

## THINKING PARTS<sup>1</sup>

Abstract: Undetached proper parts of human organisms generate a sceptical threat to the naïve thesis that we have human form. This paper introduces some new ways of dealing with the sceptical threat.

1. Introduction
  - 1.1. The Naïve Thesis
  - 1.2. A Sceptical Threat to the Naïve Thesis
  - 1.3. The Plan
2. Epistemic Solutions
  - 2.1. An Evidential Externalist Strategy
  - 2.2 Linguistic and Thought Theoretic Strategies
  - 2.3 Self-Acquaintance and Self-Monitoring
3. Psychological Solutions
  - 3.1. Spatial Parts and Causal Parts
  - 3.2 Maximality and Functional Parts
  - 3.3. Natural Function and the Mind
4. Remnant Persons
  - 4.1. The Dilemma for Animalists
  - 4.2. Creation and Exposure
  - 4.3. Continued Participation
5. Conclusion

### 1. Introduction

---

<sup>1</sup> Ancestors of this paper were presented at CEU Budapest, Cambridge, and York. I should like to thank the audiences on those occasions for their comments and questions. Special thanks to Nick Jones, Mike Martin, Lucy O'Brien, Paul Snowdon, and the other participants of a stimulating research seminar at UCL.

Empirical enquiry into the natural world can throw up results that seem to cast doubt on elements of our naïve image of our place in the world. Sophisticated contemporary physics is a source of results of this general sort. But philosophers have sometimes seen our naïve conception of things threatened by rather more modest empirical findings. Hume, for example, took it that knowledge of the occurrence of simple perceptual illusions was a sufficient basis on which to refute the naïve view that in perception we are directly acquainted with external objects. Naïve realism about perception has also been criticized on the basis of another empirical commonplace, that proximal stimulation of one's nervous system can bring about sensory experience indistinguishable from the perception of distal external objects.

In this paper I want to investigate the question of whether a modest empirical truth about the relationship between experience and the central nervous system should cast doubt upon another element of our naïve image of our place in the world.

### 1.1. The Naïve Thesis

We naïvely strike one another as things of a certain, very familiar, form. We seem to be 'humanoid'. We extend through space from head to toe, from fingertip to fingertip. We are bounded by sensitive skin. We are visible to one another, and able to come into contact with one another. In short: we have human form. Let us call the thesis that we have human form *the naïve thesis*, articulating as it does part of the way things naïvely strike us.

How does the naïve thesis relate to animalism? Animalism is the philosophical thesis that we are each fundamentally individuals of a certain biological species (*homo sapiens*).<sup>2</sup> Animalism is stronger than the simple claim that we have human form. In claiming that we are fundamentally things of the same general category as other biological organisms, animalism employs relatively sophisticated notions of *biological organism*, and *fundamental kind*. Therefore animalism goes beyond a theoretically unprejudiced articulation of how things naïvely strike us. Animalism does, however, entail the naïve view, and extends it in

---

<sup>2</sup> See Snowdon 1990. Van Inwagen 1990. Ayers 1991. Olson 1997a.

a very simple way to accommodate the discovery that we coincide with humanoid organisms, things sharing a fundamental nature and origin with other organisms.

But the naïve thesis is not the sole preserve of animalism. The naïve thesis is common ground between animalism and its principal rival in the recent personal identity debate, constitutionalism. Constitutionalist take a different view of our relation to the humanoid organisms with which we coincide, claiming that we are distinct from but ‘constituted’ by such things.<sup>3</sup> Constitutionalism also entails the naïve thesis that we have human form, because the constitution relation is understood in such a way that we share the form of the (clearly humanoid) organisms that constitute us, just as the statue shares the form of the lump of clay which constitutes it. Since the naïve thesis is a consequence of both animalism and constitutionalism, if an empirical commonplace poses a threat to the naïve thesis, then it poses a threat to both of these leading views in the personal identity debate.

