

Personal Identity and Morality

[Harold Noonan](#)

Formatted: Font: Not Bold

I

Formatted: Heading 1, Left, Line spacing: single

Formatted: Font: Not Italic

Does the true account of personal identity undermine everyday moral thinking? Do everyday moral practices presuppose a false account of our nature and persistence conditions? For example, it seems a fundamental platitude of our moral thought that only the doer of a deed merits punishment or reward for it (call this “the principle of the moral necessity of agency”). But this principle is challenged in the debate about personal identity from the mid-twentieth century. Can it be retained? Another principle challenged, also seemingly fundamental to our moral thought, is what I shall call “the principle of the intrinsicality of moral status”: no relation any sentient being has to another can deprive it, qua patient, of moral status. I shall consider the three main accounts of personal identity in the contemporary literature: the neo-Lockean psychological continuity account presented paradigmatically by Shoemaker (1963, see also Shoemaker and Swinburne 1984), animalism (as defended by Olson (1997, 2003)) and David Lewis’s perdurance account (1983, 1986). In each case it will be concluded that the account is inconsistent with what we take to be fundamental elements of our moral thought being so.

II

I begin with the contemporary neo-Lockean psychological continuity account of personal identity, the paradigm of which is Sydney Shoemaker's. This, of course, developed from Locke's account, which is intended to be in accordance with the status of "person" as a "forensic term appropriating actions and their merits, [which] belongs only to intelligent agents, capable of a law, and happiness and misery" (Essay II, xxvii.26), whence, according to Locke, "in this personal identity is founded all the right justice of reward and punishment; happiness being that for which everyone is concerned for himself, and not mattering what becomes of any substance not join to or affected by that consciousness" (Essay II, xxvii.18). Nevertheless, as Parfit's famous argument (1971, 1984) reveals, it follows from the neo-Lockean story, at least its paradigmatic form, that personal identity lacks the foundational moral status Locke affirms, but is of merely derivative importance. It is not fundamentally "what matters."

Formatted: Font: Not Italic

Formatted: Font: Not Italic

The basis for the contemporary neo-Lockean account is the argument based on the "transplant case," which makes it all but undeniable that suitably caused psychological continuity can in some circumstances be sufficient for personal identity.

Shoemaker (1963) tells the story. In the late twenty-first century surgeons can transplant brains and brain parts, specifically cerebra. Suppose that the cerebrum of a Mr. Brown is transplanted into the skull of a Mr. Robinson, with consequent transference of psychological traits. The resultant person, Brownson, is a completely healthy person, with Robinson's body, but in character, memories and personality traits quite indistinguishable from Brown, not as a

consequence of a freak accident but because of his possession of Brown's cerebrum. Now who will this person be?

Most modern philosophers, reflecting on this case, have found that they could not honestly deny that Brownson is Brown, and so they have been led to accept, as Parfit puts it (1984, 263), that "receiving a new skull and a new body is just the limiting case of receiving a new heart, new lungs and so on."

The position to which this "transplant intuition" seems to point is that personal identity is constituted by psychological facts. The essence of this is that, given the importance for our attitudes towards persons of their memories, character and personality traits, continuity in respect of these should be taken to constitute personal identity—whether or not caused by the persistence of some bodily organ, such as the brain; and absence of continuity in these respects entails absence of personal identity.

This position is called "neo-Lockean" of course, because the simplest version of it is suggested by Locke's discussions. Reading Locke to mean memory by "consciousness" we arrive at the memory criterion of personal identity.

Contemporary neo-Lockeans revise this in two ways. First, they say that other psychological facts as well as memory must be mentioned in defining personal identity. In general, any causal links between past psychological facts and present psychological facts can be subsumed under the notion of "psychological connectedness." Secondly, they introduce the notion of psychological continuity, to mean the obtaining of overlapping chains of psychological connections, to ensure conformity with the transitivity of identity.

At this point the problem of fission raises its head. Consider the variant of the Brown/Brownson case in which only half of Brown's brain is transplanted with consequent transfer of psychological traits, the other half being destroyed. The neo-Lockean must say this preserves personal identity. But now consider the case—the fission case—in which both hemispheres are transplanted, but into different heads. Both later people cannot be the same as the one original. But there is nothing to choose between them. So we must say that neither is. But then psychological continuity is not a sufficient condition of personal identity.

