

# Animal Self-Awareness

Rory Madden\*

---

Part of the philosophical interest of the topic of organic individuals is that it promises to shed light on a basic and perennial question of philosophical self-understanding, the question *what are we?* The class of organic individuals seems to be a good place to look for candidates to be the things that we are. However there are, in principle, different ways of locating ourselves within the class of organic individuals; organic individuals occur at both higher and lower mereological levels than the stereotypical bounded and physiologically unified vertebrate organism. The view that we are organic individuals smaller than the physiological organism is one that has recently been endorsed by Derek Parfit. This paper attempts to resolve a dispute between Parfit's 'embodied part' view and a contrary 'animalist' view according to which we are whole organisms. It is explained why a problem of *multiple thinkers* presents a serious obstacle to a straightforward resolution of this dispute. Parfit's own strategy for dealing with this obstacle is found to be problematic. However his strategy has a certain general shape, which can be instantiated in a different, and better, specific way. I close in the final section, not with a definite resolution of the dispute, but with illustration of how progress on the question of about which organic individuals we are requires engagement with questions about the nature and function of *self-awareness*.

---

## Keywords

animalism • Derek Parfit • embodied parts • multiple thinkers  
first-person thought • self-awareness

*Part of a special issue, Ontologies of Living Beings, guest-edited by A. M. Ferner and Thomas Pradeu*

**Editorial introduction:** One area of particular interest for both metaphysicians and philosophers of biology is the so-called 'personal identity debate'. Theorists working in this sphere aim to investigate our 'ontological character' (what kinds of beings do we think we are?) and our 'persistence conditions' (what does it take for us to persist through space and time?). Given that a plausible answer to the first question is 'organisms,' it's not hard to see how biological concerns might impact on the metaphysical discussions.

---

\*Department of Philosophy, UCL, Gower St, London, WC1E 6BT United Kingdom,  
r.madden@ucl.ac.uk

Received 2 September 2016; Accepted 20 January 2017  
doi:10.3998/ptb.6959004.0009.009



Much of Rory Madden's work falls within the ambit of this personal identity debate. His is an 'animalist' account; he holds that we are fundamentally *human animals* (or organisms), rather than fundamentally persons. In this essay, he defends this position against the well-known 'multiple-thinker problem' (and doing so hits some of the same notes introduced in John Perry's article "Identity and Self-Knowledge").

The multiple-thinker problem is a problem about individuation and self-knowledge. It emerges from the thought that there may be (at least) two thinking entities sat where you are currently sat: an organism and a brain. To which one do you refer when you say or think 'I'? Madden's article continues the discussion started in Perry's piece (while presenting an interesting alternative picture of self-awareness)—but it also tackles one of the primary issues we'd hoped to target in this collection: the notion of ontologically overlapping entities.

According to Madden, there are competing claimants to the title 'I,' and this article is an investigation into how each claimant is picked out (or 'individuated'), and to which one we refer when we *self*-refer. Thus, we can read this piece as an expression of the thought that biological reality can be divided up in different ways according to our modes of division. Moreover, Madden's discussion of self-awareness puts an interesting emphasis on specific capacities—*proprioception*, *kinesthesia*, *sense of balance* and *visual self-location*—which rarely feature in accounts of biological individuation proposed by philosophers of biology. Might self-awareness, as proposed by Madden, feature as another principle of biological individuation (like the *immunological*, *genetic*, or *functional* ones?).

Additionally, Madden's essay helpfully articulates the notions of *derivation* and *dependence*. The *metaphysical* aspect of the dependence relation between parts and wholes is somewhat overlooked in the philosophy of biology (though as Catherine Kendig shows in her "Re-engineering Metaphysics," the terms 'dependence' and 'parthood' are both mainstays of the field). Further, like Christopher Shields (in "What Organisms Once Were and Might Yet Be"), Madden discusses *teleological* relations, and demonstrates the theoretical utility of that otherwise unlovely notion.

–AF/TP

## 1

Part of the philosophical interest of the topic of biological individuals is that it promises to shed light on a basic and perennial question of philosophical self-understanding, the question *what am I?* The class of biological individuals seems a promising place to look for candidates for the things that we are.

In the contemporary metaphysics literature so-called *animalist* theories locate us within the class of biological individuals in a straightforward way. According to these theories we self-conscious human persons are one and the same as biological organisms of a certain kind (van Inwagen 1990; Olson 1997; Snowdon 2014).

The view has ancient pedigree. Aristotle took the view that we are rational animals. The view is also congenial to our modern post-Darwinian worldview: much popular biology, anthropology, sociology, and cognitive science, appears to take it for granted that we people are simply animals, of a uniquely psychologically sophisticated kind, but animals nonetheless. It would be an exaggeration to call animalism 'pre-theoretically intuitive'—it's interesting to note that developmental studies show that young children are rather slow to extend a general *animal* concept to human beings<sup>1</sup>—but for many of us nowadays, the claim that we are animals has the

<sup>1</sup>See Hermann, Medin, and Waxman 2011 for recent developmental discussion with further refer-

ring of common-sense.

