

Plenitude, Pluralism and Neo-Lockean Persons

Introduction

Recently in the debate about personal identity neo-Lockean psychological continuity accounts and animalist biological accounts have been main contenders. Animalism may be understood as the weak thesis that we (human persons)¹ are (predicatively) animals. But in the sense in which it is opposed to neo-Lockeanism it is the stronger thesis that this is so *and* that psychology is irrelevant to the persistence conditions of (the kind of) animals (we are).² This is ‘the Biological View’ of Eric Olson (1997). Some neo-Lockeans are animalists in the weak sense. Thus Shoemaker’s recently expressed view (2010, see also his forthcoming paper in Blatti and Snowdon (eds.)) is a form of weak animalism. He holds that I am (predicatively) an animal in the sense (which must be a good one) in which dogs are. Dogs are *not* entirely biologically individuated entities he holds, since Fido goes where his (upper) brain and resultant psychology goes.

We will look at Shoemaker’s view later, but for now what I wish to emphasise is merely that the debates I am concerned with are those between neo-Lockeans and strong animalists. So when I write ‘animal’ I shall mean ‘biologically individuated animal’.

These debates interweave with more general debates about the nature of constitution. In his most recent book (2007) Olson identifies a generic view he calls ‘constitutionalism’ and a more specific view about persons he calls ‘the constitution view’.

Minimally, Olson’s constitutionalism is the view that two things can be made of the same matter at a time. His constitution view of persons is that things so related may be a person (thought of as something whose persistence conditions are psychological) and an animal (whose persistence conditions are entirely biological).

¹ I.e., persons related to human beings as I am related to the human being sitting here.

² Note (i) that this does not entail that we are essentially or fundamentally (biologically individuated) human beings and (ii) that this thesis is formulated indexically (in fact, animalism has no non-indexical formulation).

As Olson's terminology is intended to suggest, generic constitutionalism and the constitution view of person go naturally together. Indeed, he suggests that they stand or fall together. I agree, I take them as a package deal and defend both.

But begin by jettisoning Olson's terminology. As Shoemaker noted forty-five years ago (1970b), it is not good English to describe the symmetrical relation of matter-sharing, such as holds, according to the psychological continuity theorist, between a person and a biologically individuated animal, or again, as Shoemaker notes, between a statue and a temporarily coincident *piece* (contrast: portion or quantity) of bronze, as 'constitution'. It is increasingly common to describe the defender of constitutionalism as a pluralist, and I shall. Equally, it is common to describe what Olson calls the constitution view as the neo-Lockean psychological continuity account. So I will be defending pluralism in general and the neo-Lockean psychological continuity account in particular.

So I align myself with the views of Lewis (1976), Parfit (1971) and Shoemaker. But I am not wholly in agreement with the latter two, and I proceed in a way that is neutral with respect to the ontological commitments of Lewis to temporal parts and many worlds. Nonetheless, I think what I have to say is entirely Lewisian in spirit.

However, pluralism comes in two forms, strong and weak. Shoemaker and Parfit are examples of strong pluralists. They hold that even all-time material coincidence is not enough for identity: An all-time coincident statue (standardly called Goliath in the literature following Allan Gibbard (1975)) and lump of clay (Lumpl) are numerically distinct. This is because they differ modally: the lump could be rolled into a ball without being destroyed, not so the statue. Weak pluralists (like Lewis (1986)) hold only that numerically distinct things can be *sometime* materially coincident. They explain away the apparent modal difference in Gibbard's example as merely apparent and due to the inconstancy of modal predication, the

capacity of modal predicates to shift in reference according to the wider linguistic context in which they are embedded. I accept only weak pluralism.

I reject strong pluralism because it entails that two wholly material things, even if not ever coincident, may differ *macrophysically* in their general modal, dispositional and counterfactual properties, though *microphysically* indistinguishable in all general respects – non-relational and relational, past and future as well as present, even modal and dispositional. If Goliath and Lumpl are distinct there may be a distant twin of Goliath, Goliath*, permanently coincident with a twin of Lumpl, Lumpl*, on a twin Earth in a symmetrical universe (it would be desperate for a strong pluralist to attempt to defend his position by rejecting the possibility of symmetry). Goliath and Lumpl*, according to the strong pluralist, differ in general modal and dispositional respects just as Goliath and Lumpl do, though alike at the microphysical level in all general respects, including modal and dispositional respects. This is akin to supposing that two in all respects microphysically indistinguishable seeds from the same type of plant, which are never planted, would have developed differently if they had been planted. To believe in such macrophysical dispositional differences in the absence of any microphysical difference at all is to believe in magic. But the strong pluralist cannot explain what is objectionable about it.

Weak pluralism, by contrast, is very plausible. It seems as obvious as anything in philosophy that temporary matter sharing by numerically distinct things is a possibility. In the case of artefacts this is because they are typically made from pre-existent material and can undergo repair and replacement of parts later. Nevertheless, some philosophers reject even weak pluralism. Typically these philosophers also deny the existence of the complex material things which are the weak pluralist's candidates for temporary matter-sharing. One of the most prominent deniers is van Inwagen (1990), who denies the existence of any complex non-living material things. In the next section I will take the time to defend weak pluralism

by criticising van Inwagen's argument for this denial. Of course, other arguments remain, but it may be of interest to expose the weakness of one of the most well-known, though the noted commonsensical status of weak pluralism should make us suspicious of any arguments against it from the start.³

After defending weak pluralism I turn next to the defence of the neo-Lockean view. Two arguments for, and one against, have been the focus of most debate. In support there is the argument from the so-called 'transplant intuition' and the argument from what has been called 'the remnant person problem'. I defend these arguments. Against neo-Lockeanism there is the variously-called 'thinking animal', 'too many minds' or 'too many thinkers' objection, most vigorously explored by Olson. This has ontological, epistemic and semantic aspects. I shall argue that once these aspects are distinguished the argument can be disarmed.

