

PHIL 260: Persons & Values
 Winter 2005; David O. Brink
 Handout #2: Classical Debates about Personal Identity

THE LOCKEAN THEORY

John Locke (1632-1704) presents his account of personal identity, as part of a general analysis of identity and diversity, in his Essay Concerning Human Understanding (originally published 1690), book II, Chapter xxvii. Locke begins with a conception of persons as self-conscious. A person is

... a thinking intelligent Being, that has reason and reflection, and can consider itself as itself, the same thinking thing in different times and places; which it only does by the consciousness, which is inseparable from thinking ... [Essay II, xxvii, §9].

Locke distinguishes the identity of persons, material substances (bodies), and immaterial substances (souls) (esp. §§10-15, 23), and claims that identity of persons consists in sameness of consciousness.

... in this alone consists personal Identity, i.e. the sameness of rational Being: And as far as this consciousness can be extended backwards to any past Action or Thought, so far reaches the Identity of that Person [§9].

THE FORENSIC ROLE OF PERSONS

As we have seen, Locke appeals to the “forensic” or normative significance of persons (§§8, 15, 17-21, 23, 26). Here, he seems to have in mind claims about the role of both persons (personhood) and personal identity in our backward-looking and forward-looking attitudes. He makes at least four claims.

- Only persons are properly praised and blamed, because only persons can be held accountable for their actions. Only persons have the power of **will**, as distinct from mere desire. What sets the will apart from desire is that the former is influenced by reflection on the consequences of different desires (II, xxi, 30-35). Joseph Butler (1692-1752) and Thomas Reid (1710-96) develop Locke’s ideas about the role of critical reflection in agency and personality further.¹ Both think that it is persons that are responsible in morality and law, because persons, unlike non-rational animals, have the capacity to distinguish between the existence and strength of desires, on the one hand, and their authority, on the other hand. Persons are able to deliberate about the suitability of their desires and regulate their intentions and actions in accordance with these deliberations.
- Person is a normative category; human is a biological category. Even if most normal adult humans are persons, the two categories are neither co-intensive or co-extensive. Non-rational humans are not persons, and rational non-human animals (e.g. rational parrots) would be persons (§8).
- Responsibility presupposes personal identity. P2 can be held responsible for P1’s actions only if P2 = P1.
- Personal identity justifies special concern. It is appropriate for P1 to have special concern for P2 iff P2 = P1.

CONSCIOUSNESS AND MEMORY

For Locke, consciousness involves the present self’s awareness of a past self’s awareness of its actions and experiences as its own. This is why his consciousness criterion is a memory criterion (cf. §§9, 25).

¹Butler, Sermons ii 13 and Reid, Essays on the Intellectual Powers of Man iii 4 and Essays on the Active Powers of Man ii 2.

We have many different kinds of memories. There is **impersonal factual** memory, as when one remembers that $2+2=4$. There is also **ability** memory, as when one remembers how to tie shoes, write in cursive, or go “top shelf” on a hockey goalie. And there is **first-person** memory in which one remembers the events that essentially involve oneself, as when one remembers one's own actions and experiences as one's own. It is this last kind of memory that is involved in consciousness and so constitutes personal identity, on Locke's view. So Locke's view seems to be:

- L1: P2 is the same person as P1 iff she remembers all of the experiences and actions of P1.

But L1 is absurdly over-demanding, as Locke seems to recognize (§§10, 20). His considered view is closer to:

- L2: P2 is the same person as P1 iff she can remember all of the experiences and actions of P1.

However, L2 is still too stringent. Locke may require something weaker still (cf. §§23, 26).

- L3: P2 is the same person as P1 iff she can remember most of the experiences and actions of P1.

Or even:

- L4: P2 is the same person as P1 iff she can remember some sufficient number of the experiences and actions of P1.

LOCKE'S JUSTIFICATION OF THE CONSCIOUSNESS CRITERION

Locke offers two main reasons for claiming that personal identity consists in consciousness: one metaphysical, one normative.