However, there are, in principle, competing ways of locating ourselves within the class of biological individuals. Biological individuals occur at both higher and lower mereological levels than the stereotypical bounded and physiologically unified vertebrate organism. On the one hand, a *group* of organisms can be counted as a single, albeit scattered, biological individual e.g. a herd of horses, a cloud of aphids, or, on one understanding, a species.<sup>2</sup> On the other hand, the articulated limbs, systems, organs, cells, of an individual organism can each be counted as biological individuals in their own right. These too are countable, non-universal, particulars, belonging to the biological world.<sup>3</sup>

Now the idea that *I* might be a group of physiologically unified organisms is not one that has been seriously discussed in the metaphysics literature, and I do not intend to discuss it here. The converse idea, however, that *I* might be some biological individual smaller than the physiological organism is one that has recently been endorsed by Derek Parfit, among others. He has made a serious case for the view that we are ‘embodied parts’: ‘We are ... the conscious, thinking, and controlling parts of human beings’ (2012, 14)—roughly the *brains* of human organisms.

On this *Brain View* the remaining, complementary, parts of the whole human organism stand to us as we might ordinarily think of the clothes we are wearing, or an automobile in which we are seated: we have intimate control over this housing structure, but our boundaries are, strictly speaking, less extensive.

For the purposes of this paper I want to set aside other possible views of our nature—such as the dualist view that we are immaterial souls—and focus on the question of how to adjudicate between these two views of which biological individuals we are: the *Embodied Part View* that we are brains, and the *Animalist* view that we are whole organisms. How, if at all, are we to decide between these views?

It seems to me that the accommodating and context-sensitive character of ordinary language renders the appeal to what we would *ordinarily say* about ourselves of limited help here. For example, the view that we are brains is sometimes dismissed quickly on the basis that it is intuitive that one has arms and legs, intuitive that we can see each other, and that one can see oneself in the mirror. Yet the forms of ownership expressible by possessives like ‘has’ are very varied: think of the different statements one could make from context to context with the words ‘I have a horse’ or ‘this horse is mine’—that one is its legal owner, its jockey, that one has captured it, bet one’s money on it, and so on. The insistence that the satisfaction of ‘having arms and legs’ made in ordinary circumstances strictly requires that arms and legs stand to one in the relation of proper part-hood in particular, rather than any other of a range of proprietorial relations (cf. ‘I have a flat tyre’) imputes to ordinary language a determinacy it is unlikely to possess.

Appeals to ordinary judgments about *visibility* seem equally inconclusive: under ordinary circumstances one can truly be said to see a riot policeman even if no part of his skin is visible, or see a meteor in the sky even if only its ambient envelope of hot gases is emitting light. Satisfaction of ordinary uses of the term ‘see’ does not always require that seen objects have exposed surfaces.

So how to proceed? We take a small step in the right direction by making the following,

---

ences.

<sup>2</sup>See Ghiselin 1974 and Hull 1978 for defence of the view that species are scattered individuals with organisms as parts.

<sup>3</sup>This is a relatively undemanding notion of a biological individual. More theoretically heavyweight notions, such as those requiring autonomy, or a role as unit of natural selection, would be more contentious in their application to biological groups or parts.

innocuously Cartesian observation: that whatever we are, we are subjects of mental states, events, and processes; in short, we *think*. This is introspectively manifest, and I think it makes good interpretive sense of much of the metaphysical debate about what we are. It is an inquiry into the nature of the beings that think our thoughts. So if it could be shown that animals, but not brains, think, then that would settle things in favour of the Animalist view. Conversely if it could be shown that brains but not animals think, then that would establish Parfit's Brain View.

However, one might worry that this appeal to thinking is just another ordinary expression of our self-image and no more dialectically effective than the statement 'I have arms and legs'. Just as one might worry that the satisfaction of ordinary uses of 'having arms and legs' is too permissive to discriminate between strict proper part-hood and other intimate relations of attachment, sensation and control, and hence too permissive to discriminate between the brain view and the whole organism view, so one might worry that ordinary statements that one 'thinks a thought,' or 'is a subject of an experience' are equally indiscriminating. Perhaps *both* the organism and the brain can truly be said to think one's thoughts or be the subjects of one's experiences.

In fact, the hypothesis that both of these biological individuals think gets support from a more theoretical regimentation of our ordinary thought and talk about the mind. According to *functionalism* as standardly developed, our ordinary thought and talk about the mind implicitly defines mental states to be parts of a complex causal structure, mediating between inputs, outputs, and other mental states of the same system.<sup>4</sup>

On the face of it, both animal and brain are systems containing structures with the sort of complex causal role in mediating between input and output and other states of the system that is implicitly specified by our ordinary thought and talk about mental states. There is of course a physical difference between the two systems: the most proximate inputs and outputs to the brain are afferent and efferent nerve signals, whereas the most proximate inputs and outputs to the animal are slightly more distal: stimulations of sense organs and movements of the body. But why should this matter? It is plausible that the systems instantiate the same abstract causal structure even if it is implemented, or realized, in these slightly different ways. On a functionalist view of the mind, then, these two nested systems should each have mental properties.

What consequence would the truth of this 'two-thinkers hypothesis' have for the philosophical question to which Animalism and the Brain View are each answers—the question *what am I?*

Arguably it would render the pursuit of the question, at least as it stands, fruitless. It could do this in either of two ways depending on how first-person thought works. Here are two models.

**Model 1:** If  $x$  thinks a first-person thought  $t$ , then  $x$  thereby refers to the unique thinker of  $t$ , if  $x$  refers to anything.