I think the pluralist, true to his name, should regard the multiplicity as unobjectionable and view the situation as similar to how, with Lewis (1976) and Robinson (1985), I wish to view the situation before the fission in the familiar thought experiment.

Pluralism Defended

I start by criticising van Inwagen's argument against the existence of non-living complex material things. As we shall see, although this argument does not succeed, it makes it clear that a believer in complex material objects must accept that there many more than one might at first think, that is, that a plenitudinous ontology of complex material things is implicit in ordinary thought.

³ Opponents of weak pluralism like van Inwagen, who deny the existence of any complex non-living material things, or the still more extreme opponents (mereological nihilists) who deny the existence of any complex material things at all, are in the same position as the sceptic who denies that even if things are, in fact, as we believe they are we do not *know* that they are. We do not have to listen to the sceptics and the opponents of weak pluralism unless they provide arguments, and their arguments have to have premisses which are as initially attractive as the propositions they are attacking. It is not ruled out a priori that there can be such arguments, which is to say that our everyday view involves philosophically interesting tensions, but the arguments need to be seen and scrutinised.

Van Inwagen's argument, as he presents it, goes as follows. If I shake hands with someone nothing new comes into existence in consequence of our brief contact. If we become momentarily paralyzed as we shake still nothing new comes into existence. The same is true if we are glued or tied together, or even if our flesh is melted so that no boundary is discernible between us. In all these cases of contact, or various kinds of bonding, of living organisms, nothing new is brought into existence as a result of our coming to be so related.

But if there are any complex non-living things then in some of these cases novel complex things *would* come into being as a result of our coming to be related in the way described. For the relations just identified include all the kinds of causal and spatial relations which are created between bits of wood, for example, when, as we would ordinarily say, a chair is made. And van Inwagen states, 'whether certain things ... compose a larger object does not depend on anything besides the spatial and causal relations they bear to one another. For example, nothing outside the region of space containing some bricks is relevant to the question whether there is anything they compose; and, in particular, the attitudes and interests of any persons are irrelevant' (1990: 12) (this is van Inwagen's tenth numbered constraint on his theorizing).

You may say that it could still be that you and I, however bonded, do not compose a third object, whilst the bits of wood glued together do, because it matters for whether composition takes place not only what relations obtain between the putative parts, but also what *kind* of thing they are – bonding relations sufficient for composition when holding between things of one kind, e.g., bits of wood, may not suffice for composition when holding between things of other kinds, e.g., people. But this rejoinder would be deeply implausible, van Inwagen would say, for as he asks, 'if the operation *fastening* has the power to turn inanimate objects into the parts of a whole, why doesn't it have the same power with respect to living organisms?' (1990: 68). The plausible thought behind the rhetorical question here is

that if you can make an artefact out of inanimate components of a certain type, you can make an artefact of that same kind out of animate components, or inanimate components of any other type, so long as they are capable of performing the right functional job.⁴

But if a new thing is *not* brought into existence when bits of wood are bonded together in the sort of way we describe as ‘making a chair’, there never have been any chairs, and if there never have been any chairs, there never have been any complex non-living material objects at all.

This is van Inwagen’s argument. But it is easy to resist. Van Inwagen thinks that nothing new comes into existence when you and I are bonded together because ‘it is pretty clear that one cannot bring a composite object into existence by bonding ... human beings – or ... living things of any sort – to each other’ (1990: 62). But it is not pretty clear at all. Van Inwagen adds in a rhetorical endnote, ‘Try to imagine bringing something into existence by gluing hamsters or snakes together’ (1990: 287).

However, nothing is easier. Van Inwagen himself imagines at one point weaving a very long, thin, snake into a hammock. Imagine instead gluing together two or more shortish, flexible snakes to make a whip. If there are any non-living complex material things at all we can do this – and it is the *conclusion* of van Inwagen’s argument, not one of its premisses, that there are not. But if something new comes into existence when some snakes are bonded together the same must be true when human beings are bonded together in the same way.

So van Inwagen’s argument for the non-existence of non-living complex material objects – which is, when stripped of its surrounding paraphernalia, just the argument outlined – is unconvincing.⁵

⁴ These theses of van Inwagen’s commit him to the possibility of making composite things out of living organisms, by combining them in the way simples are combined when they compose an organism. But this is consistent with van Inwagen’s main claim that there are only simples and living organisms, and his rejection of arbitrary undetached parts.

However, it appeals to two plausible general principles. These are, to repeat, that whether certain things compose a larger object does not depend on anything besides the spatial and causal relations they bear to one another ... in particular, the attitudes and interests of any persons are irrelevant (1990:12) and that if you can make something out of inanimate components of a certain type, you can make a thing of that kind out of animate components, or inanimate components of any other type, so long as they are capable of performing the right functional job.

I think that these principles are implicit in our everyday thought. But if we take them seriously we must either deny, with van Inwagen, the existence of such everyday things as chairs, or acknowledge the existence of many more complex material things than we ordinarily speak of. I think that the right conclusion is that such a plenitudinous ontology is implicit in our ordinary thought.

To see how far down this path we can be led consider some other examples.⁶ If there are chairs there are children's toys, and if so I can bring such an object, say, a wooden brick house, into existence merely by bringing some smaller objects into contact. Van Inwagen would argue that I cannot since creating the relevant causal and spatial relations between living things cannot bring anything new into existence. Reversing his reasoning, but retaining his two general principles, we must conclude that whenever two things of any kind or kinds, living or non-living, come into contact, whether or not as a result of intentional activity, in such a way as to create the appropriate internal causal and spatial relations between them, then some new complex material object comes into existence.

Again, it is evidently a part of ordinary thought that there are many complex wholes composed of parts that are not in contact. Rock gardens, bird cages consisting of magnetically

⁵ Van Inwagen has another argument (1981) which assumes the rejection of DAUP, the Doctrine of Arbitrary Undetached Parts and appeals to other assumptions too. But its assumptions have little initial plausibility, certainly less than the initial plausibility of the proposition that there are indeed chairs.

