conscious act of appropriation, then personal identity seems ill equipped to determine just punishment. After all, we are not presently aware of all we have done.<sup>11</sup> So, as long as the identity of the self is determined from a subjective point of view rectification is impossible, for there is nothing objective for God to appeal to in determining whether or not we should be punished.<sup>12</sup> Leibniz, though sympathetic to the claim that

---

<sup>10</sup> Ibid.

<sup>11</sup> The various problems reduce to these two: first, it seems that we are justly punished for what we mistakenly attribute to ourselves, and second, we are not punished for what we cannot attribute to ourselves. So, we’re either unjustly punished or we unjustly escape punishment. Behan, “Locke on Persons and Personal Identity”, pp.74-5, Flew, “Locke and the Problem of Personal Identity”, p. 164, and Winkler, “Locke on Personal Identity”, pp. 168-72, specifically point to the insufficiency of Locke’s theory in this regard.

<sup>12</sup> Winkler, “Locke on Personal Identity” pp. 170-71, suggests that Locke might think that personal identity consists in something that underlies consciousness as a state of awareness, which “drives a wedge between the real self and the self I take myself to be... There is a *suggestion* of this kind of view in § 13, where Locke says that we do not

consciousness is a necessary condition for personal identity argued, in response to Locke in the *New Essays*, that it was not enough. Leibniz insisted on the necessity of a substantial “real identity” underlying the phenomenal states we subjectively experience.<sup>13</sup>

There is currently no widely accepted single interpretation of Locke’s theory of personal identity, and those considered most influential are seriously flawed in one way or another. The general problem seems to be that Locke needs an objective criterion for the continued existence of the person in order to avoid circularity and to be sufficient for the theory of divine rectification. Those interpretations that ultimately have substance do this work violate Locke’s explicit intent to provide a theory that need not rely on it. Hence, it seems worth exploring a very different and new interpretation, rather than concluding that Locke’s theory is incoherent. I will argue that Locke’s theory includes a metaphysical fact of a continuing consciousness, which renders it non-circular and sufficient for his theory of divine rectification. Moreover, there is no appeal to a traditional notion of substance to explain the continuity of consciousness.<sup>14</sup> At the same time, the interpretation takes account of the importance of memory and the first personal nature of our conscious states. To be clear this is not also an argument that once Locke is

---

know ‘whether the consciousness of past Actions can be transferr’d from one thinking Substance to another’ (E II.xxvii.13: 337), a claim that would be extremely odd if Locke thought that consciousness was something whose nature was transparent, and without a foundation that might be better known by God than by ourselves. The rest of § 13 strongly suggests that consciousness is *not* transparent—that it has (though this may be misleading) an unknown real essence”. Rather than abandon the appropriation theory, the insufficiency with divine rectification leads Winkler, “Locke on Personal Identity”, pp. 164-65, to go beyond Locke yet remain in his words “Lockean” by appealing to substance (“time slices”) to make sure that God has something to look to in determining punishment: “Suppose we allow that substance-stages—temporal slices of the enduring things Locke has in mind when he speaks of substances—can be the subjects of thoughts, feelings, and actions of earlier ones, giving rise to relations of psychological continuity.” But surely Locke himself would not make such an appeal to substance to ground the identity of persons.

<sup>13</sup> Leibniz, *New Essays*, p. 236. See Curley, “Locke and Leibniz on Personal Identity”, pp. 302-26, for an interesting comparison of Locke to Leibniz on this issue.

<sup>14</sup> See Mackie, *Problems From Locke*, p. 200, Winkler, “Locke on Personal Identity”, p. 170, McCann, “Locke on Identity” p. 75-6, and Margaret Atherton, “Locke’s Theory of Personal Identity”, *Midwest Studies in Philosophy*, vol. 8 (1983) p. 287-89 for different ways of suggesting that consciousness has an underlying substantial real essence.

rescued from incoherence we should find Locke's theory of personal identity a convincing one, even though Locke's contribution to contemporary accounts of personal identity is substantial.

The clue to the interpretation is that Locke's conception of consciousness involves an ambiguity. In the chapter devoted to personal identity, Locke seems to see consciousness as 1.) *a mental state inseparable from an act of perception by means of which we are aware of ourselves as perceiving*, and as 2.) *the ongoing self we are aware of in these conscious states*.

The first sense of consciousness, which is the one that provokes the objections already mentioned, is a momentary psychological state. This sense of consciousness allows for a momentary subjective experience that the self presently perceiving is the same as the self that remembers having once had a past thought or action. 'Subjective', here, is meant to describe something psychological, which is epistemically available only from a first personal point of view. The second sense of consciousness, which is needed to answer the objections, is the objective fact of an ongoing consciousness. 'Objective' is meant to describe something that is epistemically available from a third personal (maybe only God's) point of view. The ongoing self that I am aware of in being conscious of past and present thoughts and actions seems also to have an objective continued temporal existence through any gaps in my successive states of awareness of myself. Disambiguating these two senses of consciousness allows for a metaphysical fact, what I am also calling an 'objective fact', of my diachronic existence. Thus, there is no problem of circularity, and there is a metaphysical ground for Locke's theory of divine rectification. God need only look to all that I, as a continually existing consciousness, have done to determine my just punishment and reward.

I will present a textual argument for why Locke thinks there is a metaphysical fact of the continued existence of consciousness that does not appeal to substance. I will also say a little

something about what that metaphysical fact could consist in. I will then argue that the metaphysical fact is revealed to us as a phenomenological fact. Given the right perceptual situation we will *experience* ourselves as temporally extended. Finally, I will provide an analysis of memory that helps to support the metaphysical fact of consciousness as well as show why the denial that substance is the criterion for personal identity does not also serve to deny that there can be a metaphysical fact of a continuing consciousness.

## 2. The Metaphysical Fact of a Continuing Consciousness

The problems of circularity or of providing enough by way of consciousness to satisfy the demands of divine rectification seem insurmountable as long as consciousness is interpreted to be merely a momentary psychological state of awareness. There are passages in II.xxvii, however, suggesting that consciousness (the I, self, or person) is something that persists through our momentary conscious states of ourselves.<sup>15</sup> For example, Locke sometimes speaks of consciousness not only as a distinct thing, but also as something that continues through time. He talks about there being “two distinct incommunicable consciousnesses” alternating by day and night (II.xxvii.23), “the same consciousness being transferr’d from one thinking Substance to another” (II.xxvii.13), “the same consciousness being continued in a succession of several

---

<sup>15</sup> Aaron, *Locke*, pp. 150, 152-53, argues that although the “analysis is not satisfactory” Locke’s conception of consciousness is “not merely that we are aware of the passing perceptions, but of an abiding identical I...I am not merely conscious of a series of perceptions, for I am conscious of a permanent self, an I who experiences these perceptions and who is now identical with the I who experienced perceptions yesterday”. Martha Brandt Bolton, “Locke on Identity: The Scheme of Simple and Compound Things”, in *Individuation and Identity in Early Modern Philosophy*, eds. K. F. Barber and J. J. E. Gracia, (Albany: State University of New York Press, 1994), p. 116, states “the notion of ‘consciousness extended backward is a well-known locus of difficulty” for the interpretation that Locke has a memory theory. Mackie, *Problems From Locke*, p. 178, attributes to Locke the use of ‘consciousness’ as both a “a concrete noun” and an “abstract verbal noun” in the sense that there can be “distinct incommunicable consciousnesses” (II.xxvii.23) as well as states of “being conscious”. In addition, in those conscious states we have some awareness of an I or of “perceiving ideas from the inside”. Nevertheless, he cannot find a way to explain how there can be an objective fact of an ongoing consciousness. This is why, I think, Mackie comes to the conclusion that instead of a theory of personal identity, Locke has only a theory of action appropriation.

Substances” (II.xxvii.10), and in several places, the “same consciousness extended” (II.xxvii.9, 10, 16), “reaching” (II.xxvii.9, 17), and “continued” (II.xxvii.25) into the past or future.