## 1.2 A Sceptical Threat to the Naïve Thesis

What is the empirical commonplace that poses a threat to the naïve thesis? It is a finding that has emerged from naturalistic investigation of the subjective states we ascribe to ourselves and to others. What has emerged is that the local activity of certain *proper parts* or *sub-systems* of a humanoid entity is sufficient for the presence of a conscious perspective of the sort one now has. The condition of more peripheral parts of a humanoid has no necessary role in bringing about a conscious perspective of this sort. For example it turns out that an amputee with merely ‘phantom’ limbs will have a perspective subjectively indistinguishable from one’s own, just as long as the local activity of its central parts matches the local activity of one’s own central parts.

What are these ‘central parts’? The relevant parts of the humanoid coincide roughly with what we crudely visualize as ‘the brain’, although perhaps they comprise some smaller part of the central nervous system, and perhaps they comprise some larger part of the central nervous system. From time to time slightly different parts are likely to play the role of guaranteeing a conscious perspective of the sort one has at that time. What is

---

<sup>3</sup> See Shoemaker 1999. Baker 2000.

empirically extremely plausible is just this: that the presence of a conscious perspective of the sort one has right now is guaranteed by the local activity of parts significantly smaller than the whole humanoid. For schematic convenience let us call these parts the *thinking-parts* or *T-parts* of the humanoid. What has emerged from naturalistic enquiry is that to possess a conscious perspective of the sort one has now, it is enough to ‘have’ or in some sense ‘contain’ T-parts in the right local state.

How does this empirical commonplace threaten the naïve thesis? It threatens it in the following way. Since it turns out that the T-parts are smaller than the whole humanoid, it turns out that the whole humanoid is not the only thing that contains T-parts in a certain local state of activity. Corresponding to an indefinite number of divisions of the humanoid which contains the T-parts, we can pick out an indefinite series of objects, each of which overlap, and in that sense contain, the T-parts. Let us call these objects *overlappers*. The undetached head of the humanoid is an example of an overrapper that is salient to us, striking us as an object possessed of some degree of causal unity of its own, its parts moving together rigidly and to some extent independently of the rest of the humanoid. But we seem quite able to select for attention overlappers that do not usually attract our gaze: the ‘upper-half’ of a humanoid for example, or all of a humanoid but for its left leg. As a limiting case we might consider the T-parts themselves, hidden within the humanoid. These overlappers each lack some peripheral parts that the whole humanoid possesses. But, as the possibility of amputees in various states of mutilation reveals, a lack of peripheral parts cannot by itself prevent something from possessing a conscious perspective indistinguishable from one’s own; it is enough to ‘have’ or ‘contain’ T-parts in a certain local state of activity. But the overlappers no less than the whole humanoid meet this condition. So, one might worry, each must have a conscious perspective indistinguishable from one’s own.

If this is so, then we have on our hands a dramatic sceptical challenge to the naïve thesis that we have human form. If the overlappers possess a point of view subjectively indistinguishable from one’s own, then if one is presented with any evidence for any proposition, the overlappers will each seem to themselves to be presented with evidence of the same kind. Suppose that one has the best possible evidence that one could possibly have for the naïve thesis that one has human form. Then the overlappers will each seem to be presented with evidence of the same kind. But the vast majority of

them—such as the undetached head—do not even approach human form. So when they believe themselves to have human form, on the basis of what seems to them to be the best possible evidence, they will be in error. So, the sceptic will argue, one's basis for believing that one has human form is unsafe. Even if one does in fact have human form, one's basis for believing this is a basis which in the vast majority of cases leads to error. Knowledge cannot have a basis of this kind.

So the empirical commonplace that the activity of only relatively small parts of a humanoid is needed to guarantee the presence of a conscious perspective threatens the knowability of the naïve thesis that we have human form. On mild assumptions about the closure of knowledge under entailment, this in turn threatens the knowability of both animalism and constitutionalism.