⁶ Recall also that it is certainly no part of ordinary thought that there can be no artefacts apart from those created by human beings; the notion of an 'alien artefact' is an easy one to grasp.

suspended and spatially separated bits of wire, space stations with separated living and ablution blocks, the USA – all these consist of disconnected parts. And, of course, as van Inwagen notes, ‘it is undoubtedly true that, if there are any composite material objects at all, they are composed of elementary particles and elementary particles that compose a given material object are not in contact’ (1990: 34). Appealing to van Inwagen’s general principles we must suppose that smaller things of many kinds, whether animate or inanimate, can compose wholes whether or not they are in contact and whether or not the relations they stand in are a creation of intentional activity.

It does not follow that whenever there are distinct material objects there is a complex object they compose, i.e., Universalism does not follow.⁷ But given the plenitudinous ontology now argued to be implicit in everyday thought it is hard to see a reason to object. But Universalism, entails (weak) pluralism when taken together with the implication of the arguments above that whenever two things (e.g., snakes or bricks) come into the appropriate spatial and causal relations something *new*, of which they are parts, comes into existence,.

However, as I have already said, weak pluralism does not really need to be argued for; at most it needs to be defended against arguments, like van Inwagen, that attack its commonsensical presuppositions. That is what I have attempted in this section.

The Psychological Continuity Account of Personal Identity

Pluralists can endorse a psychological continuity account of personal identity and accept the possibility of numerically distinct but (temporarily) coincident persons and biologically individuated animals. But why should they?

⁷ I see no *deductive* route from, for example the proposition that children can create toy houses by arranging wooden bricks, together with the two general principles identified in van Inwagen’s argument, to the conclusion that something new comes into existence when two people shake hands, or that something existed all along of which they were parts.

The attractiveness of neo-Lockeanism derives from the puzzle cases which seem to make it evident that my history is only contingently coincident with that of a biologically individuated animal.

The standard puzzle case is Shoemaker's case of Brown and Brownson. In this the brain, or just the cerebrum, the upper brain (from now on I shall just say 'brain'), of Brown is transplanted into the skull of Robinson. Let us suppose that the result is a completely healthy being, Brownson, with Robinson's body but in character, memories and personality indistinguishable from Brown and this not as a consequence of some freak accident but because of his possession of Brown's brain. Now who is this person?

Most philosophers have found it difficult to deny that Brownson is Brown. As Parfit puts it, they have found it hard to deny that 'receiving a new skull and a new body is just the limiting case of receiving a new heart, new lungs and so on' (1984: 253).

This belief is the so-called 'transplant intuition'. The first argument for neo-Lockeanism and against animalism is that only the former accords with it.

The second argument for neo-Lockeanism I want to discuss is that it does not face what has come to be called 'the remnant person problem'.

If transplantation of a brain is possible so is the brain-in-a-vat scenario, in which the brain is taken out of a skull and preserved in a vat of nutrients as an intermediate stage. It seems undeniable that if appropriately stimulated such a brain would support a mental life; it would be, or be coincident with, a person.

The neo-Lockean can apparently take this in his stride. The animalist has problems. It can hardly be denied that there is a person in the vat. And, anyway, animalists, who do not deny the existence of (non-human) persons other than animals, would have no business doing so. But there is no animal present, they say. So the animalist must say that when the brain is envatted a conscious being is present who (which?) is not an animal, but came into existence

at the time of envatting, or existed beforehand, in either a non-conscious or conscious state (located in a proper part of the conscious animal). Similarly, he must say that when the brain is then transplanted the conscious being present in the vat either ceases to exist or comes to be partly coincident in location with a human animal who has acquired a new brain, but not a new identity, and in doing so either remains conscious or ceases to be conscious.

So the extended process first creates and then destroys a conscious being, or renders a previously non-conscious being conscious, later reducing it to non-consciousness again. Or else, either before being envatted or after transplantation, or both, there are two conscious beings present, one identical with or coincident with a proper part of the other – which is an alternative the animalist must avoid at all cost on pain of losing his best argument against the neo-Lockean, the ‘too many thinkers’ objection.

Mark Johnston (2007) has identified the problems the animalist has with this scenario, which have been dubbed by Olson (forthcoming) the creation and destruction problems. Animalism requires that you can *create* a person, or at least bring to personal consciousness a non-conscious being, simply by removing a sustaining head and torso. Again, it requires that you can *destroy* a person, or at least reduce one to non-consciousness, simply by adding a sustaining head and torso.

The neo-Lockean apparently faces no such difficulties. He can just say that when the brain is envatted, it is the person who is transferred into the vat, severely mutilated, but still conscious. No other conscious being need be supposed present and no conscious being need be supposed to have ceased to exist. On subsequent rehousing of the brain into the new cranium the person goes where it goes and acquires a new set of kidneys etc. No conscious being ceases to exist as a result or becomes non-conscious.

In defence of animalism Eric Olson has responded to the transplant intuition by pointing out that in real life where there is psychological continuity there is the same animal.

So he says that it is hardly surprising that when we contemplate the brain transplant case we mistakenly conclude on the basis of what is in real life conclusive evidence that the brain recipient is the brain donor, even though he is not, since he is not the same animal.

I think this response unpersuasive for the same reasons as Parfit (2012). In this case standardly used criteria come apart. Yet we find ourselves strongly favouring one rather than the other. This is what needs explaining. A finger-print criterion of personal identity cannot be defended as giving the right result in thought-experiments in which its deliverances come apart from those of the psychological continuity and brain-identity criteria on the ground that in real life cases they coincide.

More interesting is Olson's second response to the transplant intuition, which is to try to explain it away on the basis of Parfit's thesis (Parfit 1971) that identity is not what matters in survival. The recipient, he says, has what matters in the survival of the brain 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 morally responsible for the donor's actions and it is the recipient rather than the surviving donor (even if provided with a replacement brain) who should be regarded as for all practical matter the same person. So we do believe mistakenly that the recipient is the donor. This is because we recognise that he is his *Parfitian* survivor (the person who has what matters in his survival) and we mistakenly believe that identity is what matters.