Most psychological continuity theorists, including Shoemaker (in Shoemaker and Swinburne 1984), respond to this problem by revising the psychological continuity account to make psychological continuity sufficient for personal identity only absent an equally good rival candidate. This is to reject the principle, the only x and y principle (Noonan 2019), implicit in a seminal argument given by Bernard Williams (1956–7, see also his 2014: Chapter 39), that, to put it roughly, whether an individual x is identical with an earlier individual y can depend only on facts about them and the relations between them; no facts about any individuals other than x or y can be relevant (for less rough formulations see Noonan 2019).

To see the intuitive objection to the revised continuity account, consider the fission case again. Suppose I am told that my brain is to be divided into two and then the two halves transplanted into different heads. Then, according to the revised psychological continuity account, I know that I will cease to exist, and two new people will come into being. However, I know that if someone destroys the right brain hemisphere before it is transplanted, thereby eliminating the plurality of candidates, I will continue to exist and be the recipient of the left-brain hemisphere. Thus, according to the revised psychological continuity criterion, in this case my continued existence is logically dependent upon the non-existence of someone, the person

resulting from the right brain hemisphere transplant, who would not be me even if he were to exist. But how can my continued existence thus be logically dependent on the non-existence of someone else?

Reflecting on the fission case with these considerations in mind leads Parfit to his remarkable thesis that identity is not what matters in survival. Thinking about this will lead us on to the question whether the neo-Lockean psychological continuity account accords with our everyday moral thinking.

To get clear before turning to the argument for it we must first see what Parfit's thesis means.

What it means is that, contrary to what we are all inclined to think, we do not have a non-derivative concern for our own future existence. What is of fundamental importance to us is that there be in the future people related by psychological continuity to ourselves as we are now—call these our Parfitian survivors. My having a Parfitian survivor tomorrow guarantees, as things are, that I will exist tomorrow, but it does not entail it, and in conceivable circumstances, like fission, the two come apart. Parfit's thesis is that given our fundamental desires we would have no reason, in such a conceivable circumstance, to prefer a situation in which we continued to exist to one in which we had merely a multiplicity of (equally comfortably off) Parfitian survivors.

Intuitively this is very implausible. We can imagine a society in which *Star Trek* variety teletransportation is in general use as a means of transportation of inanimate objects and food animals—even though it is acknowledged that it is not transportation at all but, in reality, the destruction of one object and the creation elsewhere of a mere replica. But if we try to imagine

that the people in this society, *whilst continuing to acknowledge that this is what the process really involves*, nonetheless allow themselves and their loved ones to be teletransported and regard teletransportation as a convenient alternative to travel, we run into immediate difficulties; at first sight it seems as if we have succeeded only imagining a society of madmen.

But according to Parfit we have not. These people are acting as it would be rational for us to act, *given our actual non-derivative desires*, if we lived in their society and shared their beliefs in the destructive nature of the teletransportation process. For our fundamental desires do *not* include a desire for our own continued existence but merely one for the future existence of a Parfitian survivor.

Parfit's argument for this thesis arises from reflection on the fission case. And to understand it we need to make a distinction between two types of opinion reflection on such puzzle cases about personal identity generates. First, there are opinions about how the language of personal identity is to be applied to the case, what the true statements about personal identity to be made about it are. These opinions reflect our mastery of our language and particularly those parts expressive of the concept of personal identity. In short, they reflect our semantic intuitions. They are akin to our opinions about identity in puzzle cases about non-persons, such as, for example, the case of the Ship of Theseus (Noonan 2019). But thinking about puzzle cases about personal identity generates opinions of a second sort. These are opinions about how it is rational for the people whose identity is at issue in the case to behave, given the beliefs they are described as holding. These opinions reflect our fundamental desires. For we arrive at them by imagining *ourselves* to be involved in the case and asking how *we* should then rationally behave.