Moreover there is a threat to a more general epistemological thesis, the thesis that one may come to know which particular thing in the world one is. Why? Consider identity statements of the form  $I = x$ , where  $x$  ranges over the things overlapping one's T-parts. On the naturalistic assumption that one is conscious only if one does in fact overlap T-parts—that one is not for example an unextended soul, or a single tiny material atom—then one of these statements is true. But for any statement to the effect that one is identical with a particular overlapper, one's belief that one is identical with that overlapper is unsafe even if true—for whatever basis one has for that belief is a basis that leads to error in the vast majority of cases in which a believer relies upon that basis. Therefore the truth about one's identity is unknowable. One's identity is a verification-transcendent fact.<sup>4</sup>

### 1.3 The Plan

Some animalists have explicitly addressed worries about thinking proper parts of animals but they have tended to be drawn towards the *eliminativist* position of denying the very existence of large undetached proper parts of animals.<sup>5</sup> I do not wish to dismiss this eliminativist position lightly, as it can be to some extent motivated by independent general concerns about material constitution and the conditions under which

---

<sup>4</sup> See Madden 2010 for a somewhat different argument that first-person reference can generate a divergence between what is true and what is verifiable.

<sup>5</sup> Van Inwagen 1981, 1990, Olson 1995, 1997b, Merricks 2003.

composition occurs. But for the purposes of this paper I shall proceed on the assumption that the eliminativist position is at best an unattractive last resort. The thesis that there is no such thing as an undetached head, or brain, or finger, or eye, runs counter to *semantic* appearances, and also to *perceptual* appearances.

First, we seem to have the ability to coin a syntactically singular term with the overriding intention to refer to one such thing—‘let us call this attached finger ‘Freddy’—and then go on to use an atomic sentence containing that term to literally express a truth—‘Freddy is now bent’. If these semantic appearances are correct then ‘Freddy’ achieves singular reference, and we may infer that Freddy exists. If Freddy does not exist then these semantic appearances must be mistaken in some way.

Second, objects such as attached fingers engage our primitive visual capacity to sense collections of low cardinality (1, 2, and 3), a capacity apparently present in even very young children and animals. Arguably the exercise of this ‘numerosity sense’ constitutes our most basic acquaintance with the cardinal numbers.<sup>6</sup> If, as eliminativism insists, a visible ‘three raised fingers’ does not really count for three objects, then our pre-linguistic capacity to sense cardinality is simply misfiring. Eliminativists have sometimes sought to sweeten their pill by means of charitable paraphrases of ordinary language.<sup>7</sup> But this effort is beside the point when the conflict lies with pre-linguistic perceptual appearances.

My aim in this paper is to introduce, and to some extent assess, some new non-eliminativist ways of defending our naïve conception of ourselves against the sceptical threat. The aim is to show that we can know that we are possessed of human form, without having to deny the existence of undetached heads and the rest. Eric Olson has helpfully drawn a distinction here between ‘psychological’ and ‘epistemic’ species of non-eliminativist solution. A psychological solution would show that ‘brains, heads, and other spatial parts of human organisms cannot think, or at least not in the way that you and I can’ (Olson 2007: 217). An epistemic solution in contrast would accept that the parts in question ‘think our thoughts’ (ibid.) but show that nevertheless we ‘somehow know that we are thinking animals rather than thinking heads or brains’ (218). All versions of these

---

<sup>6</sup> See Giaquinto 2001 for argument for this thesis, and references to relevant empirical studies.

<sup>7</sup> Van Inwagen 1990.

solutions that Olson surveys he finds unpersuasive, and for this reason considers the problem of thinking proper parts of animals—rightly in my opinion—to be a challenge to animalism ‘considerably more troubling’ (216) than the more familiar ones based on counter-intuitions about our persistence in science-fiction scenarios, such as brain-transplantation, tele-transportation, or bionic replacement.

The plan is as follows. In §2 I shall introduce two new epistemic solutions: an evidential-externalist strategy, and a thought-theoretic strategy. The last in particular seems to be worthy of further investigation, although as things stand it is hostage to the resolution of some difficult issues about first-person thought. I shall then move on in §3 to look more closely at the reasons for supposing in the first place that overlappers have a conscious perspective, recommending for further research a *function* based approach to distinguishing overlappers from humanoids. It will be suggested that this kind of psychological solution is more promising than the more common ‘maximality’ approach to such problems. However, the recommended approach does not by itself refute a recent ‘remnant person’ argument due to Johnston and Parfit, an argument which aims to show that animalists in particular must face the sceptical threat of at least one conscious proper part of the humanoid. I finish the paper in §4 by explaining how animalists can reject the key presupposition of that argument.