Of course, this explanation is only available if Parfit's conjunctive thesis *that identity is not what matters in survival, but we believe that it is* is correct. But this is far from clear, and it is additionally not clear whether an animalist can endorse Parfit's argument for his thesis.

What Parfit means by his slogan is we do not have a non-derivative concern for our own future existence and well-being (though we believe we do). It is this, if anything, that the

argument from fission establishes. The best formulation of the argument goes as follows. Looking forward to an imminent future in which my cerebral hemispheres will be divided and transplanted, so that two future people will be psychologically continuous with me as I am now, I should not think that I have anything to lose that I would preserve if the division were to be prevented by the botching of the transfer of one hemisphere, say the right one. But the correct description of the latter, 'botched', fission is that I continue to exist, albeit with only one hemisphere, and of the former, the successful division, that I cease to exist and two new people, numerically non-identical with me, come into existence. This is how we must describe the cases using the language of identity. So if my continued existence matters to me, I do have something to lose in a successful fission and should reject it if given a choice between these options. So my continued existence does not matter to me.

This argument rests on two claims. The first is that accepting the description of fission as my ceasing to exist, and the description of the 'botched' fission as my continuing to exist, albeit with one cerebral hemisphere, I should, given what I care about, regard myself as having nothing to lose by fissioning. The opposing opinion is nicely expressed by van Inwagen: 'Suppose I were given a choice between being totally destroyed a year from now and being partitioned now. Then I think I would choose to be destroyed in a year' (1991: 212). Jerome Shaffer expresses the same sentiment: '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).

So the first claim on which Parfit's argument rests is contestable.

The second claim on which it rests, also contestable, is that someone ceases to exist in the fission. An alternative description of the case is that it involves multiple occupancy: there

are two conscious beings, coincident before the fission, who continue to exist afterwards and to remain conscious, but are then no longer coincident. Parfit's description of the case requires that one accept a 'best candidate' or 'no rival candidate' account of personal identity according to which whether I continue to exist depends on whether there is anyone *uniquely* best qualified to count as me at the later time. (As will be the case if the fission is 'botched' and will not be if it is successful.) This is a highly contentious form of account. It conflicts with the principle formulated by Bernard Williams that whether a later individual is identical with an earlier one can depend only on facts about them and their relationship; no facts about any other individual can be relevant to whether they are one. This has been called this the Only x and y principle (Noonan 1989). More immediately relevant, it is not at all obvious that an animalist should endorse a best candidate or no rival candidate of personal identity. I have referred to van Inwagen's fundamental tenth constraint on his theorizing. Rejection of such accounts goes with acceptance of this. Van Inwagen writes: 'A similar point [i.e., to his tenth constraint] ... applies to identity across time. If object A is at place x at t1, and if object B is at place y at t2, then nothing besides the causal processes or chains of events that connect what is going on at x at t1 with what is going on at y at t2 is relevant to the question whether A and B are the same. Closest continuer and best candidate theories of identity across time provide examples of theories that violate this principle' (1991: 12). Eric Olson writes of the 'non-branching' or *Uniqueness Requirement* (that you survive only if you stand in some relation to exactly one future or past being): 'It is a startling claim ... but no one accepts the Uniqueness Requirement because it sound 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' (1997: 49). However, if the animalist is to endorse the argument for the thesis that identity is not what matters in survival, so he can call on it to explain away the seeming

evidence of the transplant intuition, he needs to endorse a best candidate or no rival candidate account of fission and accept the Uniqueness Requirement.

The transplant intuition thus seems to remain a very strong argument for neo-Lockeanism.

The same is true, I think, of the remnant person problem.

The animalist can say nothing about this which is not wholly implausible or commits him to (what he views as) the awful prospect of ‘too many minds’. Focus just on the initial envatting of the brain. The animalist says that the original person (= animal) is not involved; no animal ends up in a vat. So either some new conscious being is created, or some pre-existing non-conscious being is rendered conscious by the process, or some conscious being other than the original person (= animal) was already present – too many minds.

The most the animalist can hope to defend is the claim that the remnant person problem is a problem for anyone – even the neo-Lockean.

According to the neo-Lockean before the procedure begins the person is not a living organism, but is coincident with, that is, shares its matter with, one. After the envatting the person is not a brain, but similarly shares its matter with a brain.

Olson argues that this merely relocates the remnant person problem. ‘Although removing your brain from your head would not give it consciousness or the power to think, it would give it the power to constitute a conscious, thinking thing. Your brain is now prevented from constituting a thinker by its fleshy surroundings and putting it back where it belongs after its removal would prevent it from doing so once more.... And that seems absurd’ (forthcoming).⁸

⁸ Of course, Olson is here using ‘constitute’ in the minimal symmetrical sense (of which Shoemaker disapproves) of ‘be composed of the same matter as’.

He goes on to say that any explanation of why a brain could *constitute* a thinking thing when isolated but not in its natural habitat would serve equally well as an explanation of why a brain could think when isolated but not embodied.