First, he contrasts organisms and material or immaterial substance. Organisms (e.g. animals) should not be identified with material or immaterial substances. The persistence of an organism is a matter of identity or **continuity of life**. This continuity can and does unite different material substances; the composition of the organism can and typically does change over time (§§3-8, 10). Thus, the relationship of organism to material substance is one-many. But Locke thinks that sameness of person stands to sameness of organism (e.g. man) as sameness of organism (e.g. man) stands to sameness of material substance (e.g. atoms). So the relationship between person and man is, at least potentially, one-many; indeed, he thinks that it is also potentially many-one. A particular person might span different bodies or men, as would be true if the consciousness of a prince was transferred to the body of a cobbler (§15; cf. §§16, 19). And several persons might inhabit (successively) the same body (§§20, 22), as in the case of Day Man and Night Man (§23).

Locke's second reason appeals to the normative significance of persons. Locke thinks that both forward-looking attitudes, such as special concern, and backward-looking attitudes, such as responsibility, pride, and shame, track consciousness, rather than sameness of man, sameness of material substance, or sameness of immaterial substance. He claims, for example, that one would feel special concern for Nestor's fate or assume responsibility for Nestor's actions if and only if one had consciousness of Nestor's experiences and deeds, and not simply if one shared material or immaterial substance with Nestor (§14).

1. Sameness of person justifies special concern.
2. Sameness of material substance (e.g. atoms), as such, does not justify special concern.
3. Sameness of immaterial substance (e.g. immaterial stuff), as such, does not justify special concern.
4. Hence, personal identity \neq sameness of material substance or sameness of immaterial substance.
5. Sameness of consciousness does justify special concern.
6. Hence, personal identity = sameness of consciousness.

THE VIRTUES OF THE CONSCIOUSNESS CRITERION

Locke thinks his account of personal identity has several advantages. First, it is in certain ways metaphysically neutral or agnostic. Locke thinks that his view is compatible with materialism but also with the existence of immaterial substances. Moreover, it is compatible with Christian doctrines of the resurrection and immortality (§§7, 15, 23-27).

Second, the memory criterion seems to provide a metaphysical criterion that explains the reliability of our epistemic criteria; memory is the principal way that we are aware of ourselves persisting through time.

Third, Locke thinks that appeal to consciousness better explains the forensic role and importance of persons than does appeal to material or immaterial bodies (§§14, 18, 26).

REID AND THE NEED FOR A TRANSITIVE CRITERION

Reid defends what appears to be a form of non-reductionism about personal identity in his Essays on the Intellectual Powers of Man (originally published 1785) in part by criticizing Locke's memory criterion. Among other worries, Reid objects that the memory criterion is not transitive, as the identity relation is.

Suppose a brave officer to have been flogged when a boy at school for robbing an orchard, to have taken a standard from the enemy in his first campaign, and to have been made a general in advanced life; suppose also ... that when he took the standard he was conscious of having been flogged at school and that when made a general he was conscious of his taking the standard, but had absolutely lost the consciousness of his flogging [357/114].

The structure of his argument is a reductio ad absurdum.

1. Locke's memory criterion (L3/L4).
2. P20 can remember most of P5's experiences; P50 can remember most of P20's experiences; but it's not the case that P50 can remember most of P5's experiences.
3. Hence $P5 = P20$. [by (1) and (2)]
4. Hence $P50 = P20$. [by (1) and (2)]
5. Hence $P5 = P50$. [by (3), (4), and the transitivity of identity]
6. Hence $P5 \neq P50$ [by (1) and (2)]
7. Hence, $P5 = P50$ and $P5 \neq P50$ [by (5) and (6)]
8. Hence $\neg L3/L4$.

The argument for (7) is valid. If (2) is true, (1) must go. A criterion of personal identity must be transitive, as Locke's criterion is not. However, this defect can be remedied. Locke's criterion is memory **connectedness**.

Any persons P_x and P_y are memory connected iff P_y can remember most of the experiences and deeds of P_x (or vice versa).

We can now define memory **continuity** in terms of memory connectedness.

Any persons Px and Py are memory continuous iff they are elements in a series of persons such that temporally contiguous pairs of persons are memory connected.

We can then revise L3/L4 as follows:

L5: P1 and P2 are identical iff P2 is memory continuous with P1.