Although it is true that Locke more frequently speaks of consciousness as a momentary state of awareness of myself as thinking and acting, we can’t simply ignore these other passages. The difficulty is to see how consciousness can be seen to have a continued metaphysical existence without making some appeal to substance, which Locke has taken extraordinary pains to deny.<sup>16</sup>

There are two ways, Locke suggests, that we can think about consciousness as having a continuing existence. We can try to explain it insofar as it can be seen to fit with the traditional ontology of substances, modes, and relations or we can try to explain it insofar as it continues to exist as the objective I, self, or person we experience when we are conscious we are perceiving ideas. As for explanation by virtue of the traditional ontology, Locke tells us in II.xxvii.25 that were we to consider consciousness in this way, “the more probable Opinion is, that this consciousness is annexed to, and the Affection of one individual immaterial Substance”. Insofar as consciousness could be made to fit into the traditional ontological scheme, Locke’s best guess is that it would be considered a mode of a single individual thinking substance. This view would

---

<sup>16</sup> The only exception I have found is Yaffe, “Locke on Ideas of Identity and Diversity”, p. 226, who argues that we should attribute to Locke “the susceptibility-to-punishment theory of personal identity”. First, consciousness should be understood as the awareness of pleasure and pain. Second, x and y have the same consciousness “just in case the earlier’s actions are a potential source of pleasure or pain for the later”. Third, to avoid the objection that someone else’s action might be a source of my future pain, Yaffe includes the element of desert. So, whether I am the same person now as I was in the past is determined by the fact that I am susceptible to punishment now for what a past person did. Yaffe (pp. 228-29) then argues that Locke replaces the metaphysical fact of personal identity with the moral fact of just desert: “the metaphysical facts—the facts about who is the same person as whom—just are the moral facts; they are facts about who is appropriately punished or rewarded for whose past acts”. Although I think Yaffe’s interpretation cleverly finds an objective criterion on which to base personal identity, it is not clear that Locke would endorse the priority of the moral fact over the metaphysical fact. Although Locke does say that “almost all” our perceptions of ideas are accompanied by perceptions of ideas of pleasure and pain, not all our thoughts and actions are moral; that is, they do not all involve being compared to laws and deserving of reward or punishment. Therefore, the self does not continue to exist between times of nonmoral actions—or no actions at all. And although persons must be “capable of laws” (II.xxvii.26) they need not always be engaging in moral action to be persons.



be characteristic of Locke's time. But this is not the kind of metaphysical explanation Locke is interested in, for in the next sentence he says, "But let Men according to their divers Hypotheses resolve of that as they please" (II.xxvii.25). Since we can't know the real constitutions of things or the particularities involved in the relations between substances and their modes, Locke means to leave that kind of explanation up to the conjectures of others.

Rather, the immediately following passage reveals that Locke is interested in a different kind of explanation for consciousness. He begins by acknowledging the continually existing thing we are conscious of, namely the I, self, or person, when we are conscious we are thinking:

This every intelligent Being, sensible of Happiness and Misery, must grant, that there is something that is *himself*, that he is concerned for, and would have happy; **that this self has existed in a continued Duration more than one instant**, and therefore 'tis possible may exist, as it has done, Months and Years to come, without any certain bounds to be set to its duration; and may be the same *self*, by the same consciousness, continued on for the future. And thus, by this consciousness, he finds himself to be the *same self* which did such or such an Action some Years since, by which he comes to be happy or miserable now. **In all which account of the same self, the same numerical Substance is not considered, as making the same self: But the same continued consciousness**, in which several Substances may have been united, and again separated from it, which whilst they continued in a vital union with that, wherein this consciousness then resided, made a part of that same *self*. (II.xxvii.25, my emphasis in boldface)

Locke seems to be saying that despite our inability to penetrate to the real metaphysics of the external world or even our own minds, we have experience of a continually existing consciousness. Any being capable of happiness and misery "must grant", he says, that there is something enduring that he is aware of as himself and there is something for which he is concerned and wants to see happy rather than miserable. How, though, should we understand consciousness as having a continued existence?

Locke acknowledges the continued existence of consciousness by telling us that it has a "duration", which we know from his discussion of time as a technical term standing for the "distance between...the appearance of any two *Ideas* in our Minds" (II.xiv.3). According to

Locke, the duration designated by a single perception of an idea is an “instant”. He says, “Such a part of Duration as this, wherein we perceive no Succession, is that which we may call an *Instant*; and is *that which takes up the time of only one Idea* in our Minds, without the Succession of another, wherein therefore we perceive no Succession at all” (II.xiv.10). Furthermore, Locke asserts that ‘duration’ is a synonym for ‘continued existence’: “we call the Existence, or the Continuation of the Existence of our selves, or anything else, Commensurate to the succession of any *Ideas* in our Minds, the *Duration* of our selves, or any such other thing co-existing with our Thinking” (II.xiv.3). And even though we can have only a momentary idea of our own duration or of the duration of any other thing, Locke seems to be saying that things endure through our successive perceptions of ideas of them. We get an idea of the length of that duration by using the ideas we have as marks for measurement. As Locke says, “Duration in it self is to be considered, as going on in one constant, equal, uniform Course: but none of the measures of it, which we make use of, can be known to do so” (II.xiv.21). Furthermore, God can know the continued existence of things as they persist through those ideas by which we measure them. Locke tells us that “God’s infinite Duration being accompanied with infinite Knowledge, and infinite Power, he sees all things past and to come; and they are no more distant from his Knowledge, no farther removed from his sight, than the present” (II.xv.12). So, even though we cannot perceive the actual duration of things through the gaps in our perceptions of ideas, God can.

Also clear from II.xxvii.25 is that Locke understands consciousness as something that has “a continued duration more than one instant”. This confirms that Locke thinks it has an ongoing existence through each instant in which we are conscious of ourselves perceiving ideas. So, duration is the ongoing temporal existence of any thing as measured by a single perception of an

idea, as measured between two immediately successive ideas, or even as measured between any gaps in my awareness of myself as thinking – say between the last idea I had prior to falling into a dreamless sleep and the first idea I had upon waking. And even though there are gaps in my awareness of my own continued existence, the full extent of my duration is known by God. Therefore, in II.xxvii.25 we are told, first, that consciousness is something that endures through each momentary instance of a conscious state of awareness of itself, and second, the way in which it endures is not to be explained by appeal to the underlying metaphysics of substances and their modes.

More textual support that consciousness is something that endures through my momentary conscious states of myself is found in II.xxvii.10:

For as far as any intelligent Being can repeat the *Idea* of any past Action with the same consciousness it had of it at first, and with the same consciousness it has of any present Action; so far is it the same *personal Self*. For it is by the consciousness it has of its present Thoughts and Actions that it is *self* to it *self* now, and so will be the same *self* as far as the same consciousness can extend to Actions past or to come; and would be by distance of Time, or change of Substance, no more two *Persons* than a Man be two Men, by wearing other Cloathes to Day than he did Yesterday. (II.xxvii.10)

Locke tells us that as far as we have a unified experience of ourselves as thinking in the past and in the present we *actually are* the same ongoing consciousness. And we shouldn't think that there is no fact of a continually existing consciousness because the thoughts and actions that we are conscious we are perceiving are momentary successive mental states, just as we shouldn't think there is no fact of a continually existing man because he wears different clothes on successive occasions. Furthermore, it is the ongoing consciousness that "unites" successive perceptions of ideas: "yet 'tis plain consciousness", he says, "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, as well as it does the Existence and the Actions of the immediately preceding moment"

(II.xxvii.16). Whether the perception of the idea was in the distant past or whether it immediately precedes the present one, Locke seems to be saying that the same ongoing consciousness persists through and links together those momentary successive “Existences and Actions”.

We see the same sort of thinking when Locke links consciousness to accountability and concern:

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 which is founded in a concern for Happiness the unavoidable concomitant of consciousness, that which is conscious of Pleasure and Pain, desiring, that that *self*, that is conscious, should be happy. (II.xxvii.26)

For Locke, without consciousness there is no self to be concerned, accountable, happy, or miserable, for our perceptions of ideas of pain and pleasure would be no more to us than if they were anonymous or belonged to someone else: the thinking would not be ours. He says, “For if we take wholly away all Consciousness of our Actions and Sensations, especially of Pleasure and Pain, and the Concernment that accompanies it, it will be hard to know wherein to place personal Identity” (II.i.11). So, just as the consciousness of a present reflection on a past action is an experience of an ongoing self that is presently accountable for its own past, the consciousness of a present reflective consideration of a future action is an experience of an ongoing self presently concerned for its future. Unless consciousness actually is the same continuing thing, there is nothing to consider accountable now for what might have been done in the past and nothing to consider as moving toward the future to be concerned for. Therefore, Locke must be thinking that our accountability for the past and concern for the pleasures and pains we anticipate perceiving depends on the continued existence of the self we experience

when we are conscious we are perceiving ideas.<sup>17</sup> How, though, should we understand the metaphysics of a continuing consciousness without appealing to substance?

Perhaps Locke's analogy between consciousness and the continued existence of other things is helpful here. In II.xxvii.1-8, Locke argues that a particular thing is individuated when at a particular place and time there begins to exist the condition that makes it the kind of thing it is and that excludes all other things of the same kind from occupying the same time and place. What identifies it through time is the continued existence of that condition. A particular living thing, say a plant, animal, or human being, is individuated when its life begins to exist, and it continues to exist as the same animal as long as its life continues.<sup>18</sup> Now Locke is clear in II.xxvii.10 that he intends his analysis of the individuation and identity of animals to be in some way analogous to the individuation and identity of persons:

The Question being what makes the same *Person*, and not whether it be the same identical Substance, which always thinks in the same *Person*, which in this case matters not at all. 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. (II.xxvii.10)

It is reasonable to conclude that a person is individuated when consciousness begins to exist, and that the identity of the person corresponds to the continued existence of that particular consciousness.<sup>19</sup> So, Locke must be thinking that consciousness has a continued existence.

---

<sup>17</sup> We see the same priority of consciousness to happiness and misery and thus to concern in Cudworth, who argues that consciousness allows the self to be present to itself, which makes states of happiness and misery possible. Moreover, Cudworth, in *True Intellectual System*, p. 159, holds that this 'Duplication' of the self as subject and object via consciousness makes a being "to perceive it self to Do or Suffer, and to have a *Fruition or Enjoyment* of it self. More importantly, consciousness as the most fundamental referring of oneself to oneself makes other forms of self-reference possible." See Thiel, "Cudworth and Seventeenth Century Theories", p. 91.