But this is not so. The fact that the neo-Lockean needs to appeal to in order to explain why a brain can be composed of the same matter as a thinking thing when isolated but not embodied is merely the fact that a person can survive mutilation – and thus come to be composed of the same matter as something which was previously one of its proper parts – and remain conscious. If you cut off my foot I do not cease to exist or to be conscious. But I come to be composed of the same matter as, i.e., to be coincident with, something that I was not coincident with before – something that was previously merely a proper part of the two-footed animal I was then coincident with and was, therefore, just because of its fleshly surroundings, not then an animal or conscious being or composed of exactly the same matter as any animal or conscious being. The removal of the brain for the neo-Lockean is merely a more extreme mutilation. What happens is just that a continuously conscious being is reduced in size. This is what the animalist cannot say happens since he cannot say that any conscious being *persists* through the mutilation. Compare the case of the cat on the mat. Tibbles is sitting on the mat. Tib is a proper part of Tibbles (all of Tibbles except its tail = its tail-complement). It is not a cat. Tibbles' tail is amputated. Tibbles does not cease to exist. But now Tib coincides with Tibbles. We are not obliged to say either that Tibbles has ceased to exist and been replaced by a new cat, or that Tibbles has ceased to exist and Tib has become a cat, or even that Tib is now (predicatively) a cat as well as Tibbles – for we can say that whether something is a cat depends on its history as well as its current state. For the neo-Lockean what happens to you when you are reduced to the size of a brain-in-a-vat is logically analogous to what happens to Tibbles when his tail is amputated.

Of course, this leads us on to the problem of too many thinkers. The person coincides with a biologically individuated animal beforehand and with a brain and perhaps a mutilated biologically individuated animal as well afterwards. Granted that these coincident entities are not (predicatively) persons (just as Tib is not a cat), the questions still remain. Are there two thinkers before or is the animal not a thinker at all? And are there two/three thinkers afterwards?

The Too Many Thinkers Objection

The objection, briefly, is that it seems undeniable that normal healthy adult human animals are thinkers. But so, by definition, are persons. However, according to the neo-Lockean, persons are not (biologically individuated) animals. So the neo-Lockean account entails the existence of too many thinkers. Moreover, it creates an irresolvable epistemic problem. How do I know I am the person and not the animal thinking falsely that it is a person? Finally, if normal healthy human animals are thinkers, they are persons, since their thoughts are sufficiently sophisticated, so the neo-Lockean's attempt to identify the persistence conditions of persons collapses, since he must acknowledge the existence of different kinds of person with different persistence conditions.

The too many thinkers problem has various aspects. The first is ontological. The objection is that neo-Lockeanism entails an objectionable multiplicity of coincident thinking things. I think that the neo-Lockean can just respond by saying the multiplicity is not objectionable. Numerically distinct things may coincide, at least temporarily, and hence share many properties. Where the coincident statue and piece of clay is there are two things of the same shape, size, weight and colour, and where the person and animal are there are two things of the same shape, size, weight and colour. He must admit that it is odd to say these things. But he can insist that they are true anyway. So the claim that it is *odd* to say that there

are two subjects of psychological states, two subjects of pain, for example, where I am, should cut no ice with him.

It is important to keep in mind that persons are not the only subjects of psychological states. So are dogs. Now the transplant intuition is just as powerful as an argument that brain transfer is identity-preserving if we suppose the experiment carried out on dogs instead of human beings. So the neo-Lockean must say – as Shoemaker (2010 and forthcoming) does – that even before the transplant there are two coincident, dog-shaped and dog-sized entities present, one of which goes where the brain goes. This entity, the one that goes with the brain-transplant, is a subject of psychological states. Why is the other, which shares so many of its features, not? Sometime in the future, perhaps many dog-years away, these coinciding entities will go their separate ways when the transplant is done. But why is it not true to say *now* that there are two perceiving, desiring and believing entities present, just as there are, the pluralist neo-Lockean must say, two digesting, breathing, running entities? How can this merely future difference make such a difference *now* in the case of the psychological, but not the non-psychological, predications?

If we think about the case of non-human animals, I think it is obvious that the too many thinkers problem provides no reason to reject neo-Lockeanism unless it provides a reason to reject pluralism in general.

But humans are merely psychologically more sophisticated than dogs, not entities of a completely different ontological category. So the neo-Lockean should feel no more pressure to deny the evident truth that human animals are the subjects of psychological states than to deny that (the animals coincident with) dogs are. He can simply say of the situation in which Mr Brown is going to undergo a transplant exactly what he can say of the situation in which this is true of Fido. Each is currently coincident with an entity with which it currently shares, along with much else, its psychology.

But persons have first-personal thoughts and so we come to the other aspects of the too many thinkers problem. How do I know I am the person and not the animal, mistakenly thinking it is a person? Fido does not face this problem.

Now perhaps all persons and certainly only persons are objects of first-person reference. Objects of first-person reference are a sub-class of persons (anything which can think of itself as 'I', that is, is a 'thinking intelligent being with reason and reflection which can think of itself as itself, the same thinking being' (Locke), is a person). Whether or not analytic this is obviously true. So we can define the philosophical topic of personal identity, as it has been discussed since Locke, without using the word 'person'. We need three concepts: First, the concept of a word whose meaning is that of the first-person pronoun, secondly the concept of a token utterance, and thirdly the concept of reference. The class of interest is the class of entities that are the referents of token utterances of first person-pronouns, the class of *selves*, we may say (perhaps, as the word is ordinarily used, there are persons that are not selves, but they are not the focus of the philosophical debate). The philosophical question about the persistence conditions of persons over time is a question about some of the necessary conditions of membership in this class – of selfhood.

Now the neo-Lockean position is that all persons (along with creatures of less complexity) have psychological persistence conditions. So, since it is obviously true that all objects of first-person reference are persons, this entails that all objects of first-person reference have psychological persistence conditions. But some normal adult human animals do not, as everyone agrees. So they are not objects of first-person reference according to the neo-Lockean. But all normal adult human animals are *thinkers* of true first-person thoughts. Or so I have effectively just argued, against e.g., Shoemaker, that the neo-Lockean must accept, on pain of making an unacceptable divide between Fido and Mr Brown, or denying that the biologically individuated animal coincident with Fido shares (albeit temporarily) his

psychology. So some normal adult human animals are thinkers of true first-person thoughts of which they are not the objects, i.e., which are not thoughts about them, but about the psychological continuers with which they coincide. They are *thinkers* of first-person thoughts, but not *objects* of first-person reference. This is what the neo-Lockean must say on pain of denying that human animals are subjects of psychological states at all.