Parfit's argument for his thesis can be explained as follows. First, he describes a fission case. Next, he argues that (a) the correct description of the case, the one which accords with our

semantic intuitions, is that the original person ceases to exist but would not have done so if only one of the fission products had existed, and that the fission products are new existents (in accordance with the revised psychological continuity criterion of personal identity), but that (b) the correct opinion about what it is rational to do in such a case, the one in accord with our fundamental desires, is that it would be quite irrational, if you were the original in the case, to be concerned about your impending fission as you would be about an impending heart attack, or to think that you could gain anything by preventing the fission by ensuring only one hemisphere was transplanted (if you were given the choice), even though it is correct there were you to do so you *would* ensure your own future existence. (See also Shoemaker's exposition in Shoemaker and Swinburne 1984_119) where this choice situation is emphasised). The conclusion is then that this *combination* of opinions can be explained only by accepting that the fundamental desires we have are not the ones we think we have and do not include a desire for our own future existence; that desire, which we do have, is a merely derivative one, and what it is derivative from is a desire for a future Parfitian survivor, which as a matter of fact, as things are, we can ensure only by ensuring our own continued existence.

The argument can be resisted by denying that the description of the fission case given in (a) above, the description implied by the revised psychological continuity criterion, is correct. An alternative is to say that no one ceases to exist when the fission takes place. Rather, two people who have been spatially coincident now become spatially distinct. But if no one ceases to exist when the fission takes place of course it must be absurd to view it as death. This multiple occupancy view is the straightforward consequence of accepting Bernard Williams's only x and y principle. Given the intuitiveness of this principle it cannot be regarded as an absurd view, but it is a minority one.

Another way of resisting the argument is to deny that we do have the type (b) opinion about fission. Few have taken this line; one who does is Jerome Schaffer who writes:

Psychological continuity is important where there is identity, but not otherwise. Returning to our case of the man who splits, we would say that since identity is not preserved even though psychological continuity is preserved, the man should feel quite differently about it from the way he should feel about single transplantation. (1977, 157)

However, the difficulty with this is just that it is just so plausible that we have the (b)-type opinion.

But it is obvious that if we accept Parfit's conclusion we must abandon the principle of the moral necessity of agency, that only the doer of a deed merits punishment or reward for it, as a fundamental element of our everyday moral thinking.

Of course, as things are, fission does not happen. But imagine a society in which it does—and where it is frequently occurring and can be voluntarily chosen. And suppose that in the society people still have only the fundamental desires which according to Parfit we actually have, that is, they do not have a non-derivative desire for their continued existence. Then a legal system and societal norms in accordance with the principle of the moral necessity of agency would be wholly inappropriate.

Locke writes, in support of his claim that 'person' is a forensic term: 'in this personal identity is founded all the right justice of reward and punishment; happiness and misery being

that for which everyone is concerned for himself' (Essay II, xxvii.18). Here what follows the semicolon is intended to give a reason for what precedes. But if Parfit is right, what follows the semicolon is false. Hence in the society imagined, in which, perhaps merely because of the advanced state of medical technology, fission is frequent, reward and punishment founded on personal identity would be purposeless. For people would not be much motivated to act well by the prospect of reward given only to the doers of good since fission would too often supervene before any reward was possible; nor would they be motivated to refrain from wrong-doing by the prospect of future punishment meted out only to evil-doers since they could easily avoid it by choosing fission. Thus, if Parfit is right, the principle of the moral necessity of agency is not a fundamental part of our moral thinking but merely a rule of thumb the utility of which depends very largely on the present limited state of medical technology.

Formatted: Font: Not Italic

III

The same conclusion can be derived if animalism is accepted. The reasoning is parallel.

Animalism is the claim that we, you and I, and any other readers of this book at least, are animals of a certain kind, Human Beings. So we have the persistence conditions of human beings; and these involve *no* sort of psychological continuity whatsoever. Psychological continuity is irrelevant to personal identity because it is irrelevant to the persistence of animal identity generally. Hence for each of us there was a time when he was a fetus, but not a person, possessed of no psychology whatsoever, and again for some of us there will be times in the future, after brain damage or decay when we will still exist, but merely as human beings, not

persons, wholly devoid of any psychological life. In addition, in the cerebrum transplant case, the person does not go with the cerebrum since no animal does; he remains, devoid of a cerebrum and entirely without psychology. However, if the entire brain, including the brain stem, the biological control centre of the human animal, is transplanted, the person [that is, the animal], who is the original owner of the brain, goes with it.