## 2. Epistemic Solutions

A sceptical scenario threatening one’s putative knowledge that  $p$  is a case meeting two conditions, a ‘subjectivity’ condition and a ‘failure’ condition. Designating the subject of a given case by ‘one’ and whichever case you are now actually in as ‘this case’, a sceptical scenario is a case in which (a) things seem to one subjectively no different from this case, and yet (b) one is mistaken in believing  $p$ . For example, a classic brain-in-a-vat case threatening one’s putative knowledge that one has hands is a case in which (a) things seem to one subjectively no different from this case, and yet (b) one is mistaken in believing that one has hands. As introduced above, the cases of overlappers of T-parts seem to include many cases that meet both the subjectivity and failure conditions for a sceptical scenario threatening one’s putative knowledge that one has human form. The

overlapper cases would include many in which (a) things seem to one no different from this case, and yet (b) one is mistaken in believing that one has human form.

Now, one way of responding to an alleged sceptical scenario would be to try to show that the case does not really meet the subjectivity condition (a)—to show that the case is not really one in which things seem no different from this case. What Olson calls a ‘psychological solution’ to the problem of thinking parts in effect takes this line, in denying that overlappers think at all. But this is not the usual way in which philosophers have tried to respond to arguments by sceptical hypothesis. It is rare to see the heroic denial that a brain-in-a-vat case could really be one in which things seemed no different from this case. It is far more common to try to find other ways to meet the challenge. In this part of the paper I shall introduce some new epistemic solutions, an evidential externalist strategy, and thought-theoretic strategy, neither of which deny that overlappers have subjective perspectives.

### 2.1. An Evidential Externalist Strategy

An evidential-externalist strategy does not deny that the case of a T-parts overlapper would seem subjectively no different from this case. It also does not deny that the T-parts overlapper would mistakenly believe that it has human form. What the evidential-externalist denies is that this creates an obstacle to one’s knowledge that one has human form. The sceptic assumes that it follows from the fact that each overlapper’s case is subjectively indistinguishable from one’s own case that the evidential basis each has for believing that it has human form is the same as one’s own evidential basis for believing that one has human form. If that assumption is correct, then, since these countless beings would be mistaken to think that they have human form, the evidential basis one actually has is a basis that leads to error in a vast majority of cases. So one does not know that one has human form.

The evidential-externalist challenges the sceptic’s assumption. What follows from the subjective indistinguishability, which we are supposing to be guaranteed by their possession of T-parts in the same state as one’s own, is, in the first place, that the overlappers will each *seem* to possess evidence of the sort that one now possesses. It



simply does not follow from this alone that they all *do* possess evidence of the sort that one now possesses. So it does not follow that the evidence of the sort that one now possesses leads to error in most cases.

But what evidence could one possibly possess which the overlappers do not all possess? Here is a simple example: I know that I have hands. This is a piece of knowledge which I have gained on the basis of perfectly ordinary empirical channels of visual and proprioceptive self-knowledge. It is one part of my basis for believing that I have human form. But it is not evidence which most of the overlappers possess. Why not? They do not possess this evidence because it is not even *true* that most of the overlappers have hands. *A fortiori* they are not in possession of the knowledge that they have hands. Thus there is an evidential asymmetry between my case and the case of a handless overlapper. Moreover I can exploit this evidential asymmetry, to discriminate my case from the case of every handless overlapper. The evidence I possess—that I have hands—is evidence which clearly settles that I am not a handless overlapper.

A serious assessment of this kind of reply to the sceptic would require a general discussion of the nature of evidence, which cannot be pursued here.<sup>8</sup> What I can do is indicate a couple of points about its application to the problem of thinking parts specifically, and a general reason why it is likely to be unsatisfactory.