But how can a normal adult human being, capable of first-personal thought, be incapable of referring to itself in the first person in thinking such a thought, since the currently psychologically indistinguishable psychological continuer is? How can it lack this capacity?

To ask this is to mistake the commitment of neo-Lockeanism. The claim is just that the following is a de dicto necessary truth: ‘All objects of first-person reference are psychological continuers’.

It does not follow that if human animal A is not a psychological continuer it is essentially or necessarily something whose first-person thoughts are not about itself. No claim of de re necessity follows (though many, unlike me, with natural essentialist intuitions, including many neo-Lockeans, will find it plausible, particularly strong pluralists). To be a psychological continuer is to have a certain history and A could have had a different history. So A is not essentially something whose first-person thoughts are not about itself; it is just something whose first-person thoughts are not, in fact, about itself.

The worry may be re-expressed. According to the neo-Lockean something that is not a psychological continuer cannot be an object of its own first-personal thoughts. It cannot think that it itself is F, by thinking a thought it can correctly express in the words, ‘I am F’, even if it can think the thought. But why should this be?

However, the explanation is simple. To be a psychological continuer is to have a certain kind of history. It is like being a once and future king. But whether something has

psychological states now and how sophisticated they are, if so, cannot depend on its indefinitely distant past or future. So no matter how sophisticated a creature's current thoughts, they cannot guarantee that it is a psychological continuer. Consequently, *if* the neo-Lockean thesis that all objects of first-person reference are psychological continuers is correct, they cannot ensure that it is an object of its own current first-personal thoughts.

So there is no epistemic problem for neo-Lockeanism. There is no possibility that I am not the object of self-reference but the animal mistakenly thinking that it is the object of self-reference. Since I am thinking that I am an object of self-reference, the animal is thinking the thought it would express by saying 'I am an object of self-reference', but in doing so it is not thinking about itself, but about me.

But there is a final worry. Given that I and the human animal temporarily coincident with me are psychologically indistinguishable at present, it must be that whatever I think, the animal thinks the same. So the animal must think that it is an object of self-reference, since I do, so it must be mistaken.

The trouble with the argument is that it ignores an ambiguity in 'thinking the same'. The worry is that I the animal must be psychologically indistinguishable, so we must be thinking the same thoughts, so the animal must be thinking that it is an object of self-reference if I am.

But if you and I both pronounce the Cogito in one sense we think the same thought, in another sense not. In one sense two thinkers think the same thoughts only if they think of the same things and think of them in exactly the same way. In this sense – we may call it the Fregean sense – I and the animal are thinking exactly the same thoughts. Each of us is thinking of me in a first-person way and neither of us is thinking of the animal. So the animal is not thinking of itself in a first-person way (though, of course, its thought 'I am thinking about myself in the first-person way' is true).

So the answer to the question ‘why is the animal not thinking of itself since I, coincident with it, am?’ is that this must be so given that we are psychologically indistinguishable in *this* respect: we are thinking the same Fregean thoughts.

Hence the question facing the neo-Lockean who is not willing, with Shoemaker, to deny that human animals think is not ‘How can the animal and I *differ* in what we think, given our coincidence?’, but why, given our numerical distinctness, we are cognitively identical, thinkers of the same Fregean thoughts.

But it is not difficult for the neo-Lockean to answer this, given his physicalist stance. I and the animal may be completely microphysically indistinguishable except in irrelevant long-distant past and future respects; it would be a mystery if we were not so cognitively identical, were thinking about different things or thinking of them in different ways.

It is often said that the meaning of ‘I’ can be explained by saying ‘it is the word each one uses to speak of himself’. But no word’s meaning can be such that any user of it is *guaranteed* to make singular reference to himself when he uses it. A user of ‘I’ refers to himself if he makes it true that he satisfies the open sentence ‘x refers to x’ – something the slayer of Laius will do if he begins a sentence ‘The slayer of Laius ...’, whether or not he knows he is the slayer of Laius. But his use of ‘I’ cannot guarantee that he does refer to himself. The illusion of such guaranteed self-reference is what lies behind the thought that the epistemic and semantic aspects of the too many thinkers problem are objections to neo-Lockeanism.

To see that a guaranteed self-reference for ‘I’ is an illusion consider the familiar fission case and compare it with the (whole) brain transplant case. In the fission case when the two hemispheres are transplanted, according to the ‘no rival candidate’ or ‘best candidate’ neo-Lockean accounts of Shoemaker (1970a) and Parfit (1971), the original person ceases to exist and two new persons come into existence. If we reject these accounts because of their

conflict with the Only x and y principle we must, with Lewis (1976) describe fission as multiple occupancy. Before the fission two persons, two conscious beings, Lefty and Righty, are coincident. They cease to be coincident with the fission, but continue to exist and continue to be conscious. On the neo-Lockean view the whole brain transplant should be understood similarly. Before the transplant two conscious beings are coincident. They cease to be coincident with the transplant, but continue to exist, though only one continues to be conscious. The difference between the cases is just that fission is entirely symmetrical.

Now in the fission cases, if described as cases of multiple occupancy, utterances of 'I' before the transplant are not guaranteed a determinate singular reference. What is Lefty referring to when says (simultaneously with Righty) 'I am hungry'? He cannot be speaking of himself alone since he has no way of uniquely identifying himself – he is a person, hence a psychological continuer, thinking the thought 'I am hungry', but so is Righty. Either he fails to refer or, as Lewis (1976) says, he thinks a plural thought, with the content 'We both ...' or 'At least one of us ...'. If he does not know what is going to happen most plausibly his thought has the content (to use Lewis's language now) 'the maximal sum of pairwise R-related person-stages tokening this thought ...', in which case he fails to refer since there are two such maximal sums.⁹ The case is essentially no different from the following: I use 'Tom', as I think, to refer to one of my acquaintances; in fact two identical twins, Dick and Harry, have been fooling me.¹⁰ Hence, if it is known that fission is going to take place there

⁹ A single token of 'I' may be produced by two speakers with differing intentions so that two singular references take place. Compare the case (Johnston 2006) in which two people with different intentions together create and put in place the ambiguous road sign 'Begin Highway' -- one intending it as the name of the highway ('the (Menachem) Begin Highway') and the other as an instruction ('Start highway'). This makes sense because we can imagine the sign constructed and erected by a single person (Mr Janus) who has both intentions (and gets paid twice). In the fission case, however, the two people do not have two different reference-determining intentions which a single person could have simultaneously had.