This animalist position is, of course, wholly opposed to the neo-Lockean account, and the strongest argument against it is the main argument for the latter, the transplant argument. Animalists must therefore respond to it.

Eric Olson does so by an appeal to Parfit's thesis that identity does not matter. The recipient of the cerebrum transplant, he says, does have what matters in the survival of the donor. Therefore, the donor has the same reason to care about the welfare of the recipient as you have to care about your own welfare. It is the recipient who should be held responsible for the donor's actions, and, rather than the surviving donor (if provided with cerebrum replacement), who should be regarded for practical reasons as the same person. These facts cause us to believe, mistakenly, that the recipient is the donor. We recognize that he is the Parfitian survivor, and so we mistakenly believe he has what matters for identity. Olson suggests that we use "the same person" in ordinary speech in a "practical" sense in which it expresses not numerical identity but just the holding of those relations of psychological continuity and connectedness which according to Parfit matter in survival. His "bold conjecture" is that (in a cerebral transplant version of Locke's cobbler-prince example): "The fact that Brainy is the same person after the operation as Prince was before it in this practical sense of 'same person' is the main source of the transplant intuition" (1997, 69).

Of course, this explanation depends on Parfit's thesis being correct. But as we have seen it is controversial; additionally, it is not clear that an animalist can endorse it.

Parfit's argument rests on the claim that our opinions about the fission are in accord with a 'best candidate' account, so that whether I continue to exist depends on whether there is anyone who is *uniquely* best qualified to count as me at the later time (as will be the case if the fission is botched but not if it is successful). This conflicts with Williams's only x and y principle.

But it is not at all obvious that an animalist should endorse the uniqueness requirement on a sufficient condition of personal identity. Olson himself writes "no one accepts the Uniqueness Requirement because it sounds right. The transplant intuition has led us into a quandary and the Uniqueness Requirement is seen as the best way out; it is a theoretical necessity" (Olson 1997, 49). However, if the animalist is to endorse Parfit's argument for his thesis, so that he can call on it to explain the seeming evidence of the transplant intuition, he must accept the uniqueness requirement. Thus, the transplant intuition seems to remain a very strong argument against animalism.

But there are also strong arguments for animalism, in particular, the thinking animal argument, also called the "too many thinkers" argument, which points out that since it is undeniable that human and other animals are thinkers it appears to follow that, according to the neo-Lockean position, there are two thinkers where I am, the person and the thinking animal, and that there are at least twice as many highly sophisticated thinking beings on this planet as there are persons.

So let us assume the correctness of animalism and ask what it implies about everyday moral thinking. The first thing to note is that the animalist can argue for Parfit's thesis in his own way without needing to appeal to the problematic uniqueness requirement on the condition of personal identity that Parfit's own argument requires.

Imagine that: "Your brain stem is replaced by an inorganic substitute ... bit by bit. The rest of you is left intact ... there is never a period when your life-sustaining functions are left without an organ to coordinate them, or when your cerebrum is not aroused and coordinated in a normal way by a brainstem. As a result, there need be no interruption of consciousness throughout the operation (suppose the surgeon uses only a local [anaesthetic](#)). The result would be a conscious rational being with your mind.... [Animalism] entails that you would not survive this. The resultant being would not be you.... it would not be a human being at all. For something with an inorganic brainstem would not be an animal at all.... If you are an animal you will not survive without an [organic] brain stem" (Olson 1997, 141-2).

Now suppose that you are faced with a choice between this operation (inorganic replacement) and an operation in which your brain stem is repaired and patched up, retaining its biological identity (maybe your brain stem is riddled with cancer so either replacement or repair is necessary). In each case the outcome is a person psychologically continuous with you. The difference between brain stem repair and patch up and inorganic replacement makes no difference to that. According to animalism, our opinion about this case, if we reflect on it carefully, should be (a) that we do not persist through inorganic replacement of brainstem but do persist through repair. But reflection on this case also generates, Olson would insist, and very plausibly, the opinion (b) that it would be quite irrational if you were facing this choice to prefer repair to inorganic replacement (as irrational as preferring to have your heart repaired rather than

replaced if both procedures were equally safe and effective) or to think that you could gain anything by choosing repair, even though it is correct to say that by doing so you would ensure your own future existence. The conclusion is then Parfit's: to reconcile these opinions we must say that we in fact have no non-derivative desire for our continued existence.