**Model 2:** If  $x$  thinks a first-person thought, then  $x$  thereby refers to  $x$ .

Each model captures the datum that one cannot think first-personally about anyone but oneself.

Now suppose in accordance with the two-thinkers hypothesis that when the question is voiced 'what am I?' the whole organism  $a$  and the embodied part  $b$  both think a first-person thought  $t$ . According to Model 1, neither would succeed in thereby referring to anything, for there is no unique thinker of  $t$ . The question, then, would lack content. According to Model 2 there would be no failure of reference but a double-success: in thinking  $t$ ,  $a$  thereby refers to  $a$ , and in thinking  $t$ ,  $b$  thereby refers to  $b$ . The shared first-person thought would be a vehicle for each thinker to inquire about itself. However, there would, I believe, remain an *epistemic*

<sup>4</sup>See Lewis 1972; Shoemaker 1984.

form of indeterminacy. Neither thinker would be able to know which of the two thinkers it is. Any evidence putatively discriminating between the hypothesis that one is the embodied part and the hypothesis that one is the whole organism would be presented to the one thinker if and only if it is presented to the other thinker. So neither thinker could possibly know the answer to the question it manages, according to the second model, to ask of itself. On either model of first-person thought, then, pursuit of the opening philosophical question would be fruitless.

The issue of multiple thinkers, then, appears to be a serious obstacle in the road to answering to the question of what I am. In the next section of the paper I turn to Parfit's recent strategy for dealing with this kind of multiple-thinkers problem on behalf of the Brain View. His discussion is problematic but his strategy has a certain general shape which can in principle be instantiated in different specific ways. In the final section of the paper I propose a more promising specific implementation of the general strategy, one which appeals to self-awareness. This promises to resolve matters in favour of Animalism, or at least a variant thesis in the close vicinity of Animalism.

## 2

In his defence of the Brain View, Parfit in fact endorses the two-thinkers hypothesis. He takes the view that both the brain and the whole animal think. However, he appeals to a distinction between *derivative* and *non-derivative* property possession in order to support the conclusion that we are not whole animals but the smaller parts. What is this distinction? And, in the light of the apparent sceptical consequences of the hypothesis that there are two thinkers, how is the distinction supposed to support Parfit's definite answer to the question *what am I?*

Parfit introduces the general idea of derivative property possession by analogy with some other properties of the animal:

Animals digest their food by having a part, their stomach, that does the digesting. Animals sneeze by having a part, their nose, that does the sneezing .... Human animals think, we can similarly claim, by having a part that does the thinking. (2012, 14–15)

Further into his discussion Parfit uses the terminology of 'direct' and 'indirect' thinking: contrasting 'the conscious, thinking, controlling part of this animal, which is directly thinking this thought' with 'the animal that is indirectly thinking this thought, by having a part ... that does the thinking' (21). Elsewhere other language is used: the part alone thinks in 'a strict, or first-rate, sense' not in a 'loose, second-rate, sense' (25); and the part thinks 'non-derivatively' whereas the whole animal thinks only 'in a derivative way' (26).

We can attempt to make sense of these claims. Let us say that an object  $x$  is *derivatively*  $F$  just in case  $x$  is  $F$  in virtue of the fact that some distinct  $y$  is  $F$ —and  $x$  is *non-derivatively*  $F$  just in case  $x$  is  $F$  and  $x$  is not derivatively  $F$ . The 'in virtue of' here is intended to express non-causal explanation, or grounding. For example, when I am in Paris, Paris spatially contains me in virtue of a certain smaller human-shaped region  $r$  nested within Paris containing me, the smaller region containing me in a more basic way. So Paris, France, and Europe, each contain me derivatively whereas  $r$  contains me non-derivatively. Another example: a tomato is visible to me in virtue of the fact that the tomato's facing surface is visible to me. The tomato is derivatively visible to me whereas the tomato's facing surface is visible to me non-derivatively.<sup>5</sup> In both cases

<sup>5</sup>This is how Frank Jackson distinguishes 'immediate' and 'mediate' objects of perception (1977: 19–20).

there is a basic way in which something is  $F$  (containing me by being wholly filled by me, or being seen by fixing my visual phenomenal character) and then some relation to another thing (inclusion of one spatial region by another, or parthood in the case of the tomato and its surface) which together ground the fact that the related thing is also  $F$ .

It seems to me fair to understand Parfit as claiming that the whole animal is in this sense a derivative thinker whereas the smaller part of the animal is a non-derivative thinker. There is some basic way in which the part of the organism thinks, and then in virtue of standing in a relation of mereological inclusion to this thinker, the organism also thinks.<sup>6</sup>

The next question is whether these claims about thinking are plausible. Suppose, along the functionalist lines suggested above, that being a thinker is a matter of instantiating the causal structure implicitly specified by our ordinary thought and talk about the mind. On this supposition it is plausible that both brain and animal think. But there is arguably an asymmetrical dependence here. The animal could not instantiate this causal structure without having a smaller part that also instantiates the causal structure. The converse seems not to hold: the brain could, in principle—with nourishment and some very fortunate stimulation—instantiate this causal role *in vitro*, in the absence of the animal.<sup>7</sup> We can also suppose that the brain's instantiating the causal structure does not depend in turn upon some still smaller part of the organism instantiating the causal structure. The brain is the mereologically minimal realizer of this structure.