¹⁰ I am here accepting Lewis's view that 'I' uttered pre-fission when fission is not being envisaged has the status of an improper description. It would serve my purpose equally well to regard it as determinately referring to the current shared stage. What I deny is that in this situation the persisting person Lefty is successfully referring to Lefty alone and the persisting person Righty is simultaneously referring to Righty alone. This has been urged by Saunders and Wallace (2008, 2008a) and contended by Tappenden (2008, 2010). I am with Tappenden.

is no intelligible doubt Lefty can express by uttering ‘I wonder whether I am Righty or Lefty?’ Lefty faces no epistemic problem.

In the whole brain transplant case, by contrast with the fission case, there is an asymmetry. There is only one maximal sum of R-related stages, only one psychological continuer. So the person, the maximal sum, *can* determinately refer pre-transplant to himself using ‘I’. But the animal cannot. It can only refer to the person. (It can, of course, determinately refer to itself using the description ‘the animal’ and when it does so the person also refers to the animal.) In the fission case, if described consistently with the Only x and y principle, neither conscious being who pre-exists the fission and (literally) survives it succeeds in making a singular reference to itself at all. In the transplant case both conscious beings make singular references with ‘I’, but they do not make singular references to *distinct* individuals. Hence only the person, the maximal sum of pairwise R-related person-stages, makes a singular reference to himself. On the opposing view, which underpins the animalist’s objection to neo-Lockeanism, the use of ‘I’ guarantees a reference, but if so Lefty and Righty determinately refer in the fission case, thought of as involving multiple occupancy, and Lefty is asking a good question if he asks ‘Will I be going off to the left or the right?’

If the fission case is not thought of as involving multiple occupancy the possibility of singular reference by the single conscious being present before the fission returns. But for the neo-Lockean the brain transplant remains a case of asymmetric multiple occupancy unless it is denied that the biologically individuated human animal is a conscious being. So the person and the animal are not expressing different singular thoughts when they utter ‘I’ and the too many thinkers problem has an obvious solution. In a sense, we may say, the solution is that they *do* face an epistemic problem, the same epistemic problem that the possibility of fission

presents if the Only x and y principle is accepted. But this is just the ‘problem’ that even reference with ‘I’ is not guaranteed.¹¹

Alternative Solutions to the Too Many Thinkers Problem

Shoemaker and Parfit agree that the too many thinkers problem is not an insuperable objection to neo-Lockeanism, but they differ from me in how they answer it (since they do not wish to say that before the transplant the biologically individuated human animal is a conscious being who cannot refer to himself in the first-person).

I want to finish by rejecting their answers.

Shoemaker denies that biologically individuated animals, human or non-human, have psychological states at all. He therefore denies that dogs are animals in this sense (and says that in a brain transplant the dog goes with its brain). (So he is able to say, despite his opposition to (strong) animalism, that we are predicatively, animals, i.e., animals in the sense in which dogs are.)

He distinguishes thick and thin properties: thin properties can be shared by (temporarily) coincident entities with different persistence conditions; thick properties cannot. Size and shape are thin properties, psychological states are thick. A thick property may be thought of as a conjunction of a thin property with a sortal property which determines a set of persistence conditions. So a person and coinciding biologically individuated human animal share all their thin properties but not their thick properties because they differ in their persistence conditions. And this is so even if we suppose them to have had the same origin – even if we suppose the animal to have been created fully grown, rather than developing from

¹¹ As well as the fission case there is the possible case of conjoined twins sharing a cerebrum but not a brainstem nor any other vital organs involved in the life-processes thought to individuate organisms (Metz 2001: 289-90). The animalist must say that these are two thinkers, each of which is thinking a single thought about itself when they simultaneously token ‘I’ (like, he must say, pace Lewis, Lefty and Righty in the fission case). This must be so if an ‘I’-thought has a guaranteed reference, as he requires for his argument against the neo-Lockean. But it does not seem plausible to say that in this case either animal will be able to refer to itself or know that it was one rather than the other.

a foetus. In fact, according to Shoemaker qua strong pluralist, even if the person and animal are permanently coincident they will still differ in their persistence conditions. So the animal will still lack any psychological properties.

I disagree with Shoemaker. I do so, first, because I reject his strong pluralism. But nor do I accept that whether something is now a possessor of any psychologically properties can depend on its future or, except causally, on its past in the way he requires. Some properties are possessed by things at times partly in virtue of what happens at other times. It is implausible that the matter only momentarily coinciding with a person or animal thereby qualifies as either. It is not implausible that psychological properties can only be possessed by things with careers of non-zero temporal extent. The possession of a psychological property by a thing at a time must cause or be caused by its possession of some psychological property at another time. This is part of Shoemaker's position. But the difficulty is this. I and the animal with which I am currently coincident have been together a long time and, God willing, will continue to be so for a few years yet. So if I am now thinking that it is raining and the animal is not is this difference inexplicable, or is to be explained by differences in non-actualized potentialities or grounded in some long past or future difference? None of these options is attractive. Although it seems absurd to say that the matter now momentarily coincident with me is thinking that it is raining, it seems far from absurd to say this of the animal with which I have been and will be, so long coincident. Granted thinking, like breathing and digesting and running, takes time – which is why it is absurd to think of the momentarily coincident piece of matter as doing any of these things – it does not seem that the period of time need be years long. David Lewis's (1981: 76) day-long person- stages can count as thinking and walking and talking, even though, as he notes, they cannot do everything a person can do, since they cannot do what a person can do only over a longish

period. The biological animal coincident with me can do at least whatever my current day-long person-stage can do.