A MORE INCLUSIVE PSYCHOLOGICAL CRITERION

But memory connectedness and even continuity might fail, yet we might want to recognize persistence so long as there are other forms of psychological continuity, such as retention of impersonal factual memory and retention of skills. For this reason, we might seek a form of psychological continuity that does not focus exclusively on memory. Moreover, if we are impressed by Locke's appeal to the forensic role of persons, we might be led to focus on the sort of psychological connections involved in deliberation, forming plans based on deliberation, and performing actions that implement such plans. Therefore, to explain persistence, we might appeal to other backward-looking and forward-looking psychological connections and appeal to the **deliberative** control of intentional states (e.g. beliefs, desires, intentions, and actions).

Any persons Px and Py are psychologically connected iff the mental states and actions of Px are counterfactually dependent in the appropriate way on the mental states and actions of Py (or vice versa).

We can then define psychological continuity in terms of psychological connectedness.

Any Persons Px and Py are psychologically continuous iff Px and Py are elements in a series of persons such that (e.g. temporally) contiguous pairs of persons are psychologically connected.

Psychological reductionism could then be formulated as follows:

Psych-R1: P1 and P2 are identical iff they are (sufficiently) psychologically continuous.

REID ON SELF-PREDICATION

Reid also thinks that the Lockean reductionist analyzes personal identity in terms of mental states, and this, he thinks, confuses subject and predicate.

My personal identity, therefore implies the continued existence of that indivisible thing which I call myself. Whatever this self may be, it is something which thinks, and deliberates, and resolves, and acts, and suffers. I am not thought, I am not action, I am not feeling; I am something that thinks, acts, and suffers. My thoughts actions and feelings change every moment; they have no continued, but a successive, existence; but that self, or I, to which they belong, is permanent, and has the same relation to all succeeding thoughts, actions, and feelings which I call mine [341/109].

I have thoughts and experiences; hence, I am neither thoughts nor experiences.

How compelling is this argument? Consider three different non-reductionist claims Reid might be making.

- a) I am not a mental state.
- b) I am not a set or series of mental states.
- c) I am not an ordered set or series of mental states.

(a) seems clear, but no reductionist denies it. Some reductionists, such as Parfit, who are reductionists about persons, as well as about personal identity, deny (b), or at least (c) (RP: 21-12). But it may be less clear that neither (b) nor (c) is defensible. (c), especially, seems not obviously mistaken. So, if a reductionist embraces (c), this may not be a problem. But, moreover, reductionism about personal identity does not require reductionism about persons. Shoemaker does not endorse this further reductionist claim (endorsing an adverbial analysis of mental states as essentially modifications of persons, similar to the way in which dents are essentially concave surfaces in larger structures), and our reductionist criteria (L1-Psych-R1) have not.

"STRICT" AND "LOOSE" KINDS OF IDENTITY

Like Reid, Bishop Butler defends a form of non-reductionism against Lockean reductionism. Butler makes his case in his "Dissertation on Personal Identity" (originally published, 1736). Butler and Reid both complain that Locke's view implies that there is personal identity through time in only a "loose and popular," not in the "strict and philosophical" sense (B: 263-5/100-103; R: 112, 116-7). Butler writes

... from hence it must follow, that it is a fallacy upon ourselves, to charge our present selves with anything we did, or to imagine our present selves interested in anything which befell us yesterday, or that our present self will be interested in what will befall us tomorrow; since our present self is not, in reality, the same with the self of yesterday, but another like self or person coming in its room, and mistaken for it; to which another self will succeed tomorrow [267/102].

Butler thinks Locke's account of the identity of organisms over time (e.g. the identity of an acorn and an oak) is not strict identity.

And if they have not one common particle of matter, they cannot be the same tree, in the proper philosophic sense of the word same; it being evidently a contradiction in terms, to say they are, when no part of their substance and no one of their properties is the same ... [265/101].

Butler thinks the same problem afflicts Locke's analysis of personal identity through time.

But they cannot, consistently with themselves, mean, that the person is really the same. For it is self-evident that the personality cannot really be the same, if, as they expressly assert, that in which it consists is not the same [267/103].