First, it is worth noting a *prima facie* dialectical worry about the evidential-externalist reply for animalists in particular. The problem of thinking parts is potentially especially embarrassing for animalists because it parallels the following ‘problem of the thinking animal’, which animalists have used to motivate their own position against their constitutionalist rivals.<sup>9</sup> It is plausible that there is a human animal where one is right now, and that such a human animal, like other advanced primates, can think. But if so, challenges the animalist, how could one possibly tell that one is distinct from this animal, even if one is in fact distinct from it? The *prima facie* worry for animalists now is that any sound response to the problem of thinking parts will *mutatis mutandis* furnish a reply to this problem of the thinking animal, which animalists have hoped to use against the constitutionalist view that we are distinct from animals.

---

<sup>8</sup> Williamson (2000: Chs 8 and 9) argues that there is an evidential asymmetry between ‘good’ and ‘bad’ cases, based on the thesis that one’s evidence is precisely what one knows.

<sup>9</sup> Snowdon 1990. Ayers 1991. Olson 1997a.

In this case, I think, such a dialectical worry would be unfounded. An evidential-externalist reply to the thinking-animal problem on behalf of the constitutionalist looks relatively unpromising. Why? Simply because there is no ordinary piece of self-knowledge—analogue to the perceptually-based knowledge that one has hands—which constitutionalists could claim that we, but not thinking animals, possess, and on the basis of which to rule out the epistemic possibility that one is the thinking animal. One’s ordinarily empirically accessible properties match those of the animal far too closely. A constitutionalist evidential-externalist reply to the problem of the thinking animal could only be based on intuitions about one’s possible divergence from the animal in certain science-fiction cases, these being the only respects in which the person and the animal are supposed to fail to match. But, surely, such intuitions do not have the secure status of one’s belief that one has hands. The latter is a deliverance of ordinary channels of empirical knowledge. There is a ready answer the question of how one knows that one has hands. There is no ready answer to the question of how one knows that one would diverge from the animal in certain science-fiction cases. It is hardly a ‘Moorean’ foundation upon which to eliminate the possibility that one is the thinking animal. Therefore an evidential externalist reply to the problem of the thinking animal looks much weaker than an evidential externalist reply to the problem of thinking parts.

Here is another clarification of the evidential-externalist reply to the problem of thinking parts. The knowledge that one has hands would allow one to discriminate one’s case from every case of a *handless* overlapper, such as an undetached head. But there are many overlappers which do in fact have hands, lacking instead some other peripheral parts (the ‘complement’ of my left leg, for example). The knowledge that one has hands, then, is not by itself a sufficient basis to discriminate oneself from these subjects. In order to know that one has human form one would need to amass evidence which discriminates one’s case from every case of an overlapper who in some way falls short of human form. However, it seems plausible that one can amass such evidence. If one can come to know in an ordinary way that one has hands, then one can know in an ordinary way that one has a left leg, and so on for other peripheral parts.<sup>10</sup>

---

<sup>10</sup> Note that I shall not in this paper consider putative overlappers which differ from a whole humanoid only microscopically. The possibility of such overlappers raises the spectre of Unger’s (1980) ‘Problem of the Many’, and this requires completely separate treatment—for example along the following lines: individual macroscopic continuants are not related one-to-one to

The principal concern about the externalist-evidentialist reply is a concern one might have about any such response to the sceptic, the concern that it takes a somehow *insouciant* attitude to the sceptical challenge. Even if one does in fact have evidence that the overlappers do not possess, evidence which settles that one is the whole humanoid distinct from the various T-parts overlappers, it is difficult not to feel that in the intellectual context of answering the sceptical challenge one does not have the right simply to rely on that evidence. Even if one does in fact possess the evidence, in relying upon it to answer the sceptic one would be taking for granted something which is in dispute. One would in some sense be begging the question.

It is very hard to tell whether in the end this is a reasonable worry about evidential-externalism. On the other hand, neither does there seem to be any straightforward way to lay the worry to rest. So I submit that an externalist-evidentialist reply to the problem of thinking parts is likely to leave us feeling unsatisfied.