Derek Parfit's response to the too many thinkers objection is different. He denies that human animals think in the way that persons do, not because they have merely biological persistence conditions, but because they are too big. They contain parts not required for thought so they do not think. You are in fact, brain-sized. He calls this view the 'Embodied Parts View'. One version of this is that you are your brain. But Parfit thinks a better version is that the thinking embodied part of an animal, i.e., the part you are, is not the brain, but something related to a brain in a way roughly similar to that in which an animal is related to its whole body. This is the Embodied Persons View. You are an entity distinct from but coincident with the brain contained within your skull. So when the transplant takes place you go where the brain goes and when your brain is envatted so are you. No new conscious being is brought into existence, nor is a previously non-conscious being rendered conscious by the mere removal of fleshly surroundings.

Parfit endorses the Embodied Persons View as a way of defending neo-Lockeanism. He formulates what he defends as follows:

The Narrow, Brain-Based Psychological Criterion: If some future person would be uniquely psychologically continuous with me as I am now, and this continuity would have its normal cause, enough of the same brain, this person would be me. If some future person would neither be uniquely psychologically continuous with me as I am now, nor have enough of the same brain, this person would not be me.

My objection is that the Embodied Persons View is as vulnerable to the too many thinkers objection as the standard neo-Lockean view.

It is, first, obvious why Parfit is not willing to identify persons with their brains. These are biological organs, whose persistence conditions are entirely non-psychological.

Like biologically individuated animals their continued existence does not require any psychological function. So if I am my brain I existed before I had any psychological properties and may do so again. And if my brain is transplanted I will go where it goes regardless of whether it carries any psychology with it. This is not what Parfit wants to say.

But the Embodied Persons View avoids these consequences only if the persistence conditions of the thinking part which is coincident with the brain are at least partly psychological. But then the brain and the thinking part will not necessarily have the same history. They will not come apart in a brain transplant, but they will be capable of coming apart in other circumstances. Indeed, they *will* have different histories, except in science-fiction cases in which human beings come into existence fully developed without any foetal stage.

So Parfit faces the too many thinkers objection. He must say that the brain is not a thinker, though it is coincident with one and has the same surroundings and virtually the same history. He cannot explain this by saying that this is because it is too big, with parts not required for thought. So he is left with the same problem Shoemaker faces.

I conclude that neo-Lockeans should not deny that biologically individuated animals think. They should acknowledge that these animals think, accept the Only x and y principle and celebrate their pluralism by endorsing a multiplicity of thinkers.

References

Gibbard, A. 1975 Contingent Identity. *Journal of Philosophical Logic* 4: 187-222.

Johnston, M. 2006_Hylomorphism. *Journal of Philosophy* 103: 652-698.

Johnston, M. 2007 Human Beings Revisited: My Body is Not an Animal. *Oxford Studies in Metaphysics* 3:33-74.

- Lewis, D. 1976 *Survival and Identity*. In Amelie Oksenberg Rorty (ed.), *The Identities of Persons*. University of California Press. 17-40.
- Lewis, D. 1981 *Philosophical Papers*, vol. 1. New York: Oxford University Press
- Lewis, D. 1986 *On the Plurality of Worlds*. Oxford: Basil Blackwell.
- Metz, W. 2001 *Ultrasound in Obstetrics and Gynecology*. New York: Thieme Medical Publishers.
- Noonan, H. 2003 *Personal Identity*, 2nd edition. London: Routledge.
- Olson, E. 1997 *The Human Animal: Personal Identity Without Psychology*. Oxford: Oxford University Press.
- Olson, E. 2007 *What are We? A Study in Personal Ontology*. Oxford: Oxford University Press.
- Olson, E. forthcoming *Animalism and the Remnant Person Problem*. In J. Gonçalves (ed.) *Metaphysics of the Self*. Instituto de Filosofia da Linguagem, Universidade Nova de Lisboa: Peter Lang
- Parfit, D. 1971 *Personal Identity*. *Philosophical Review* 80 (January):3-27.
- Parfit, D. 1984 *Reason and Persons* Oxford: Oxford University Press
- Parfit, D. 2012 *We Are Not Human Beings*. *Philosophy* 87:5-28.
- Robinson, D. 1985 *Can Amoebae Divide Without Multiplying?* *Australasian Journal of Philosophy* 63:299 – 319
- Saunders S. and Wallace D. 2008 *Branching and Uncertainty*. *British Journal for the Philosophy of Science* 59 (3):293-305.
- Saunders, S. and Wallace, D. 2008a *Saunders and Wallace Reply*. *British Journal for the Philosophy of Science* 59 (3):315-317
- Shoemaker, S. 1970a *Persons and their Pasts*. *American Philosophical Quarterly* 7:269-85
- Shoemaker, S. 1970b *Wiggins on Identity*. *Philosophical Review* 79:529-544.

- Shoemaker, S. 2010 Reply to my critics. *Philosophical Studies* 148: 125-32.
- Shoemaker, S. forthcoming Thinking Animals without Animalism. In S. Blatti, and P. Snowdon (eds.) *New Essays on Animalism: Persons, Animals and Identity*, Oxford: Oxford University Press.
- Tappenden, P. 2008 Saunders and Wallace on Everett and Lewis. *British Journal for the Philosophy of Science* 59 (3): 307-314.
- Tappenden, P. 2010 Varieties of Divergence: A Response to Saunders and Wallace.
http://philsci-archive.pitt.edu/5384/1/Varieties_of_Divergence.pdf
- Van Inwagen, P. 1981. The doctrine of arbitrary undetached parts. *Pacific Philosophical Quarterly* 62:123–137.
- Van Inwagen, P. 1990 *Material Beings*. Ithaca: Cornell University Press