<sup>18</sup> See II.xxvii.4-6.

<sup>19</sup> See II.xxvii.9.

Furthermore, the continued existence of consciousness cannot be merely the awareness of past thoughts and actions. Locke defines a person as “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” (II.xxvii.9). Since thinking is a succession of momentary ideas, each individual perception of an idea has no continued existence of its own.<sup>20</sup> Thus, the only way a person can be the same from thought to thought is if consciousness has a continued existence from thought to thought. The analogy would be that just as consciousness as a state of awareness of myself is a constituent of an act of thinking yet something psychologically different from an act of thinking, consciousness as the ongoing self bears some sort of relation to the train of successive perceptions of ideas, yet is something metaphysically different from them.<sup>21</sup> To use Locke’s words, just as the “Identity of the same *Man* consists...in **nothing but** a participation of the same continued Life, by constantly fleeting Particles of Matter, in a succession vitally united to the same organized Body” (II.xxvii.6, my emphasis), the identity of the same person consists in *nothing but* a participation of the same continued consciousness, by constantly fleeting perceptions of ideas. Although the analogy is not strict in that Locke seems to think that life just is the organized particles of matter while consciousness is something more than organized perceptions of ideas, the point is that both life

---

<sup>20</sup> Locke says, “Only as to things whose Existence is in succession, such as are the Actions of finite Beings, v.g. *Motion* and *Thought*, both which consist in a continued train of Succession, concerning their Diversity there can be no question: Because each perishing the moment it begins, they cannot exist in different times, or in different places, as permanent Beings can at different times exist in distant places; and therefore no motion of thought considered as at different times can be the same, each part thereof having a different beginning of Existence” (II.xxvii.2).

<sup>21</sup> See Shelley Weinberg, “The Coherence of Consciousness in Locke’s *Essay*”, *History of Philosophy Quarterly*, (2008) 25:1, 21-40 and Uriah Kriegel and Angela Coventry, “Locke on Consciousness”, *History of Philosophy Quarterly* (2008) 25:3, 221-42 for views that consciousness is kind of self-referential perception distinct from other acts of perception. See also Larry M. Jorgensen, (2010). “Seventeenth Century Theories of Consciousness” in *The Stanford Encyclopedia of Philosophy* (<http://plato.stanford.edu/>, 2010).

and consciousness are something metaphysically distinct from the fleeting things in which they participate.<sup>22</sup>

The problem is that we still don't know anything about the metaphysics of a continuing consciousness. Notice, though, that neither (for Locke) do we know anything about the metaphysics of a continuing life. All we know is that plants, animals, and human beings continue living as the same things, despite their constant change of substance. Locke explains that in plants "there is an Organization of those parts, as is fit to receive, and distribute nourishment", such that it continues as that plant. The organization "is that individual Life, which existing constantly from that moment both forwards and backwards in the same continuity of insensibly succeeding Parts" (II.xxvii.4). In the next two sections, Locke confirms that "The Case is not so much different in *Brutes*" (II.xxvii.5) or in human beings (II.xxvii.6). The important point is that we do not need a full account of the "life" of an animal in order to know that it has (there is) one and that it is something distinct from the fleeting particles of matter. Indeed, Locke thinks a full account of what life consists in is beyond our understanding: "We see Animals are generated, nourished, and move... These and like Effects we see and know: but the causes that operate, and the manner they are produced in, we can only guess, and probably conjecture" (IV.xvi.12). I suggest that when Locke tells us that the continued existence of consciousness is analogous to the life of animals, he is saying just this sort of thing. We don't need a full account of consciousness in order to know that we have (there is) one and that it is distinct from our fleeting perceptions of ideas.

---

<sup>22</sup> Atherton, "Locke's Theory of Personal Identity", p. 288, gives two reasons why what she calls the "integrative unity" of thinking is "not the same as the unity of consciousness": first, that we can imagine some "underlying cause" of the unity of our thoughts and ideas that allowed for two distinct consciousnesses, and second, that given his arguments against Descartes, Locke would not want to say that we know that what provides an essential integrative unity to our "thoughts, desires, and powers to move" is conscious thought.

Locke acknowledges that not everything we know to exist has a tidy metaphysical explanation. Winkler argues to a different purpose that Locke does not confine himself to traditional ontological categories, but rather sees them as “worthy of philosophical investigation in their own right”. Locke felt a greater allegiance “to certain features of the lived or commonsense world” than to traditional categories and philosophical problems.<sup>23</sup> Indeed, Locke’s interlocutors desiring more explanation of the how ideas – something else not neatly categorized – fit the substance/mode ontology were met with responses like the following:

Supposing ideas real spiritual things ever so much, if they are neither substances nor modes, let them be what they will, I am no more instructed in their nature, than when I am told they are perceptions, such as I find them. (W ix: 220)<sup>24</sup>

This therefore may be a sufficient excuse of the ignorance I have owned of what our ideas are, any farther than as they are perceptions we experiment in ourselves; and the dull unphilosophical way I have taken of examining their production, only so far as experience and observation lead me; wherein my dim sight went not beyond sensation and reflection. (W x: 256)

If we take Locke at his word that “’Tis not the Unity of Substance that comprehends” personal identity (II.xxvii.7), and he intends to leave the explanatory details of a continuing consciousness “up to the conjectures of others” (II.xxvii.25), then anything other than the metaphysical fact of consciousness must be left undetermined. But, as with ideas, the lack of a satisfying ontological categorization does not preclude the fact of an existence of which we are aware on the basis of experience and observation.<sup>25</sup>

---

<sup>23</sup> Winkler, “Locke on Personal Identity”, pp. 165ff.

<sup>24</sup> *The Works of John Locke*, 10 vols. (London, 1823). All additional references will be in the body of the paper by volume number: page number.

<sup>25</sup> William L. Uzgalis, “Relative Identity and Locke’s Principle of Individuation”, *History of Philosophy Quarterly*, vol. 7, no. 3 (July 1990), p. 287, relies on the analogy to the life of animals to argue that Locke means persons to be mixed modes. Although Uzgalis’s proposal is tempting since it allows Locke to have immaterial substance play a role, but not as constitutive of the identity of the person, I do not think consciousness can be a mixed mode for the reason that mixed modes are mental constructions that have no continued existence of their own. Locke is clear on this point: “All other things [but substances] being but Modes and Relations ultimately terminated in Substances, the Identity and Diversity of each particular Existence of them too will be by the same way determined: Only as to



Locke has the same attitude with respect to the existence of the external world. The lack of full rational warrant that external objects really exist in the moment we are sensing them does not keep Locke from the bold claim that we need not be concerned with the skeptical possibility. Locke is clear that we have an experience of the existence of external objects even if we do not have a full account concerning that fact. He says, “Several effects come every day within the notice of our Senses, of which we have so far *sensitive Knowledge*: but the causes, manner, and certainty of their production...we must be content to be ignorant of. In these cases, we can go no farther than particular Experience informs us of matter of fact” (IV.iii.29). With respect to the continuing consciousness we can say that just as an objectively existing world is present and certain to me in sensitive knowledge, an objectively existing ongoing self is present and certain to me in virtue of being conscious of myself as perceiving ideas.

Despite Locke’s own unwillingness to provide more, we can perhaps say something about what the metaphysics of the ongoing consciousness could be from other things Locke says, particularly about the metaphysical possibility that matter thinks. He asserts, “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 degrees of sense, perception, and thought” (IV.iii.6). There are several interpretations of what Locke could mean by God’s “superaddition” of thought to matter, especially given his IV.x. argument for the existence of God in which he seems to maintain that thinking cannot be material. To provide an

---

things whose Existence is in Succession, such as are the Actions of finite Beings, *v.g. Motion and Thought*, both which consist in a continued train of Succession, concerning their Diversity there can be no question: Because each perishing the moment it begins, they cannot exist in different times, or in different places, as permanent Beings can at different times exist in different places; and therefore no motion or thought considered as at different times can be the same, each part thereof having a different beginning of Existence” (II.xxvii.2). Mixed mode sortals, therefore, will capture either the continuity of the substance in which they terminate or they are momentary perceptions of ideas. Therefore, Uzgalis has accounted for the continuity of consciousness by appeal to thinking substance or he has not accounted for it at all.

explanation that reconciles these apparently contradictory claims it is important to see what the superaddition of thinking to matter could amount to in that it must be more than just the way in which particles of matter are in motion. Lisa Downing characterizes one group of interpretations as advocating an “extrinsic powers” reading of superaddition,<sup>26</sup> which can be seen as either the arbitrary attaching (or “gluing”) of thinking powers onto bodies<sup>27</sup> or the attaching to bodies thinking powers that are governed by divinely established laws of nature.<sup>28</sup> On the other hand, we can see superaddition as involving the way in which God has created the real constitutions of things such that they give rise either mechanically<sup>29</sup> or non-mechanically to thought. Since it would take us too far afield to go into each of these views, let’s focus just on Downing’s account, since I think it gives us more to work with insofar as consciousness is concerned, and it is a plausible reconciliation of superaddition and Locke’s proof for the existence of God.