Imagine now a society with advanced medical treatment technology in which brain stem replacement by inorganic substitutes is a frequent occurrence and easily obtainable. Also suppose the people still have only the fundamental desires which, according to Parfit we actually have, that is, they do not have a non-derivative desire for their own continued existence. Then a legal system and societal norms according with the principle of the moral necessity of agency would be wholly inappropriate. In this society, given the advanced state of medical technology, reward or punishment founded on personal, that is, animal, identity would be purposeless. For people would not be much motivated to act well by the prospect of rewards given to the doer of good deeds since brainstem replacement would too often intervene; nor would they be motivated to refrain from acting wrongly by the prospect of future punishment meted out only to evildoers, since they could easily avoid it by electing for brain stem replacement. Thus, if animalism is correct, the principle of the moral necessity of agency is again revealed as a mere rule of thumb, the utility of which depends very largely on the present limited state of medical technology.

IV

I now turn to the Lewis's perdurance account of personal identity. Perdurantism is, in fact, neutral between the dispute over the role of psychological continuity in the explanation of

personal identity. A perdurantist can think this crucial or totally irrelevant. According to the perdurantist, persons persist by having temporal parts at different times. For our present purposes the crucial thing is that for him their temporal parts are the *subjects* of mental states, as persons are. Lewis explains this very clearly:

A person stage is a physical object, just as a person is.... It does many of the ... things that a person does: it walks and talks and thinks, it has beliefs and desires (Lewis 1983: 76)

The apparent conflict between the perdurance view and our everyday thinking is immediate. We think it unproblematic that in some circumstances at least, when there are no extenuating circumstances or excuses, and no other person (or animal) will suffer as a consequence, an evildoer should be punished for his deeds. But according to the perdurantist view, whenever punishment is meted out, some subject of psychological states, who only fails to be a person because he is part of a temporally longer one and is entirely innocent of the crime since he did not even exist when it was committed, is punished, or at least made to suffer, along with the morally responsible agent.

The apparent conflict is more general, as Olson (2010) and Johnston (2016, 2017) bring out. Olson calls the psychologically endowed temporal parts of persons the perdurantist posits “subpersons.” Johnston calls them “personites.” They note that, for example, if there is a subperson /personite now coinciding with me which will no longer exist tomorrow this renders morally problematic my planned visit to the dentist today since the subperson/personite, unlike

me, will suffer the pain today but not live long enough to experience gain. The same reasoning renders morally problematic spending time learning a difficult language for a trip abroad. And irritatingly, in accordance with this reasoning, the child who claims that making him do his homework isn't fair is arguably right. Moreover, since not only subpersons/personites as well as persons exist if perdurantism is the case, but also, for example, what we might call 'subdogs' or 'caninites', and these also need to be counted in the moral calculus, at least if dogs do, it follows that just as it is morally problematic to force a lazy child to do homework, it is morally problematic to put an obese dog on a diet. Underpinning the reasoning here is the principle of the intrinsicity of moral status, the thought that no relation one sentient being can have to another can deprive it of the right to be counted in the moral calculus (qua patient). *Being the wife of* cannot, *being a child of* cannot, *being the creation of* cannot. Nor can *being a temporal part of*. Thus, it seems that the perdurantist's ontology does not accord with our fundamental moral thinking. It renders morally problematic activities which according to our ordinary thinking are utterly unproblematic.

As Olson and Johnston note the conflict is not merely between perdurantism and ordinary moral thought. Any "generous" ontology which includes short-lived person-like entities or short-lived dog-like entities etc., coincident with longer-lived persons or dogs, generates the same conflict; it does not matter whether they are correctly thought of as parts of longer ones; it matters only that they can suffer.