So we seem to have the one-way non-causal dependence needed to make sense of Parfit's claims that a part of the organism is the non-derivative thinker, while the organism is a derivative thinker.

I now want to make a couple of critical points about Parfit's strategy. First, observe that the kind of mereological explanation just given of the animal instantiating the relevant causal structure, in terms of its proper part instantiating the causal structure, is only one kind of non-causal 'in virtue of' explanation. Another kind of non-causal explanation is *teleological* explanation. Ask *why* does the animal contain a part that instantiates the relevant causal structure? It seems reasonable to say that the part instantiates the relevant causal structure only in order that the *animal* may instantiate that causal structure—so that the animal may act, perceive, problem-solve, forage, socialize, and so on. It is not implausible that the evolutionary explanation of the presence of the complex brain in the human organism will turn on the claim that it serves the survival and reproductive interests of the organism for the organism to be a sentient agent. From this teleological point of view, it is a mere side effect of the organism instantiating the causal structure that its part instantiates that structure too. There may be a sense, then, in which the brain thinks in virtue of the organism thinking. That is to say, on a certain understanding of non-causal dependence, it is the organism not the brain that is the primary or non-derivative thinker, contrary to Parfit's view.

Obviously there's a lot more to be said about this. My limited point for present purposes is just that the notion of a derivative thinker inherits complexity from the complexity of the notion of non-causal dependence upon which it relies.

This brings me to my second critical point about Parfit's strategy for answering, in the face of the problem of multiple thinkers, the question *what am I?* As described so far his strategy

<sup>6</sup>Persson (2004, 399–400) and Dainton (2007, Ch 7) make similar claims.

<sup>7</sup>I assume here, along with most parties to the debate about animalism, that the animal would indeed be absent in this brain-in-a-vat scenario (though it is not actually obvious to me that animalists should concede this. See Madden 2016 for discussion). Note also that it is not being denied that the brain's instantiating the structure ordinarily *causally* depends on the presence of the rest of the animal, for oxygen, nutrition, etc.

is incomplete. For all that has been said, there remains indeterminacy, semantic or epistemic, in the answer to our question. There are still two thinkers after all. In order to complete the strategy Parfit needs an additional claim about first-person thought, that first-person thought refers only to the *non-derivative* thinker. That is to say, a variant of the first of the two models of first-person thought is needed:

**Model 3:** If  $x$  thinks a first-person thought  $t$ , then  $x$  thereby refers to the unique *non-derivative* thinker of  $t$ , if  $x$  refers to anything.

If first-person thought refers only to the non-derivative thinker, then, while it is true that there are two thinkers, there may yet be a definite answer to the question *what am I?*

Is it plausible that the character of first person thought embodies a restriction to refer to non-derivative thinkers? Olson, in a piece otherwise critical of Parfit, seems to agree: ‘arguably it belongs to the content of first-person thoughts that they refer to the being who thinks them non-derivatively’ (Olson 2015, 48).

It seems implausible to me that this belongs to the content of first-person thought. To suppose that this is part of the *content* of first-person thought is to suppose that every first-person thinker has a tacit grasp of what is, as we have just seen, a rather technical notion, which can be developed in different ways, depending on which notion of non-causal dependence is chosen. It seems far-fetched to say that I would be simply *contradicting* myself if I believed that I am thinking through a relation of non-causal dependence to some distinct thing’s thinking.

Parfit, in his own discussion, actually wavers between what might be called *descriptive* and *reconstructive* proposals. At points he does seem to commit to the bold descriptive claim that as a matter of fact first-person thoughts embody a restriction to refer only to non-derivative thinkers (2012, 18–19). At other points, however, he seems inclined to agree that the reference of first-person thought is not as things stand restricted to non-derivative thinkers. Instead he ‘invites’ his reader to ‘use the word ‘I’ in the different, more precise sense’ (24) to refer explicitly to the non-derivative thinker of one’s thoughts. ‘We can then truly claim that we are not human beings in the sense that refers to animals, but are the most important parts of these animals’ (24–25).

Is the reconstructive proposal better than the descriptive proposal? *Prima facie* it seems an odd move to make. His animalist opponent might as well invite his audience to use ‘I’ to refer to the *derivative* thinker. The animalist can then also declare victory: we can truly claim that we *are* human beings in the sense that refers to *animals*, and *not* parts of these animals.

Is this a stalemate? Or is there a case to be made for making one rather than another reconstruction? Parfit does give some positive grounds for preferring his particular reconstruction of ‘I’ to refer the non-derivative thinker of one’s thoughts. This meaning, he says, ‘would have more importance’ (24). It is difficult, though, to understand why he says this. Why ought we to be so especially concerned to think about one rather than another relatum of a certain non-causal dependency relation? Is this really what we are, so to speak, *trying to get at*, in asking what we are? It is also, as I have said, unclear that such relations run in one direction only, and hence unclear that even those persuaded to adopt Parfit’s reconstruction of ‘I’ would secure a determinate answer to the question *what am I?*

I believe, however, that the general shape of Parfit’s reconstructive strategy is promising. In pursuing the question of which individuals we are in the face of the problem that more than one entity meets the conditions for thinking, we should identify some specific *way* of thinking—schematically  $\phi$ -*thinking*—such that:

1. Our real concern, when we ask what we are, is the identity of the  $\phi$ -*thinker* of our thoughts; and

2. There is a unique  $\phi$ -thinker of our thoughts.

If we can identify such a way of thinking, then we can use 'I' to refer to the unique  $\phi$ -thinker of our thoughts, and in so doing, reinstate a determinate, and hopefully tractable, question, *what am I?*

As we have seen, thinking defined in a standard functionalist way will not do. No doubt we are concerned to identify something that instantiates the causal structure implicitly defined by our ordinary talk about the mind; the problem is that there is no such unique functional thinker. *Non-derivative* thinking, as we have just seen, is also problematic: the notion of non-causal derivativeness is generic. Until it is made more specific it is not clear that condition (2) is met, that there is a unique non-derivative thinker (perhaps it is true both that the organism is the teleologically primary thinker and that some smaller part of the organism is the mereologically minimal thinker). On the other hand, once specified in one of these more technical ways, condition (1) seems even more unlikely: why should we be especially interested with the non-derivative thinker in that specific sense?

I would like to make a proposal at this point. Our concern in asking which thing we are is not simply a concern about which thing is thinking these thoughts, and having these experiences. It is a concern with which thing is thinking these thoughts, and having these experiences, in a *self-aware* way. We are concerned to identify the thing which, in having one's thoughts and experiences, is thereby immediately aware of itself. I propose, that is, that *self-aware thinking* meets condition (1). I conjecture that this gives a neutral point of agreement between parties to the metaphysical debate about what we are. If parties to the debate were told that a number of entities think one's thoughts, in a range of different senses, but that only one of these entities is thereby self-aware, I conjecture that that the latter entity would be, so to speak, the intended object of their philosophical inquiry into the identity of the self. So it would not be inept or question-begging in the context of this debate to reconstruct 'I' 'more precisely' to refer to the unique *self-aware* thinker of these thoughts (if it does not already do so). To acknowledge that this reconstruction may be more precise or determinate in meaning than the ordinary pronoun 'I,' we can use the notation 'I\*' to explicitly indicate thoughts conforming to the following model:

**Model 4:** If  $x$  thinks a first-person\* thought  $t$ , then  $x$  thereby refers to the unique *self-aware* thinker of  $t$ , if  $x$  refers to anything.

But now the big question: what about condition (2)? Can it be argued that there is a *unique* self-aware thinker of one's thoughts? If there are multiple thinkers of one's thoughts, then why are there not also multiple self-aware thinkers of one's thoughts? What's so special about self-awareness?

In the remainder of the paper I want to defend this proposal. More specifically I will argue, after a couple of false-starts, that there is a way forward for Animalism, or more precisely for the view, which we can call Animalism\*, that I\* am an animal. It is arguable that even if there are multiple thinkers, the unique self-aware thinker is an animal.

### 3

How should we understand self-awareness? A simple proposal is that self-awareness is the awareness, the presentation to the mind, of the object in the world that is identical to oneself.

But this simple proposal is surely inadequate. If my attention happens to alight upon the back of one individual in a crowd of people displayed on a CCTV monitor, and that individual

happens to be me, then while there is a sense in which I am in fact presented with the object that is identical to myself, there seems a more important sense in I am not thereby self-aware.

A better proposal is that  $x$  is self-aware just in case  $x$  is presented with  $x$  through reflexive channels. A reflexive channel is a channel such that  $x$  is presented with  $x$  if  $x$  is presented with anything at all through that channel. Looking up at a CCTV screen is not in this sense a reflexive channel: it is a channel through which  $x$  can be presented with things without *ipso facto* being presented with  $x$ .

So which channels are reflexive channels? A traditional place to begin looking is in the sources of introspective knowledge, knowledge of one's own mental properties. Perhaps there is some plausibility to, as well as tradition behind, the idea that simply through introspection of our stream of consciousness we are presented with ourselves. Even if one is not experienced as an object quite like ordinary objects of sensory awareness, one is, in introspection, in some sense just as phenomenologically present or given to consciousness as any other experienced object.

But I think those hoping for a determinate answer to the question of what I am should be wary of this line of thought. It's a very difficult question by what means exactly we are aware of our own conscious states and events, but the following seems plausible, a sufficient condition for introspective access. Simply in *having*, or *undergoing*, conscious states and events, one is thereby in a position to know about them.<sup>8</sup>

But if that sufficient condition for introspective access holds, then, once it is granted—as it is being granted here—that the animal and brain *each* have, or undergo, conscious states and events, then it seems correspondingly plausible that they are each thereby in a position to know about those conscious states and events. So if introspection is a reflexive channel then it is a reflexive channel for *both* animal and brain. But that means there is no more a unique self-aware thinker than there is a unique thinker. No progress has been made.

Can we deny the claim that introspection of the stream of consciousness is a way of being presented with oneself? There is some phenomenological support for Humean scepticism about self-awareness here. It is plausible to claim, as many empiricist philosophers have done, that a subject  $x$  who is introspecting his stream of thought and consciousness is not thereby presented with  $x$  at all.<sup>9</sup> The various elements constituting the stream of consciousness are individually accessible, but the *subject* of the stream is entirely elusive. Scrupulous reports of introspective awareness ought to take Lichtenbergian form, like 'there is experience occurring' or 'there is thinking going on.'