Downing argues for what she calls “essentialist superaddition.” First, we should see Locke as not committed to the truth of corpuscularianism, but to a “much more abstract metaphysics of real constitution and primary quality. His ontology of mind reveals these same commitments, together with the same official agnosticism about what the real constitution of bodies and minds are actually like.”<sup>30</sup> It would, however, include a commitment to essentialism<sup>31</sup> with respect to

---

<sup>26</sup> Lisa Downing, “Locke’s Ontology” in ed. L. Newman, *The Cambridge Companion to Locke’s Essay* (Cambridge: Cambridge University Press, 2007), pp. 365-69.

<sup>27</sup> G. W. Leibniz, *New Essays on Human Understanding*, trans. P. Remnant and J. Bennett (Cambridge: Cambridge University Press, 1981), pp. 65-65.

<sup>28</sup> Rae Langton, “Locke’s Relations and God’s Good Pleasure,” *Proceedings of the Aristotelian Society* 100 (2000), pp. 75-91, Edwin McCann, “Locke’s Philosophy of Body” in ed. V. Chappell, *The Cambridge Companion to Locke* (Cambridge: Cambridge University Press, 1994), pp. 56-88, Matthew Stuart, “Locke on Superaddition and Mechanism,” *British Journal for the History of Philosophy*, vol. 6 (1998), pp. 351-79.

<sup>29</sup> Michael R. Ayers, Mechanism, Superaddition, and the Proof of God’s Existence,” *The Philosophical Review*, vol. 90 (1981), pp. 210-51.

<sup>30</sup> Downing, “Locke’s Ontology”, p. 378.

<sup>31</sup> “Essentialism” is defined as “the view that the qualities and behavior of a body follow from its real constitution (some particular configuration of its intrinsic and irreducible qualities), together with the real constitutions of other

the real constitutions of things, but not with respect to what we experience of those things from our perspective (the nominal essence) or those sorted constitutions (real essences) corresponding to the nominal essences.<sup>32</sup> Superaddition, then, although it might look to be the addition of extrinsic (thinking) properties to matter is only so from our point of view. That is, thinking properties don't follow from our idea of body (the nominal essence). But that doesn't mean that there might not be a configuration of the underlying real constitutions that has thinking as an intrinsic property. Downing proposes two ways of understanding the real constitutions, such that they can give rise to thinking. It could be that "God gives some stuff a nonmechanical real constitution that allows it to manifest thought as well as extension and solidity, while he gives nonthinking material stuff a different type of real constitution, which might well be purely mechanical."<sup>33</sup> The other way to see it is that there could be "uniform nonmechanical real constitutions differently configured: the real constitutions of all material things (things that satisfy the nominal essence of matter) are nonmechanical; God configures some of them so as to allow them to think."<sup>34</sup> Note that this explanation entails that mechanism is at best an incomplete explanation of bodies.

But, Downing continues, Locke seems to draw just such a conclusion from Newton's discovery of gravity:

---

bodies and the spatial relations among bodies." Downing, *Ibid.*, p. 368, suggests that "constitutionalism" might be more accurate.

<sup>32</sup> Note that Downing, *Ibid.*, p. 370-72, uses 'real constitution' to refer to "the configuration of intrinsic and irreducible qualities responsible for all of a thing's qualities/powers, while reserving 'real essence' for constitutions relative to nominal essences, that is for whatever constitution is responsible for a set of observable properties enshrined by us as a kind." This allows us to distinguish a configuration of real constitutions making up a particular individual thing from that configuration that is denoted by our idea of the kind to which the individual belongs. See also Paul Guyer, "Locke's Philosophy of Language," in ed. V. Chappell, *The Cambridge Companion to Locke* (Cambridge: Cambridge University Press, 1994), pp. 133-34 and Owen, *Hume's Reason*, pp. 108.

<sup>33</sup> *Ibid.*, p. 372.

<sup>34</sup> *Ibid.* In both suggestions, Downing takes 'nonmechanical' to mean "not merely or strictly mechanical", that is, not exhausted by size, shape, solidity, and motion/rest."

The gravitation of matter towards matter, by ways inconceivable to me, is not only a demonstration of that God can, if he pleases, put into bodies powers and ways of operation above what can be derived from our idea of body, or can be explained by what we know of matter, but also an unquestionable and every where visible instance, that he has done so. And therefore in the next edition of my book I shall take care to have that passage rectified. (W IV: 467-68).

The crucial point I want to take from Downing's account is that Locke's agnosticism with respect to what we can know of the real constitutions of things beyond the ideas we have opens the door to metaphysical entities (configurations of real constitutions) that give rise to our ideas, but are not fully captured by our ideas. So what does this have to do with the metaphysical fact of consciousness? Just as there are real constitutions that have intrinsic to them the ability to give rise to our idea of matter or our ideas of thinking, there are real constitutions configured such that they give rise to our ideas or experience of consciousness. Thus, there would be a metaphysical fact of consciousness that is not reducible either to thinking (or to thinking substance) or to my experience of being conscious of myself. In addition, if we interpret Locke's view to be that the real constitutions have powers, then just as there are real constitutions with a power to think that is not always manifest in actual thinking, there are other real constitutions with a power to be conscious that is not always manifest in the consciousness that I am thinking. This sort of view is also consistent with Locke's claims that consciousness is "inseparable" from thinking. That is, it is consistent that whenever I am thinking I am conscious that I am thinking *and* that thinking and consciousness are distinct metaphysically.

### 3. The Phenomenological Fact of a Continuing Consciousness

In the first section, I argued that Locke uses the term 'consciousness' ambiguously to refer both to the state of awareness of myself as thinking and also to the ongoing self I am aware of in that conscious state. I then argued, in section 2, why we should read Locke's theory of personal

identity to include a metaphysical fact of a continuing consciousness. In this section, I will explain how the metaphysical fact of consciousness is revealed to me in my experience of my own thinking. In other words, how are we conscious of a continuing self in virtue of a succession of conscious states of ourselves as perceiving ideas? First, in his version of the *cogito* argument, Locke is clear that we are conscious of our own synchronic existence whenever we have an individual perception of an idea: “In every Act of Sensation, Reasoning, and Thinking, we are conscious to our selves of our own Being; and, in this Matter, come not short of the highest degree of *Certainty*” (IV.ix.3). Second, Locke seems to think that the consciousness we have of ourselves perceiving at any one time can be “joined” to the consciousness of ourselves perceiving at another time to establish some sort of experience of a continuing self:

That with which the *consciousness* of this present thinking thing can join it self, makes the same *Person*, and is one *self* with it, and with nothing else; and so attributes to it *self*, and owns all the Actions of that thing, as its own, as far as that consciousness reaches, and no farther; as every one who reflects will perceive. (II.xxvii.17)

The temptation, which should be resisted on pain of circularity, is to read this passage as saying that the continuing consciousness is constructed from successive conscious states. A more sympathetic reading sees the passage as suggesting that we have an *experience* of continuity in successive states of consciousness: we experience our conscious states of ourselves as linked in a continuing stream of consciousness.

How should we see our experience of ourselves as linked in consciousness? Descartes held the view that we can have ideas in the mind simultaneously, and there is no reason to think that Locke considered it differently.<sup>35</sup> The composition of complex ideas of substances requires that we have several simple ideas in the mind simultaneously as does reflecting on the perception of

---

<sup>35</sup> According to Descartes “the soul is capable of thinking of more than one thing at the same time, and of continuing with a particular thought which it has”. See “Conversation with Burman”, vol. III, p. 335; (AT V: 149).

an idea and perceiving a relation between ideas.<sup>36</sup> If ideas are perceived in the mind simultaneously, then we are also conscious of ourselves as perceiving those ideas. Consider first an example in which we have knowledge in virtue of perceiving an agreement involving simultaneous ideas. “Thus the Mind,” Locke says, “perceives that *White* is not *Black*, that a *Circle* is not a *Triangle*, That *Three* are more than *Two*, and equal to *One* and *Two*. Such kind of truths the Mind perceives at the first sight of the *Ideas* together” (IV.ii.1). To take just one of these examples, I am conscious I am perceiving an idea of the number three and I am conscious I am perceiving an idea of the number two. But my experience in being conscious that I am perceiving an agreement of ideas seems to be that I am conscious of the same self as perceiving both ideas. I experience the two perceptions of ideas as held together by one consciousness. The resulting conscious state is of myself as knowing intuitively that three are more than two. The experience of the unity of consciousness is such that there can be one and the same consciousness linking two perceptions of ideas in a proposition known to be true.<sup>37</sup>

The experience of the unity of consciousness is also evident in cases where we engage in demonstrative reasoning. For example, first there is a perception of the idea ‘P→Q’ followed by the perception of the idea ‘P’. It seems that in order to intuit ‘Q’ the perception of the idea ‘P’ must be brought into contemporaneity with the perception of the idea ‘P→Q’. Perceptions of ideas would, in a sense, overlap one another in my consciousness in something more like an

---

<sup>36</sup> In the mental operation of “*COMPOSITION*” the mind “puts together several of those simple ones [ideas] it has received from Sensation and Reflection, and combines them into complex ones” (II.xi.6). “Reflection”, Locke explains, “is the *Perception of the Operations of our own Minds* within us, as it is employed about the *Ideas* it has got” (II.i.4), which implies that the reflective perception is in the mind at the same time as the perception that is its object. Additionally, in comparing ideas “the Mind so considers one thing, that it does, as it were, bring it to, and set it by another, and carry its view from one t’other” (II.xxv.1). In all three cases, ideas are perceived in the mind simultaneously.