It's not always clear whether Butler and Reid have in mind compositional or qualitative diversity, though their clearest statements of the objection seem to appeal only to compositional diversity. As Reid writes

Our consciousness, our memory, and every operation of the mind, are still flowing like the water of a river, or like time itself. The consciousness I have of this moment can no more be the same consciousness I had last moment, than this moment can be the last moment. Identity can only be affirmed of things which have a continued existence. Consciousness, and every kind of thought, are transient and momentary, and have no continued existence; and, therefore, if personal identity consisted in consciousness, it would certainly follow, that no man is the same person at any two moments of his life; and as the right and justice of reward and punishment are founded on personal identity, no man could be responsible for his actions [360/116-17]

Later, Reid makes the same argument.

If our personal identity consists in consciousness, as this consciousness cannot be the same individually at any two moments, but only of the same kind, it would follow that we are not for any two moments the same individual persons, but the same kind of persons [361/117].

Reid's allusion to the river is a cue to the Heraclitean pedigree of this argument, which seems to appeal to compositional diversity over time.

1. P1 = consciousness-1.
2. P2 = consciousness-2.
3. consciousness-1 ...consciousness-2.
4. Hence P1 ...P2.

But the reductionist thinks that persons are series of continuous mental (e.g. conscious) states. We can read "P1" and "P2" in (1) and (2) as picking out (a) persons or (b) person slices (temporal parts of persons). If (a), then the reductionist clearly denies (1) and (2); persons are to be identified with series of continuous mental states, not some small subset of these. If (b), then, of course, (4) follows. But this just shows that distinct temporal parts of persons are not the same; it does not show that the person at t1 and the person at t2 are different.

BUTLER AND THE CIRCULARITY OBJECTION

Butler thinks that Locke makes a "wonderful mistake" (264/100). Butler seems to believe that memory already presupposes personal identity. My awareness of a past experience e is a memory iff:

1. I believe that I had experience e.
2. I had experience e.
3. My current belief is counterfactually dependent in the appropriate way on that past experience.

Condition (2) shows that memory presupposes personal identity. But then memory cannot be used to provide a reductive analysis of personal identity. The memory criterion is circular.

The standard solution to this problem (due to Shoemaker) is to introduce a notion that is otherwise like memory but does not presuppose personal identity. Shoemaker calls this quasi-memory. My awareness of a past experience e is a quasi-memory iff:

1. I believe that I had experience e.
2. Someone had experience e.
3. My current belief is counterfactually dependent in the appropriate way on that past experience.

Our memory criterion (L5) would have to be modified to avoid circularity, as in

L6: P1 and P2 are identical iff P2 is q-memory continuous with P1.

However, our more inclusive psychological continuity criterion (Psych-R1) does not require modification, for it requires only the appropriate sort of counterfactual dependence among the mental states and actions of P1 and P2; it does not assume that P1 and P2 are the same.

Q-MEMORY, CONTINUITY, AND FISSION

But because q-memory and continuity do not presuppose personal identity, they can obtain non-identical people. Defining personal identity in terms of memory may be circular. But defining it in terms of q-memory or continuity may not seem to ensure personal identity.

There is another worry here, viz. that q-memory and continuity can take branching forms, in which more than one person in the future is continuous with me now. As Reid observes

This [Locke's] doctrine has some strange consequences, which the author was aware of. Such as, that if the same consciousness can be transferred from one intelligent being to another, which he

thinks we cannot show to be impossible, *then two or twenty intelligent beings may be the same person* [357/114].

Parfit discusses such a case in **Venetian Memories** (RP: 220-21) in which someone else has a few of my q-memories. Does this cause trouble for L6 or Psych-R1? Consider this reductio. Suppose we call Tom's later self Dick and the stranger with a few q-memories of Tom's earlier self Harry.

1. L6.
2. Dick q-remembers things Tom did.
3. Harry q-remembers things Tom did.
4. Hence, Dick = Tom.
5. Hence, Harry = Tom.
6. Hence, Dick = Harry.
7. But Dick \neq Harry.
8. Hence, \neg L6.