In fact, seen in this way the challenge to the generous ontologist from our everyday moral thinking is anticipated by Locke. Locke has a tripartite ontology. He distinguishes between persons and human beings (men, as he puts it), but also between these and thinking substances, which may be material or immaterial:

It is not... unity of substance that comprehends all sorts of identity... but we must consider what idea it is applied to stand for; it being one thing to be the same substance, another the same man and a third the same person, if *person*, *man*, and *substance* are three names standing for three different ideas (*Essay II*, xxvii.7).

Hence it may be that:

Different substances by the same consciousness [are] united into one person... For it being the same consciousness that makes a man be himself to himself personal identity depends on that only, whether it be annexed to one individual substance, or can be continued in the succession of many (*Essay II*, xxvii.9)

Formatted: Font: Not Italic

So Locke notes, the question whether “if the same [immaterial] thinking substance be changed it could be the same person?” cannot be resolved “but by those who know what kind of substance they are they do think and whether the consciousness of past action can be transferred from one thinking substance to another” (*Essay II*, xxvii.13).

Formatted: Font: Not Italic

He goes on:

But why one intellectual substance may not have represented to it, as done by itself, what it never did, and perhaps was done by some other agent, will be difficult to conclude from

the nature of things. And that it never is so will by us, till we have clear views of the nature of thinking substances, be best resolved into the goodness of God, who as far is 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 that draws reward or punishment with it.

Locke's thought is that if, as his tripartite ontology allows, one thinking substance may be conscious of the actions of another, the two will make but one person. But then in punishing that person for his sins God will make suffer the thinking substance then coincident with it, even if it was never present at any past crime. So it will be punished for what it never did but was done by some other agent.

This is too much for Locke, so he appeals to God's goodness to rule out such a "fatal error." But his concern is clearly the same as Olson's and Johnston's about the generous ontology and contemporary generous ontologists cannot answer the challenge by appeal to such a deus ex machina.

V

It appears from the foregoing that to ensure absence of conflict with our everyday moral thinking and to allow the fundamental place in our moral thought of the principles we have discussed (the principle of the necessity of moral agency and the intrinsicality of moral status) an account of personal identity (a) must not involve a generous ontology, (b) cannot deny the relevance of

psychology to personal identity and (c) must be consistent with the only x and y principle. These constraints rule out the familiar accounts of Lewis, Olson and Shoemaker. What is left?

References

- Johnston, M. (2016). Personites, Maximality and Ontological Trash, *Philosophical Perspectives*, 30: 198–228.
- Johnston, M. (2017). The Personite Problem: Should Practical Reason be Tabled?, *Noûs*, 51: 617–44.
- Lewis, D. (1983). *Philosophical Papers*, vol. I., Oxford: Oxford University Press.
- Lewis, D. (1986). *On the Plurality of Worlds*, Oxford: Blackwell.
- Locke, J. (1694 (reprint 1975)). *An Essay Concerning Human Understanding*, Oxford: Oxford University
- Noonan, H. (2019). *Personal Identity*, 3d ed., London: Routledge.
- Olson, E. (1997). *The Human Animal*, New York: Oxford University Press.
- Olson, E. (2003). “An Argument for Animalism,” in R. Martin and J. Barresi (eds.), *Personal Identity*, Oxford: Blackwell.
- Olson, E. (2010). Ethics and the Generous Ontology, *Theoretical Medicine and Bioethics*, 31: 259–70.
- Parfit, D. (1971). Personal Identity, *The Philosophical Review*, 80 (1): 3–27.
- Parfit, D. (1984). *Reasons and Persons*, Oxford: Oxford University Press.
- Shaffer, J. (1977). Personal Identity, *Journal of Medicine and Philosophy*, 2:147–61.

Formatted: No underline

Shoemaker, S. (1963). *Self-Knowledge and Self-Identity*, Ithaca NY: Cornell University Press.

Shoemaker, S. and R.G. Swinburne (1984). *Personal Identity*, Oxford: Blackwell.

Williams, B.A.O. (1956-7). Personal Identity and Individuation, *Proceedings to the Aristotelian Society*, 57: 229–52.

Williams, B.A.O. (2014). *Essays and Reviews: 1959–2002*, Princeton: Princeton University Press.