## 2.2. Linguistic and Thought Theoretic Strategies

An evidential-externalist reply does not deny that the case of an overlapper meets the subjectivity and failure conditions (a) and (b). Before we move on to ‘psychological solutions’ which deny that an overlapper case meets the subjective similarity condition (a), it is interesting to ask whether there might be any epistemic solutions which agree to (a) while denying that such a case would be a case meeting the failure condition (b), that one would be in error. We can find in the scepticism literature anti-sceptical replies of this general sort. For example the semantic externalist reply inspired by Putnam’s (1981: Ch 1) discussion of brains in a vat can be seen as denying that a brain-in-a-vat case meets the failure condition. The idea is that although the brain-in-a-vat’s case is in some sense subjectively just like this case, the brain-in-a-vat would *not* mistakenly believe that it has hands. Due to causal constraints on reference, a brain-in-a-vat would be in no position to refer to hands at all. So it would not have a false belief that it has hands. If it has a

---

precise collections of microscopic parts, but are instead each composed of many slightly different precise collections of microscopic parts. These many collections do not correspond to many near-coincident macroscopic objects, but to one macroscopic object multiply micro-constituted. Jones (2010) explores this promising approach.

belief at all, it will be a true belief, to the effect that it has ‘hands-in-the-image’ or ‘virtual hands’.

The only kind of epistemic response to the problem of thinking parts of which I am aware in the literature in effect makes a similar move. It tries to deny that overlapper cases meet condition (b) of a sceptical scenario threatening one’s putative knowledge that one has human form. Olson suggests (though does not endorse) a linguistic proposal, according to which personal pronouns such as ‘I’ are governed by the convention of referring to whole human organisms and not their undetached parts.<sup>11</sup> If this proposal is correct then when the overlappers use ‘I’ they each refer to the whole human organism. So in fact the overlappers each speak the truth when they say ‘I have human form’ since they each refer with the pronoun ‘I’ to the whole human organism, which does indeed have human form. An overlapper’s case is not then a case of error after all.

Such a proposal is problematic in two respects. First, the key linguistic hypothesis does not seem plausible. There is no evident reason to think that our use of the first-person pronoun is governed by a convention of referring to any particular kind of object. It is improbable that the standing meaning of a word such as ‘I’ embodies any restriction on the kind of thing to which it may refer. (This is surely one source of the difficulty of the personal identity debate). It is much more plausible that ‘I’ and its cognates, and synonymous words in other languages, are governed by the convention of self-reference, or reflexive reference. They are words by means of which one may intentionally perform the linguistic act any  $x$  performs when  $x$  refers to  $x$ . If it is conceded that undetached parts have the general capacities to conceive, intend, and execute linguistic acts, then it is hard to deny that they could conceive, intend and execute the act of self-reference by use of ‘I’.<sup>12</sup>

A second problem with the linguistic hypothesis derives from the distinction between language and thought. In the first instance the linguistic hypothesis would establish that the conscious overlapping parts would each utter a true sentence when they uttered the English sentence ‘I have human form’. But it is not clear how this could establish that they do not nevertheless each have a mistaken first-personal, or ‘de se’, belief to the

---

<sup>11</sup> Olson 2007: 218. Zimmerman considers a parallel reply to an analogous objection to perdurantism about the persistence of persons (2003: 501-3)

<sup>12</sup> See Madden 2011b for more on the intention to self-refer.

effect that they themselves have human form. At most the linguistic proposal would demonstrate that they would be unable to manifest such a belief with a sentence of English. This is a problem because intuitively knowledge requires that the basis one employs could not easily have led to mistaken belief. The linguistic proposal does not show that one's basis could not easily have led to mistaken belief, only that it could not easily have caused the production of a false sentence.<sup>13</sup>

The second problem with the linguistic hypothesis suggests that a more satisfying epistemic solution might question the capacity of the overlappers to *think* about themselves first-personally, to single themselves out self-consciously in thought. If an overlapper cannot single itself out self-consciously in thought then it does not meet the failure condition (b) for a sceptical scenario: the overlapper would not even mistakenly believe that it has human form.

An immediate concern about the plausibility of such an 'thought-theoretic' solution is that it would fall foul of an equivalent to the first objection to the linguistic proposal: in so far as it is plausible that the word 'I' is a device of reflexive linguistic reference, it is just as plausible that a first-person thought is means for reflexive thought: in grasping a first-person thought  $x$  thinks of  $x$ . How could overlappers think first-person thoughts without thereby thinking of themselves?