<sup>37</sup> See Owen, *Hume’s Reason*, p. 47.

extended present.<sup>38</sup> This can be achieved if thinking is temporally extended in the sense that one idea can remain in the mind while another idea is perceived. In this kind of case, perceptions of ideas are united or gathered together in consciousness, perhaps we should say “joined” such that I am conscious of myself as temporally extended through these successive perceptions of ideas. Locke doesn’t specifically describe the perception of an agreement as the result of reasoning as temporally extended. But he does suggest something like it in his description of how we employ intervening ideas in a demonstration:

The Reason why the Mind cannot always perceive presently the Agreement or Disagreement of two *Ideas* is, because those *Ideas*, concerning whose Agreement or Disagreement the Enquiry is made, cannot by the Mind be so put together, as to shew it. In this Case then, when the Mind cannot so bring its *Ideas* together, as by their immediate Comparison, and as it were Juxtaposition, or application one to another, to perceive their Agreement or Disagreement, it is fain, by the Intervention of other *Ideas* (one or more, as it happens) to discover the Agreement or Disagreement, which it searches; and this is that which we call *Reasoning*. (IV.ii.2)

Locke seems to be saying that the intervening idea needed to show the agreement is brought into contemporaneity with the ideas already in the mind, which allows for the perception of agreement between the previously juxtaposed ideas. This would be the sense in which the original ideas overlap with the new intervening idea in something like an extended or continuing present. In being conscious of my individual successive perceptions of ideas as an uninterrupted

---

<sup>38</sup> Mackie, *Problems From Locke*, pp. 180-81 considers and rejects this (due to the lack of a metaphysical underpinning) as a possible revision of Locke’s theory. “Locke could have said that what makes me the same person from one moment to the next, while I’m awake, is a genuine co-consciousness of experiences, an overlapping of specious presents. What I take to be happening now is not instantaneous, not confined within a knife-edge present, but fills some short stretch of time, and these nows overlap one another. For each now there is, we might say, an I-occurrence, and successive I-occurrences will similarly overlap and fade into one another: thus these I-occurrences build up into a continuous I-history. The single person, the I, is what is taken to be there, all at once, at each moment in an I-history. This constitutes my identity throughout any one waking day, any period throughout which I am continuously conscious.” Periods of unconsciousness are filled by what Mackie calls “memory bridges”. Mackie ultimately rejects the revision because he thinks that Locke rejects it. I would argue that because Mackie is reading Locke to have an appropriation theory of sorts, he misreads Locke’s claims concerning “distinct incommunicable consciousnesses” in II.xxvii.20 to mean different collections of conscious states or memories, rather than a point about the relation between a single man and two metaphysically distinct consciousnesses.

continuous stream of thinking, which seems to happen in reasoning, I am conscious of myself as something that continues to exist, as having a temporally extended existence.<sup>39</sup>

We see the same sort of thing happening in reflection, in particular on a past thought, as the perceptual situation in which consciousness is “extended backwards” in the appropriation of a past action. When I currently reflect on a perception of a past thought or action I have in my mind simultaneously two different perceptions of ideas. I have the present reflection and I have the revived idea that is the object of the reflection. Internal to the present reflective perception is the consciousness that *I am reflecting*, and internal to the perception of the revived idea is the consciousness that *I am remembering*. If there is the same internal experience of myself as both reflecting and remembering, then there is a unified experience of myself as perceiving from the past to the present,<sup>40</sup> which can be expressed by saying *I am conscious I am perceiving an idea I*

---

<sup>39</sup> The problem of establishing how objects of consciousness can have a temporal existence comes front and center in later attempts to give descriptive analyses of the nature of conscious experience. Dan Zahavi, *Subjectivity and Selfhood*, p. 56, describes the problem as Husserl undertook it in this way: “In his *Vorlesungen zur Phänomenologie des Zeitbewusstseins*, Husserl asks how it is possible for us to be conscious of temporal objects. His well-known thesis is that a perception of a temporal object, as well as the perception of succession and change, would be impossible if consciousness provided us only with access to the pure now-phase of the object and if the stream of consciousness itself was a series of unconnected points of experiencing, like a line of pearls. Had our perception been restricted to being conscious of that which exists right now, it would have been impossible to *perceive* anything with temporal extension and duration, for a succession of isolated, punctual, conscious states does not, as such, enable us to experience succession and duration. Since we are obviously conscious of succession and duration, we must acknowledge that our consciousness, one way or another, can encompass more than that which is given right now. Although we can be co-conscious of that which has just been, and that which is about to occur, the crucial question remains: *how* can we be conscious of that which is no longer, or not yet, present to our consciousness? Some have suggested that imagination or memory might play a crucial role. There is however an obvious difference between seeing a movement (that necessarily extends in time) or hearing a melody, and remembering or imagining either. Husserl’s own alternative is to insist on the *width of presence*.”

<sup>40</sup> In II.i.10 Locke seems to equate consciousness with sensibility, and in Locke’s time consciousness was thought to involve a “sentiment interior”. Davies, ‘*Conscience as Consciousness*’, pp. 30-32, notes that in his *Examination of Malebranche, The Complete Works of Locke*, (London, 1812), vol. ix, p. 245, Locke uses the term ‘consciousness’ as equivalent to Malebranche’s ‘interiour sentiments’, which for Malebranche refer to a kind of imperfect knowledge we have of our souls that comes from what we “feel within ourselves”. Moreover, Locke’s French translator, Pierre Coste, who is known to have worked closely with Locke, translated ‘consciousness’ in different ways, but he often used the words ‘sentiment’, ‘sentir’, and even ‘sentir interieurement’. This last example is employed to capture what Locke meant in I.iv.20, when he says “Whenever memory brings any idea into actual view, it is with a consciousness [sentir interieurement] that it had been there before”. In each case, though, as Davies points out, Coste is emphasizing the interior nature of the experience of our own thinking.



*had before.*<sup>41</sup> Thus, I have a unified experience of being conscious of two perceptions of ideas as opposed to an experience that would be described as including two perceptions of ideas with two consciousnesses. Moreover, when I have this kind of unified experience of myself as thinking in both the past and the present, I have the experience of my consciousness as “reaching” or “extending” into the past. Similar arguments can be given for cases of reflection that do not involve memories.

Remember also that Locke defines a person not as a substance, but as a “thinking intelligent Being, that has reason and reflection” (II.xxvii.9). Moreover, as we’ve just seen, reflection is one way – most explicitly when we reflect on memory – in which the continued existence of a consciousness is made present to itself. Because of reflection, I can be conscious of myself, or my person, as something that exists diachronically. Therefore, Locke’s claim that a person must have reflection is evidence that he thinks consciousness is something that continues to exist through momentary conscious states of itself as thinking. In our conscious states of ourselves as thinking the metaphysical fact of our continued existence is revealed to us as a phenomenological fact: we have the experience of ourselves as existing diachronically – as temporally extended.

#### 4. Consciousness and Memory – Two Questions

I began by outlining the problems left unsolved by leading interpretations of Locke’s theory of personal identity. I have argued for an interpretation that is textually supported and solves these problems. But there still remain two questions. First, aren’t those passages arguing against the view that substance is the criterion for personal identity equally powerful against the view

---

<sup>41</sup> Section 4 contains a full explanation of an instance of remembering.

that a continuing consciousness is the criterion for personal identity? That is, what about Locke's emphasis on occurrent mental states (memory, or what I can attribute to myself) as necessary for personal identity? Locke argues that substance cannot be the criterion, since being conscious of one's past actions (having occurrent mental states) is what generates concern, moral responsibility, and a sense of oneself through time. Therefore, when it comes to personal identity it seems that Locke is really interested in a psychological, not a metaphysical criterion. Second, what is the relation between consciousness and those occurrent mental states (memory)? These two questions are related in that those passages arguing that personal identity does not consist in substance are mostly the same passages in which Locke refers to "consciousness" as "consciousness of past actions", namely occurrent (or remembered) mental states. Although this is textual evidence that Locke sees the criterion of personal identity as only psychological, I will argue that Locke's considered view of the relation between consciousness and memory cuts against this reading by requiring an actual consciousness/mind/body/composite to generate those mental states required for a first personal concern for oneself and one's actions. The importance of our psychological states for our experience of our own identity, therefore, does not preclude that Locke's theory also includes a metaphysical criterion.

As for the first question, why would Locke allow a metaphysical criterion, a continuing consciousness, as an objective criterion for personal identity when he argues so vehemently against the view that substance can play the same role? I have four responses. First, I suggest that Locke is not against a metaphysical criterion, but only against allowing substance to serve as that criterion. As we know, Locke is agnostic as to the ultimate metaphysical nature of the world, and what we can know of it is only through our experience, namely through the ideas we have. Moreover, Locke is interested in the spatio-temporal identity of things, which suggests

that he is interested in actual things as they persist through time. Indeed, the first eight sections of II.xxvii are devoted to an analysis of what we should consider when we ask about the spatio-temporal identity conditions of different kinds of things. In II.xxvii.8, Locke makes the point that old definitions of ‘human being’ should not rule our considerations, because the traditional idea (‘rational animal’) is ill equipped to capture what we really think human beings are. Prince Maurice’s story of the rational parrot that we would not call a man, since it doesn’t have the right body shape, is evidence that Locke wants to reconsider identity, including personal identity, conditions, unhindered by traditional ideas.