A solution to these two problems is for the Lockean to recognize that continuity is scalar and have identity track maximal continuity. Note that Dick and Harry are not equally continuous with Tom. Harry just has a few q-memories of Tom, whereas Dick is systematically continuous with Tom. Psych-R1 appeals to sufficient psychological continuity. Harry seems to be insufficiently continuous with Tom. We could get the same result, I think, by appealing to a version of the **closest continuer theory** (cf. Nozick).

Psych-R2: P2 is identical with P1 iff P2 is the most psychologically continuous with P1 of those persons whose continuity rises above a certain threshold.

Something like Psych-R2 promises to deal with cases of **uneven branching**. Of course, there is no guarantee that all branching will be uneven. In principle, we could imagine a case of perfectly even or symmetrical branching. One such case is the case of **fission**. Fission raises interesting and difficult problems, which we will discuss later. For now, let's just note that the reductionist can and typically does deal with such cases by adding a **nonbranching** clause.

Psych-R3: P2 is identical with P1 iff P2 is a (sufficiently) closest continuer of P1 and there is no other continuer of P1 that is as continuous with P1.

We will return to the significance of this nonbranching clause later.

BUTLER AND HUME ON SELF-CONSCIOUSNESS

Butler thinks he is introspectively aware of his own self-consciousness. He might mean (i) awareness of a persistent **subject** of experiences (when I compare any two distinct awarenesses of myself, I am immediately aware that these awarenesses are had by the same self, viz. me) or (ii) awareness of a persistent **object** of experiences (when I compare any two distinct awarenesses of myself, I am immediately aware that these selves are the same).

But mightn't two self-experiences, which I now compare, be had by different subjects or concern different selves, as in Branchline Cases or Venetian Memories?

In A Treatise of Human Nature (originally published 1739) David Hume (1711-76) gives an account of personal identity that admits of both skeptical and reductionist interpretations. Hume's skepticism focuses, in part, on Butler's claims to be immediately aware of his own identity. Hume denies that we have any direct awareness of a persisting subject of experience (Treatise 251-52).

If any impression gives rise to the idea of self, that impression must continue invariably the same, thro' the whole course of our lives; since self is suppos'd to exist after that manner. But there is no impression constant and invariable. Pain and pleasure, grief and joy, passions and sensations succeed each other, and never all exist at the same time. It cannot, therefore, be from any of these impressions, or from any other, that the idea of self is deriv'd; and consequently there is no such idea [T 251-52].

He continues:

For my part, when I enter most intimately into what I call *myself*, I always stumble on some particular perception or other, of heat or cold, light or shade, love or hatred, pain or pleasure. I never can catch *myself* at any time without a perception, and never can observe anything but the perception. ... If any one upon serious and unprejudic'd reflexion, thinks he has a different notion of *himself*, I must confess I can reason no longer with him [T 252].

Hume claims that the idea of identity requires (1) the "idea of an object that remains invariable and uninterrupted through a supposed variation of time" (253).

1. A-A-A-A-A

By contrast, the idea of diversity involves either (2) the idea of "several different objects existing in succession, and connected together by a close relation" (253)

2. A A A A A

or (3) the idea of a "related succession as variable and interrupted" (254).

3. A A B A A

We never have any direct evidence for (1). The closest we can come is (2), but (2) involves diversity not identity. In fact, Hume thinks, we tend to confound (2), and even (3), with (1). This is how we come to create the **fiction** of personal identity.

Though he thinks that there is no such thing as personal identity, in the strict and philosophical sense, or at least that we cannot know that there is, Hume does have an **explanatory** account of how and why ascribe identity through time. And this explanatory account becomes a kind of fallback account of the truth-conditions of ascriptions of identity, once we have rejected the initial unsuccessful accounts. He believes that we think there is a substance that unites all of our various experiences. According to Hume, however, we have no impression of this self; all we ever observe is distinct experiences standing in relations of resemblance, contiguity, and causation to each other. Where these relations are especially strong, we impute all these experiences to a single enduring subject. In this way, (2) and (3) become our truth conditions for (1). Thus, Hume comes to accept something very much like psychological reductionism, but he treats it, unnecessarily, as a fallback position.