Arguably the present assumption that there are multiple thinkers reinforces this Humean scepticism. On this assumption, we should think of a conscious event as 'housed' within multiple thinkers simultaneously. Simple being introspectively aware of the conscious event or state itself does not distinguish any one of these housing thinkers—any more than merely scrutinizing a snooker ball distinguishes the room in which the ball is located from the still larger building in which both ball and room are located.

But, of course, this Humean scepticism is not yet a happy stopping-point. If it is true that introspection is never a way of being presented with oneself then indeed there are not thereby multiple thinkers self-presented through introspection. But a determinate answer to our question needs it to be true that there is *one* self-presented thinker.

I can see no way for the Brain View to argue that the brain is uniquely presented with

<sup>8</sup>Which is not to say that one need actually make any introspective judgements on a given occasion, nor that introspective judgements one does make will be infallible: they will not be immune from misclassification, inattention, wishful thinking, and other general causes of error in judgement.

<sup>9</sup>See Mach 1897, Carnap 1928, and Russell 1914 for different expressions of this Humean point.

itself. But I think the opposing Animalist\* has two potential ways forward, which will repay scrutiny. First, the Animalist\* might argue that Humean scepticism about introspection is not generally correct; introspection of mental properties *is* a way of being self-presented, but, for certain reasons, *only* the animal is self-presented in introspection. A second way forward is to argue that the animal is uniquely self-presented, but the enabling reflexive channels are *non-introspective* channels.

The first way forward is suggested by some claims made by Michael Ayers and Bill Brewer. The context of their claims is in both cases a defense of animalism against Cartesian dualist alternatives. But if what they say is correct, then it should have force in the present context, as an argument that the animal is uniquely self-aware via introspection of certain mental properties. Here is Ayers:

The physical or material self ... is the *presented* subject of experience and action. The experience of pain, for example, since pain is necessarily in some part of the body, presents the bodily self from within as the subject of pain as surely as the sensation of colour presents some object or area from without as the subject of colour. Something similar is true of every bodily sensation; and, since all other forms of sensory awareness are necessarily integrated with awareness of the body, it is true of all sensory experience whatsoever. (1991, 233)

And here is Brewer arguing along similar lines:

In bodily awareness, one is aware of determinately spatially located properties of the body which are also necessarily properties of the basic subject of that very awareness. In contrast with external sense perception, a psychological property of oneself is physically located in or on the body. Therefore, rather than any mere possession, the animal body is the conscious mental subject of bodily awareness. (1995, 17)

The proposal, I think, is that some conscious events are not 'inner' items, whose housing subject is inscrutable to introspection. For bodily sensations—and if Ayers is right, for *all* forms of sensory experience—introspective awareness of these conscious events will present the 'animal body' as their subject. So, for the animal, introspection of at least some mental properties is a way for the animal to be reflexively aware of itself. For the brain, in contrast, introspection of sensations is not a way for the brain to be reflexively aware of itself; sensations are not *in* the brain in the sense that they are in the animal body. In being aware of sensations, the brain is only aware of certain mental properties of the *animal*, not of itself. So it is not introspectively *self*-aware in the same way. The animal is the unique self-aware thinker.

Is this convincing? Insofar as I can understand the reasoning, it seems to confuse 'act' and 'object' aspects of bodily sensation. On the one hand there is the mental property of the conscious subject of awareness: the *state* of awareness, of, say, a pain in one's foot. On the other hand, there is the *object* of that state of awareness, in this case a pain in one's foot, or the property-instance of hurting that is possessed by one's foot. Following C. D. Broad's terminology, we can say that while the former is a potential object of *introspection*—our way of knowing our own states of mind—the latter is a potential object of *inspection*—in this case attentive scrutiny by means of bodily-awareness.<sup>10</sup> Now the objects of inspection do distinguish the animal from the brain: they are located in the animal body, and in that sense the animal is their subject, or locus, in a way in which the brain is not. But to argue that the animal is uniquely *introspectively* aware, it would need to be shown that only the animal has access to states of *awareness* of these sensory

<sup>10</sup>Broad 1925, Ch 6.

objects. It is difficult to see how that could be shown. It is plausible that introspection of the act proceeds via inspection of the object of the act, as numerous writers on the ‘transparency’ or ‘diaphaneity’ of awareness have urged. But it does not follow that for  $x$  to introspect  $x$ ’s state of awareness via inspection of objects of sensation located in  $y$ ,  $x$  must be numerically identical to  $y$ . Why couldn’t the brain introspect its own states of body-awareness by inspecting sensations that are located beyond its boundaries in its housing animal superstructure?

I know of no more successful way of arguing that introspection of states of mind is a reflexive channel for the animal only. So I think the Animalist\* is better off cleaving to the Humean line that introspection of states of mind is not a reflexive-channel for *any* subject. Introspection of conscious states and events doesn’t present any subject to itself.

So that leaves the second way forward, which, to recall, is to argue that there are *non-introspective* reflexive channels, for the animal only. Can it be argued that the animal is uniquely self-aware by means of non-introspective channels?

I’ll begin by considering an attempt by Paul Snowdon in his recent book to reply briefly to the sort of worry about multiple thinkers which I’m addressing here.