Second, Locke thinks, at least in principle, that bodies, minds, and consciousnesses can be separated. But though Locke’s thought experiments are meant to drive our thinking about what constitutes personal identity, I take them to represent mere possibilities. So, even though it is possible that whatever configuration of real constitutions responsible for my conscious states can exist on its own, it will not be found without those configurations responsible for my mind (thinking) and my body.<sup>42</sup> Locke thinks we are mind/body composites (human beings) regardless of what minds or bodies turn out to be. I am arguing that what he really thinks is that we are consciousness/mind/body composites, where each component is ontologically distinct regardless of what the underlying real constitutions turn out to be.<sup>43</sup> Although there is a metaphysical fact of consciousness, it should not be seen as reducible to worn out notions of thinking or bodily substance.

---

<sup>42</sup> See II.i.20, where Locke says, “I see no reason therefore to believe, that the *Soul thinks before the Senses have furnish’d it with Ideas* to think on.” I take this to mean that both the body (sense organs) and the mind are necessary for thinking.

<sup>43</sup> See also Vere Chappell, “Locke and Relative Identity,” *History of Philosophy Quarterly*, 6:1 (1989), p. 78ff.

An ontological scenario something like this is consistent with Locke's claims about the relationships between our mental states. Locke tells us that we are conscious of all of our thinking (II.i.19) and that consciousness "is inseparable from thinking" (II.xxvii.9). But he also claims that we have periods of dreamless sleep in which there is no thinking and therefore also no consciousness that we are thinking. I have argued that consciousness in one sense is a state of awareness *that I am thinking*. Without consciousness, there is no experience of an *I* thinking. As has been noted in the literature, seeing Locke to understand consciousness as identical to thinking – either first order perception or second order reflection – conflicts with other important commitments.<sup>44</sup> We can, however, see consciousness as a self-conscious reflexive perception internal to (yet not identical to) any act of perception.<sup>45</sup> This relation between our mental states is fully consistent with the claim that metaphysically we are created as body/mind/consciousness composites – or single beings in which these parts are somehow united and act together. Likewise, the mental state that is consciousness is not identical to the mental state that is the perception of an idea, yet they are somehow united and work together. So, regardless of whatever mind/body composite my consciousness is annexed to, I will experience those thoughts as my own. I will be concerned for all those actions as if I had been unified with the mind/body composite that did them. This is precisely why Locke, in II.xxvii.13, appeals to the "goodness of God" not to disrupt the mind/body/consciousness composite by transferring a consciousness between thinking substances.<sup>46</sup> I suggest, therefore, that Locke's arguments against substance as the criterion for personal identity are not in the service of denying that something persists as the

---

<sup>44</sup> Mark Kulstad, "Locke on Consciousness and Reflection" in *Leibniz on Apperception, Consciousness, and Reflection* (München: Philosophia Verlag, 1991), was the first to point out apparent problems.

<sup>45</sup> See again, Weinberg, "The Coherence of Consciousness," Kriegel and Coventry, "Locke on Consciousness," and Jorgensen, "Seventeenth Century Theories of Consciousness."

<sup>46</sup> A more thoroughgoing analysis of this passage comes shortly.

criterion for personal identity that is also metaphysically responsible for our conscious states, but only that he wants to get away from an explanation that claims that just by virtue of tradition (or what we “ordinarily” think) we know what it is.

Third, I suggest that although being aware of our occurrent mental states as our own is necessary for having an experience of ourselves as temporally extended and for experiencing concern for past and future pleasures and pains, it is not sufficient for personal identity. The reason is that it affords us only first personal access to whether or not a present self is the same as a past self. To read it any other way is to uncharitably saddle Locke either with the circularity problem or with incoherence. Thus, given the textual evidence we should see Locke as also including in his theory that consciousness has metaphysical duration. The ambiguity in Locke’s use of ‘consciousness’ to refer both to a state of awareness and a continuing self can be seen as two aspects of personal identity that have separate tasks: consciousness as a mental state serves to tell me from my own first personal observation and experience about my own identity, such that I experience my concern for myself and my own moral responsibility. The unknown configurations of the real constitutions of things serve to tell anyone else who can know it from the third personal objective point of view (that is, God’s) whether I really am the same self.<sup>47</sup>

Fourth, there is a textual argument concerning the relation of consciousness to memory that supports the view that consciousness is distinct from memory and need not be understood simply as awareness of our occurrent mental states. It’s true that this argument is in apparent tension

---

<sup>47</sup> Note also that this is fully consistent with Locke’s claim that the drunkard is responsible for actions she can’t remember. Because the Drunkard cannot prove that her inability to attribute an action to herself is due to the transference of consciousness, and we are unable to consult God, we must do the best we can epistemically from the third personal objective point of view. Locke says, “Laws punish both with a Justice suitable to their way of Knowledge...For though punishment be annexed to personality, and personality to consciousness, and the Drunkard perhaps be not conscious of what he did; yet Humane Judicatures justly punish him; because the Fact is proved against him, but want of consciousness cannot be proved for him” (II.xxvii.22).

with those passages that seem to identify consciousness with memory. For example, when Locke is drawing a distinction between consciousness and the soul (or thinking substance) he will refer to consciousness as “consciousness of past Actions”, as if consciousness (and therefore the criterion for personal identity) just is occurrent memory. In the II.xxvii.14 discussion of the transmigration of souls, Locke answers “No” to the question whether I am the same person as Nestor or Thersites at the siege of Troy if I have the same soul but I have no consciousness of any of their past actions. “Can he be concerned in either of their Actions? Attribute them to himself, or think them his own more than the Actions of any other Man, that ever existed?” Regardless of the fact that I have the same soul, if I am not conscious of what another did, which here means that I have no occurrent memory of his actions, I am not the same person. So, Locke seems to be thinking that consciousness is equivalent to memory.

We find something similar in Locke’s example of the prince and the cobbler.

Should the Soul of the Prince, carrying with it the consciousness of the Prince’s past Life, enter and inform the Body of the Cobler as soon as deserted by his own Soul, every one sees, he would be the same person with the Prince, accountable only for the Prince’s actions.... But he would be the same Cobler to every one besides himself. (II.xxvii.15)

Since the consciousness that determines the Prince’s identity is the “consciousness of the Prince’s past Life,” Locke seems to be saying that the Prince’s memories are what is important about the Prince’s consciousness rather than that the Prince’s consciousness is something distinct from those occurrent thoughts. But notice also that in both cases Locke is emphasizing the first personal point of view. In the case of Nestor and Thersites, I cannot identify myself with another if I can’t from my own point of view appropriate to myself what she did. In the last line of the passage about Prince and the Cobler, Locke emphasizes that the Prince recognizes the experience of his own mental states as his own even though from the outside he is still the same Cobler to

everyone else. The interpretation I am advocating does not deny that Locke uses ‘consciousness’ ambiguously to refer both an occurrent mental state of awareness of ourselves and to the ongoing self I am conscious of in that mental state. In some passages, the state of awareness is emphasized while in others the ongoing self is emphasized.<sup>48</sup>

A passage in which we see the ambiguity clearly, yet Locke emphasizes the metaphysical (objective or third personal) reading of consciousness comes in II.xxvii.13. Here, Locke argues that were a consciousness transferred from one thinking substance to another there would be a problem of false memory. In this traditionally controversial section,<sup>49</sup> Locke begins by considering “whether the consciousness of past Actions can be transferr’d from one thinking Substance to another.”

I grant, were the same consciousness the same individual Action, it could not: But it being a present representation of a past Action, why it may not be possible, that that may be represented to the Mind to have been, which never really was, will remain to be shewn. And therefore, how far the consciousness of past Actions is annexed to any individual Agent, so that another cannot possibly have it, will be hard for us to determine, till we know what kind of Action it is, that cannot be done without a reflex Act of Perception accompanying it, and how perform’d by thinking Substances, who cannot think without being conscious of it. But that which we call the *same consciousness*, not being the same individual Act, why one intellectual Substance may not have represented to it, as done by it self, what it never did, and was perhaps done by some other Agent, why I say such a representation may not possibly be without Matter of Fact, as well as several representations in Dreams are, which yet, whilst dreaming, we take for true, will be difficult to conclude from the Nature of things. And that it is never so, will by us, till we have clearer views of the Nature of thinking Substances, be best resolv’d into the Goodness of God,

---

<sup>48</sup> Perhaps the most difficult passage for my interpretation is Locke’s example of the “*Day and the Night-man*” in II.xxvii.23, where Locke seems to equate a lack of consciousness with forgetfulness. Again, I would argue that in these cases Locke is focusing on the first-personal awareness of ourselves.