If it had been established that in deciding my identity, I am restricted to relying on something like the information that I am a thinker then the admission of a multiplicity of thinkers would create a problem about evidence. However, no obvious case has been made for that restriction. So for all that has been said I can appeal to my knowledge that I am six feet tall, weigh sixteen stone, and look unkempt. This knowledge enables me to legitimately discriminate myself from, say, my brain and my head, neither of which has those properties. I know enough to enable me to know that I am not these things. This means that the current worry is without bite (2014, 247)

What we see here is exactly an attempt to exploit self-awareness of *non-mental properties* rather than mental properties to privilege the animal. However, the details of what Snowdon says here seem problematic; he seems to underrate the difficulty of the problem. Why? Let’s ignore, as Snowdon does, the concern that multiple thinkers might cause reference-failure of ‘I’; let’s suppose that each thinker can refer to itself. That still leaves something like a dilemma for Snowdon’s reply. First, as already remarked, it is likely that there some flexibility or accommodation in what counts as satisfaction of self-ascribed physical predicates. In a familiar context the sentence ‘I cannot fit into that parking space’ says something true just in case a certain object with which I am *related*—my car—cannot fit into that parking space. I can’t infer from the felicity of the statement that I myself am literally too large to fit in the space. In the same way, it is not at all clear that simple appeal to elements of one’s physical self-image such as ‘I weigh sixteen stone’ will discriminate one from the brain quite as neatly as Snowdon suggests.

On the other hand, let us grant that ascription of something like ‘\_\_\_ weighs sixteen stone’ *does* discriminate between animal and brain, in the sense that it is just false when ascribed to the brain. The problem now is that it becomes very unclear whether the animal can be assumed to possess *knowledge* that it weighs sixteen stone. Why? It is plausible that knowledge requires that the basis for belief is such that a mistake could not easily have been made on that basis.<sup>11</sup> However, given the subjective similarity of the animal’s situation and the brain’s situation there is an intuitive sense in which each will be forming the belief that it weighs sixteen stone on the same kind of basis. So even if the animal gets it right in thinking that it weighs sixteen stone,

<sup>11</sup>This is the basic idea of currently popular ‘safety’ driven accounts of knowledge. See Williamson 2000.

there is a sense in which it is *lucky*. Error could easily have occurred; there is a very nearby situation, that of the brain, in which that same kind of basis gives rise to false belief. So it is not clear that the animal really does *know* that it weighs sixteen stone and *a fortiori* unclear that it can use the knowledge to discriminate itself from the brain.

Now a full development of this line of objection would require excursion into general questions about epistemology and scepticism. But I think it is right to say that Snowdon's reply is too quick to be convincing, at least as it stands.

However, the general idea, that some kind of awareness of one's physical properties, rather than introspective knowledge of one's mind, could distinguish oneself from one's brain, can perhaps be executed in a better way, or so I hope to suggest in the remainder of this paper.

My recommendation is to locate reflexive channels among the monitoring systems of proprioception, kinesthesia, the sense of balance, visual self-location (and indeed *inspection* of bodily sensations). For surely a case can be made that the evolved function of these sensory channels is to monitor the state of the whole organism. That's what these systems are for. The object presented to consciousness by these sensory systems is the whole organism. If that case can be made, then it would follow that there is after all a unique  $x$  such that  $x$  is presented with  $x$  through reflexive channels. The organism is that thing. The brain, in sharing all the thoughts and stream of consciousness of the organism, is also presented with something through these sensory systems, but that thing is still the organism. So for the brain these are not reflexive channels: they present to the brain an object numerically distinct from the brain. On the view of self-awareness as a matter of possessing reflexive channels of awareness, then, it is the organism, uniquely, which is self-aware. The brain, even if it is in some sense a thinker—even if it is in some sense the non-derivative thinker—is not self-aware. So we have a definite answer to the question 'what am I\*?', if not to the original naïve question 'what am I?'<sup>12</sup>

In order to develop the proposal slightly further, I'll close by anticipating an objection, an objection in the spirit of the 'predicate flexibility' horn of the above dilemma for Snowdon.

The organism-monitoring channels, goes the objection, can equivalently be conceptualized as *brain*-monitoring channels. Therefore, they can be conceptualized as reflexive channels for the brain too. So actually we still do not have a unique self-aware thinker. In more detail: whenever the organism-monitoring systems carry the information that the organism's legs are crossed, or it is unbalanced, or has a sore foot, they will *ipso facto* carry information about certain *relational* properties of the brain, such as *being the brain of an organism with crossed legs*, or *being the brain of an organism that is unbalanced*, or *being the brain of an organism with a sore foot*. In effect the same sensory apparatus simultaneously realizes two sets of reflexive channels: one delivers information about a set of properties that uniquely characterize the animal; the other delivers information about a slightly different set of (relational) properties, which uniquely characterize the brain. Each thinker, then, has a channel such that  $x$  is aware of  $x$  through that channel if  $x$  is aware of anything through that channel. So each thinker is self-aware.

Now I think the way for the Animalist\* to respond to this objection is to dispute its rather quick transition from 'monitoring,' or information-carrying, to *awareness*. Here is an analogy to make the point. When I am watching Socrates I am thereby getting information about a certain set of properties of his wife Xanthippe. When Socrates sits, I get information that Xanthippe now has a husband who sits. When Socrates stands I get information that Xanthippe has a husband who now stands. I am able to monitor dynamically a wide range of *relational* properties of Xanthippe by monitoring Socrates. But it seems to me that a situation in which

<sup>12</sup>Notice that we have here a twist on the old complaint that functionalism cannot account for subjectivity: of two functionally equivalent systems, only one is self-aware.

one is gaining information solely of relational properties of a thing is not a situation in which one is thereby *aware* of the thing, in the intuitive sense of having the thing presented to one's consciousness here-and-now. As Shoemaker observes in a discussion of perception, there is an intuitive connection between presentational awareness of a thing and awareness of its intrinsic properties.

We can perceive relations between things we perceive; but we wouldn't perceive these things at all, and so couldn't perceive relations between them, if they didn't present themselves as having intrinsic, non-relational properties. To perceive that this book is to the right of that one I must perceive, or at least seem to perceive, intrinsic properties of the two books e.g., their colors and shapes. (Shoemaker 1994, 253)

I think what the Animalist\* should say in response to the present objection is this: while the brain may be gaining information about some of its own properties through organism-monitoring systems, this reflexive monitoring is not reflexive *awareness*, because the brain is not thereby aware of any of its *intrinsic* properties. Only the organism is aware of its intrinsic properties through the organism-monitoring systems (the disposition of its limbs, its sore foot etc.). So the organism alone has reflexive channels of *awareness*. The organism alone meets the condition of being present to its own mind.

This is the beginning of a defence of Animalism\*, the thesis that the unique self-aware thinker of these thoughts is an animal. But now finally this defence is bound to return us to the question of the satisfaction of the *first* component of the schematization of Parfit's strategy, which we are here implementing in a different way from Parfit. Suppose that *self-aware thinker* does satisfy the second component; there is a unique self-aware thinker. The question now, in the light of the foregoing clarification of what self-awareness entails, is whether the strategy captures what we were really after in asking which thing we are. Were we really interested in asking which is the thing *presented* with itself via reflexive channels (with presentation understood to require awareness of intrinsic properties)?

Perhaps, perhaps not. But if we want an *answerable* question, in roughly the right *vicinity* of the question driving the traditional inquiry into our identity, then I suspect we won't be able to do better.

## Literature cited

- Ayers, Michael. 1991. *Locke, Epistemology and Ontology*. Routledge.
- Bermudez, Jose, Naomi Eilan, and Anthony Marcel. 1995. *The Body and the Self*. Cambridge, MA: MIT Press.
- Block, Ned. 1980. *Readings in the Philosophy of Psychology*, Volumes 1 and 2. Harvard University Press.
- Brewer, Bill. 1995. "Body Awareness and the Self." In Bermudez et al. 1995, 291–310.
- Broad, C. D. 1925. *The Mind and its Place in Nature*. London: Kegan Paul.
- Carnap, Rudolf. 1928. *Der Logische Aufbau der Welt*. Berlin.
- Dainton, Barry. 2007. *The Phenomenal Self*. New York: Oxford University Press.
- Hermann, Patricia, Douglas Medin, and Sandra Waxman. 2012. "When Humans Become Animals: Development of the Animal Category in Early Childhood." *Cognition* 122 (1): 74–79.
- Hull, David. 1978. "A Matter of Individuality." *Philosophy of Science* 45: 335–360.

- Ghiselin, Michael. 1974. "A Radical Solution to the Species Problem." *Systematic Zoology* 23: 536–44.
- Jackson, Frank. 1977. *Perception*. Cambridge: Cambridge University Press.
- Lewis, David. 1972. "Psychophysical and Theoretical Identifications." In Block 1980, 207–215.
- Mach, Ernst. 1897. *Contributions to The Analysis of Sensations*. Translated by Cora May Williams. Chicago: Open Court Publishing.
- Madden, Rory. 2016. "Human Persistence." *Philosophers' Imprint* 16 (18): 1–15.
- Olson, Eric. 1997. *The Human Animal*. New York: Oxford University Press.
- . 2015. "On Parfit's View That We Are Not Human Beings." In *Mind, Self and Person*, edited by Anthony O'Hear. Royal Institute of Philosophy Supplement 76, 39–56. Cambridge University Press.
- Parfit, Derek. 2012. "We Are Not Human Beings." *Philosophy* 87: 1–28.
- Persson, Ingmar. 2004. "Self-Doubt: Why We Are Not Identical to Things of Any Kind." *Ratio* 17: 390–408.
- Russell, Bertrand. 1914. "On the Nature of Acquaintance III: Analysis of Experience." *Monist* 24 (3): 435–453.
- Shoemaker, Sydney. 1984. *Identity, Cause & Mind*. Cambridge University Press.
- . 1994. "Self-Knowledge and 'Inner Sense'." *Philosophy and Phenomenological Research* 54 (2): 249–269.
- Snowdon, Paul. 2014. *Persons, Animals, Ourselves*. Oxford: Oxford University Press.
- Van Inwagen, Peter. 1990. *Material Beings*. Ithaca: Cornell University Press.
- Williamson, Timothy. 2000. *Knowledge and its Limits*. New York: Oxford University Press.

---

© 2017 Author(s)

This is an open-access article distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license, which permits anyone to download, copy, distribute, or display the full text without asking for permission, provided that the creator(s) are given full credit, no derivative works are created, and the work is not used for commercial purposes.

ISSN 2475-3025