<sup>49</sup> For example, Mackie, *Problems From Locke*, pp. 184-85, argues that Locke’s claims about false memory as the result of consciousness transference in II.xxvii.13 are inconsistent with the claim that consciousness retains memory, as we seem to find it in the example of the Prince and the Cobbler. For another view that this passage reveals an inconsistency in Locke’s theory see Anthony Flew, “Locke and the Problem of Personal Identity”, *Philosophy* 26 (1951); slightly revised and reprinted in *Locke and Berkeley*, ed. C.B. Martin and D.M. Armstrong (Garden City, N.Y.: Anchor/Doubleday, 1968), pp. 155-78. Those attempting to defend the consistency of this passage with the rest of Locke’s theory include William P. Alston and Jonathan Bennett, “Locke on People and Substances,” *Philosophical Review* (1988), 97:1, pp. 25-46, Paul Helm, “Locke’s Theory of Personal Identity,” *Philosophy* 54 (1979), pp. 173-85, and Don Garrett, “Locke on Personal Identity, Consciousness, and ‘Fatal Errors’,” *Philosophical Topics* 31:1 &2 (2003), pp. 95-125.

who as far as the Happiness or Misery of any of his sensible Creatures is concerned in it, will not by a fatal Error of theirs transfer from one to another that consciousness, which draws reward and Punishment with it. (II.xxvii.13)

Since the consciousness being transferred is described as the “consciousness of past Actions,” we could read Locke to be saying that consciousness is memory. Thus, when consciousness is transferred from one thinking substance (“Agent”) to another the memories of the actions are transferred too. So when I occupy the new substance, that substance is now responsible for all those actions that I, as the transferred consciousness, either truly or falsely remember having done. But that can’t be Locke’s view, for “punishment,” he says, “. . . [is] annexed to personality and personality to consciousness” (II.xxvii.22). Consciousness, after all, determines the identity of the person, and it is the identity of the person that serves ultimately as the necessary condition for moral responsibility and just punishment, not the substance. So that the new substance is punished by virtue of the transferred consciousness shouldn’t be problematic. There would be no reason for Locke to think, as he seems to, that God’s intervention is needed to save us from this possibility.

Moreover, this reading doesn’t seem to be consistent with Locke’s view that consciousness transference results in the problem of false memory at all. The reason is that if consciousness is nothing more than appropriated occurrent thoughts, and consciousness is also the criterion for my identity, then it seems that nothing *I* remember could be false. The problem of false memory occurs only if the transferred consciousness is held accountable for something it didn’t do, which would have to be something not already a part of its occurrent memory. Therefore, reading consciousness as metaphysically distinct from memory makes sense of the problem of false memory. So although these passages emphasize the importance of the first personal aspect of



consciousness for generating concern for my past and future self, they do not preclude that Locke also sees consciousness as something metaphysical.<sup>50</sup>

Indeed, Locke's own account of memory seems to include a metaphysical difference between consciousness, memory, and the power of the mind to revive ideas that is suggestive of a mind/body/consciousness composite in which there are distinct powers working together to result in an experience of remembering. This brings us to the second question: what is the relation of consciousness to memory? In Locke's II.x account of memory, we find strong evidence that consciousness, although essential to experiencing an occurrent idea as a memory, is not identical to memory. In the following passage, Locke provides some explanation for how we should think of memory and the role of consciousness in having memories:

For the narrow mind of Man, not being capable of having many *Ideas* under View and Consideration at once, it was necessary to have a Repository, to lay up those *Ideas*, which at another time it might have use of. But our *Ideas* being nothing, but actual Perceptions in the Mind, which cease to be any thing, when there is no perception of them, this *laying up* of our *Ideas* in the Repository of Memory, signifies no more but this, that the Mind has a Power, in many cases, to revive Perceptions, which it has once had, with this additional Perception annexed to them, that it has had them before. And it is in this Sense it is said that our *Ideas* are said to be in our Memories, when indeed, they are actually no where, but only there is an ability in the Mind, when it will, to revive them again; and as it were paint them anew on it self, though some with more, some with less difficulty; some more lively, others more obscurely. (II.x.2)

First, we should see memory is a kind of "Store-house" or "Repository" for ideas that the mind can no longer have in view. But having memories in a repository is not to be confused with having a collection of ideas as normally experienced, say as images of red and yellow, that can be plucked out as is and re-introduced into the mind. Rather, Locke explains that memories cease to be ideas – in fact they "cease to be any thing" – when they are not the objects of acts of

---

<sup>50</sup> This is not to deny that there is a textual tension, but only to show, first, that not all cases in which Locke argues that consciousness should not be identified with substance are also cases that lend to the view that consciousness consists in occurrent thought. Second, it should also increase the plausibility both that Locke's textual references to consciousness as a something distinct and having metaphysical duration are meant seriously. Finally, if we take those references at face value, then we can rescue Locke's theory from incoherence.

perception. The “laying up of our *Ideas* in the Repository of Memory,” says Locke, “signifies no more but this, that the Mind has a Power, in many cases, to revive Perceptions” that it has had before. Locke seems to be saying that an instance of remembering is the revival of something that when revived is perceived as an idea, but is not an idea prior to the revival. This is why Locke says that our ideas are “actually no where” before they are revived. Nevertheless, there has to be some more to this story, since it seems not just unsatisfactory but at the very least philosophically suspect to say that the revived idea comes from nothing. I suggest that there are two parts to the story. First, there is the question what is it to perceive an idea “from memory” as opposed to “from sensation” or as the result of another mental operation. Second, there is the question what a memory consists in prior to being revived.

Let’s consider the first question. For Locke, the power of the understanding comprises all the powers associated with perception, or thinking (II.vi.2). Having an idea in the mind just is the exercise of the mind’s power of perception in the form of a particular mental operation, say sensing, reflecting, contemplating, or reviving, as it is directed toward a mental object. In the case of sensation, there is an impression in the senses and a power of the mind to perceive, which results in the perception of an idea of sensation (II.i.24). In the case of having a memory, there is something stored away and a power of the mind to revive (perceive), which results in the perception of an idea from memory. In the previous passage from II.x.2, Locke explains that when the mind exercises the power of revival, the ideas perceived come with “this additional Perception annexed to them, that it has had them before”. The “additional Perception”, Locke confirms, is due to consciousness:

For to remember is to perceive anything with memory, or with a consciousness, that it was known or perceived before: without this, whatever *Idea* comes into the mind is new, and not

remembered: This consciousness of its having been in the mind before, being that, which distinguishes Remembring from all other ways of thinking. (I.iv.20)

These passages seem to indicate that the psychological experience of remembering is due to the consciousness, or awareness, of the way of perceiving when an idea is revived in the mind.

Consciousness provides for my experience of my own thinking – that the idea is the result of a reviving instead of, say, a sensing or a reflecting.

Locke employs consciousness in the same way in his IV.ii.14 account of sensitive knowledge.

When I perceive an idea of an external object by way of sensation I am conscious that it is from sensation. Locke says,

But yet here, I think, we are provided with an Evidence, that puts us past doubting: For I ask any one, whether he be not invincibly conscious to himself of a different Perception, when he looks on the Sun by day, and thinks on it by night... We as plainly find the difference there is between an *Idea* revived in our minds by our own memory, and actually coming into our Minds by our own Senses. (IV.ii.14)

The conscious experience of our ideas allows us to identify their perceptual source. Therefore, consciousness should neither be confused with memory itself (the “repository”) nor with the perceptual act or power of revival.

What is more, defects of memory, for Locke, are defects in thinking, that is, defects involved in reviving ideas, not defects in the consciousness that the ideas revived are *from memory*.

Locke describes the two defects of memory as “ignorance”, when the mind “loses the *Idea*” completely, and “stupidity”, when the mind acts so slowly that the ideas are not revived “when need and occasion calls for them” (II.x.8). In either case, forgetfulness has nothing to do with consciousness. True, to have the experience that I am remembering or that I am the same self now that I was in the past requires that I be able to revive an idea. But this doesn’t mean that

what is involved in reviving an idea should be identified with the consciousness that I have had that idea before.

What, then, is the metaphysical explanation for memory? How can Locke say that ideas are “no thing” prior to being objects of acts of perception? I suggest that Locke says this because we have no conscious experience of whatever precedes an idea. And, although, for Locke, all thinking is conscious, the ideas must be revived before what is in the mind counts as thinking – since thinking just is having ideas. So it is not inconsistent with Locke’s view that there is something prior to thinking that might result in thinking that is not conscious, that is, not yet a mental object of experience.

We get some indication of what this might be for Locke at the end of II.x. Locke argues that not only humans but other animals, too, have this “faculty of laying up, and retaining the *Ideas*, that are brought into the mind” (II.x.10). Not only do sounds “mechanically cause a certain motion of the animal Spirits, in the Brains of those Birds, whilst the tune is actually playing”, which then results in particular motions of the wings or body, but also birds are able to recall a sequence of notes when they are no longer directly affected through their senses or when the imitation is necessary for preservation. What explains memory in the case of other animals?

Locke says,

But which is more, it cannot with any appearance of Reason, be supposed (much less proved) that Birds, without Sense and Memory, can approach their Notes, nearer and nearer by degrees, to a Tune play’d yesterday; which if they have no *Idea* of in their Memory, is now no-where, nor can be a Pattern for them to imitate, or which any repeated Essays can bring them nearer to. Since there is no reason why the sound of a Pipe should leave traces in their Brains, which not at first, but by their after-endeavors, should produce the like Sounds; and why the Sounds they make themselves, should not make traces which they should follow, as well as those of the Pipe, is impossible to conceive. (II.x.10)

Although a little difficult to follow, Locke is suggesting that the ability birds have to repeat a pattern or imitate a tune is due to the ability to record what they have experienced as traces in the brain, which “by their after-endeavors” result in a later production of those sounds.<sup>51</sup> It is quite plausible to think that the case is similar in human beings. What is laid up or stored away “in memory” are traces in the brain that with an effort of the mind to revive become ideas perceived in the mind.<sup>52</sup>

We see a similar account of a relation between the mind and the body in the perception of ideas in Locke’s II.i.23 account of the “origin” of ideas. Having ideas in the mind, Locke says, is “coeval with *Sensation*;

which is such an Impression or Motion, made in some part of the Body, as produces some perception in the Understanding. ‘Tis about these Impressions made on our Senses by outward Objects, that the Mind seems first to employ it self in such Operations as we call *Perception, Remembring, Consideration, Reasoning*, etc. (II.i.23)

Just as there is an impression in the sense organs before there can be a perception of an idea of sensation, there is some sort of imprinting or impression made in the brain before there can be a revival of an idea from memory. The remembered idea is the result of the mind turning toward what is retained in the perceptual act of reviving.

That the precursors of revived ideas are likely located in the brain (or perhaps in the “animal spirits”) is acknowledged by Locke in his admission that brain deterioration or other diseases of the body can result in the erasure of memory:

How much the Constitution of our Bodies, and the make of our animal Spirits, are concerned in this; and whether the Temper of the Brain makes this difference, that in some it retains the

---

<sup>51</sup> Locke thinks that “*Perception* ... is, in some degree, *in all sorts of Animals*” (II.ix.12) and that the ability (or relative inability) of any creature to have ideas depends of the “fewer and duller the Impressions” and “the duller the Faculties” (II.ix.15).

<sup>52</sup> For the view that memory consists in causal traces in the brain, see Don Garrett, “Locke on Personal Identity, Consciousness, and ‘Fatal Errors’”, *Philosophical Topics*, 31:1 & 2 (2003), pp. 107, 116. I am indebted to this article as well as to a discussion with Prof. Garrett at the Eastern Division Meeting of the APA in 2008.

Characters drawn on it like Marble, in others like Free-stone, and in others little better than Sand, I shall not here enquire, though it may seem probable, that the Constitution of the Body does sometimes influence Memory; since we oftentimes find a Disease quite strip the Mind of all it *Ideas*, and the flames of a Fever, in a few days, calcine all those Images to dust and confusion, which seem'd to be as lasting, as if graved in Marble. (II.x.5)

Other passages confirm that what is retained in memory can be explained by traces or “impressions” that are made in the brain. In II.i.15, in his response to the Cartesian claim that the soul always thinks but while sleeping does not retain it, Locke considers how something like that could happen:

Perhaps it will be said, that in a waking Man, the materials of the Body are employ'd, and made use of, in thinking; and that the memory of Thoughts, is retained by the impressions that are made on the Brain, and the traces there left after such thinking; but that in the *thinking of the Soul*, which is not perceived *in a sleeping Man*, there the Soul thinks apart, and *making no use of the Organs of the Body*, leaves *no impressions on it, and consequently no memory* of such Thoughts. (II.i.15)

Although Locke responds to this suggestion by saying the if the soul can “receive, and contemplate, without the help of the Body”, then surely “it can retain without the help of the Body too,” his point is that regardless of whether thinking is material or immaterial or some combination of both, it can't be that whatever does the thinking would have the capacity for thinking and not also for retaining what it thinks. So it can't be that dreamless sleep is due to a separation of what thinks from what retains the thinking, in this case the soul from the body. Rather it has to do with the manifestation and interaction of the powers of whatever constitutes the actual metaphysical mind/body/consciousness composite.<sup>53</sup> Finally, there is one other

---

<sup>53</sup> There is an apparent tension here with Locke's example of the Prince and the Cobbler. If memories are stored in the body, then how can the “soul of the Prince, carrying with it the consciousness of the Prince's past Life, enter and inform the body of the Cobbler” (II.xxvii.15)? It would seem that there could be no memory transferred. I suggest we need not read the passage that way, for having “Princely Thoughts” can be only a matter of assigning the Prince's experience of himself to whatever thoughts he currently has. This is consistent with an ‘appropriation theory’ interpretation. Moreover, we could read Locke to be saying that for the Prince (in the Cobbler's body) to be accountable for an action he need only be having an occurrent thought of a particular action, which can be a transferred occurrent idea that does not need to be revived. Finally, were we to insist that the collection of memories

somewhat unrelated piece of textual evidence that memory consists in traces in the brain. Locke attributes hallucinations, which on his model of the mind can only be ideas retrieved from memory, to the “Reveries of a crazy Brain” (IV.iv.2).

Although Locke never says explicitly that memories are retained as traces in the brain, it is consistent with his agnosticism with respect to the nature of thinking substance and with respect to the metaphysics of causal relations not to delve into any substantive explanation. On the very first page of the *Essay*, Locke declines to “meddle” in such philosophical issues:

I shall not at present meddle with any Physical Consideration of the Mind; or trouble myself to examine, wherein its Essence consists, or by what Motions of our Spirits, or Alterations of our Bodies, we come to have any Sensation by our Organs, or any *Ideas* in our Understandings; and whether those *Ideas* do in their Formation, any, or all of them, depend on Matter, or no. (I.i.2)

So, Locke’s failure to go into any detail as to what memory consists in is due to his insistent neutrality on the nature of ideas, on the relation between real constitutions and powers, between the mind and the body, between an act of thinking and our consciousness of it, or even between an objectively existing continuing consciousness and the conscious state of itself as thinking. Moreover, that Locke proclaimed his agnosticism leaves little room for speculation. What is clear, however, is that Locke is distinguishing the causal precursors to revived ideas from those ideas as revived. And if we must be conscious of ourselves as perceiving revived ideas, because all thinking is conscious, yet the causal precursors to those ideas are “no thing” *to us*, then Locke must be thinking that consciousness is distinct yet somehow united to and working with memory.

There is one final worry. If being able to attribute a past action to our selves is necessary in order not to be “created miserable”, then it seems that on my interpretation Locke is committed

---

themselves are what is transferred and so what is necessary for attributing identity to oneself, then we have made Locke, once again, guilty of circularity.

to bodily resurrection, and even perhaps to having the same body resurrected. Locke, however, seems to explicitly deny that necessarily I appear on the “Great Day” with the same body:

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 deserve that Punishment for them. (II.xxvii.26).

Here Locke states that I can appear for just punishment in any body whatsoever; it doesn't have to be the exact body to which I was united throughout my life. But this need not be seen as in conflict with the view that stored memory consists in causal traces in the brain for two reasons. First, as far as we can know, God can restore to any particular body those causal traces corresponding to my particular experience. The reason is that God has something objective to look to in determining what those causal traces should be. Therefore, if Locke's view is that I must revive and experience as mine the appropriate idea of a past action, then there doesn't seem to be a problem. Second, if Locke's view is that I do not have to do the reviving myself, then it seems that we don't need the causal traces at all. God could just as easily restore in any moment any particular ideas I need to be conscious of in order to satisfy the criterion that I am not punished for what I have no awareness of having done.

## 5. Conclusion

Traditionally, Locke is considered to have a psychological account of personal identity that relies on consciousness as the criterion. The criticisms it has faced are numerous. If the psychological theory is taken also as a metaphysical theory, then there is a problem of circularity. If Locke is interpreted to have an appropriation theory, in which the first personal attribution to ourselves of occurrent mental states is sufficient for personal identity, then the theory is ill-equipped to meet the demands of Locke's commitment to divine rectification. There



must be something objective for God to look to in determining those actions for which we can be justly punished or rewarded. Moreover, Locke is adamant that the objective metaphysical criterion is not substance. I have argued that there is textual evidence that Locke's theory includes a metaphysical criterion, namely an enduring consciousness.

Locke can be seen to use 'consciousness' ambiguously to mean both a state of awareness of myself as thinking *and* as the ongoing self I am aware of in that conscious state. The latter sense of consciousness allows me to know and have an experience of myself as the same self continuing through time, where the former sense is the objective criterion known perhaps only to God. Where the metaphysical fact of my ongoing self is revealed to me as a phenomenological fact due to the nature of particular perceptual situations, the metaphysical fact is beyond my ken. Thus, Locke can be seen to have a coherent theory that incorporates both psychological and metaphysical aspects. Although it goes beyond anything Locke says explicitly, there is reason to think that Locke is concerned to abandon the traditional substance/mode ontology he has inherited in order to embrace a relatively agnostic view that is consistent with some version of explanation by virtue of real constitutions and qualities or powers. Consistent with this view, I have argued that Locke sees persons as mind/body/consciousness composites in which all three components are distinct configurations of real constitutions that are somehow united.

Finally, I have argued that Locke's metaphysical account of memory as consisting in causal traces in the brain helps to support the interpretation that we are mind/body/consciousness composites, and that although occurrent mental states are necessary to experience and identify myself as the same self, it is the ongoing consciousness that does the work from a third personal point of view.