At this point it is helpful to note that there is a shape of solution to the problem of thinking parts that is not clearly classified by Olson's contrast between 'psychological' and 'epistemic' solutions. Such solutions are distinguished by their contrasting attitudes to the claim that overlappers 'can think our thoughts'. But that claim has a strong reading and a weak reading. The strong claim is that overlappers can think *all* of the thoughts that we think. The weak claim is that overlappers can think *some* of the thoughts that we think. A solution could agree to the weak claim while denying the strong claim: a solution could argue that we can know that we are not overlappers on the basis that overlappers think some but not all the thoughts that we undoubtedly do think.

---

<sup>13</sup> See Kearns and Magidor 2008 for more against the idea that plausible counterfactual safety conditions on knowledge can be stated meta-linguistically, in terms of nearby possibilities of truth and falsity of sentences.

It is not quite clear whether Olson would classify such a solution as ‘psychological’ or ‘epistemic’. But whatever the terminology, it is in exactly this way that we should understand the thought-theoretic analogue of the linguistic epistemic solution. The claim is that we can know that we are not overlappers because there is one kind of thought—first person thought—that we can think but they cannot think. This forestalls the objection on the basis that first-person thoughts are reflexive thoughts. The proposal is not that overlappers think first-person thoughts but somehow fail to think of themselves, either by thinking of something distinct from themselves, or by thinking of nothing. That would be impossible: first-person thoughts are guaranteed reflexive thoughts if they are anything. The proposal is rather that overlappers do not think first-person thoughts at all.

### 2.3 Self-Acquaintance and Self-Monitoring

To defend such a thought-theoretic proposal one should need to locate some plausible necessary condition of the possibility of thinking a first person thought—and then show that overlappers fail to meet that condition. A distinguished proposal for a necessary condition on first person thought is this:  $x$  thinks a first-person thought only if  $x$  is *self-acquainted*. Some such proposal is found in the writings of philosophers from Descartes to Frege, and much more recently Kripke, who asserts that first-personal understanding requires ‘a special acquaintanceship with oneself ... more fundamental than anything purely linguistic’. (2011: 391). We can think first-personally only ‘because each of us is acquainted with her/himself in a special first person way’ (2008: 214).

This proposal is no clearer than the notion of self-acquaintance. Although Kripke is confident that ‘everyone knows what this type of acquaintance is’ (2008: 215) and ‘there is nothing mysterious about this’ (2008: 214) it would be useful in the present context to say more.

In general one is acquainted with an object only if one has a standing sensitivity to its states and doings, as, for example, one does when an object is immediately present to one’s senses. So on one natural filling out of the notion of self-acquaintance, one is self-acquainted only if one has a standing sensitivity to the states and doings of the object

that is oneself. It is plausible that such sensitivity is mediated by various distinctively first-personal channels of self-awareness: one monitors one's bodily states from the inside, one is visually aware of one's own location in space relative to other objects, and so on. By these various channels one counts as acquainted with the object that is oneself.<sup>14</sup>

But is it plausible to insist that  $x$  can think a first-person thought only if  $x$  is self-acquainted in this sense? One might here raise the sort of objection that Anscombe made to the proposal that the word 'I' could be explained as a demonstrative meaning *this body*.<sup>15</sup> She imagines someone struck blind, physically anaesthetized, floating in a sensory-deprivation tank of tepid water. It seems intuitive to say that such a subject would remain able to think first-personal thoughts. But the sort of channels of self-monitoring gestured at above are all shut down. So—one might think—self-acquaintance cannot be a necessary condition for first-person thought.

But this objection merely invites clarification of the notion of acquaintance. It can be true that one is acquainted with an object even if one has gaps in one's episodes of awareness of the object. This is obviously the case with one's acquaintance with people and places. So long as one had initial exposure to the object, and one retains standing sensitivity to its states and doings, then one counts as acquainted with the thing. It is only after a long period of separation from a person or place that it can no longer truly be said that one is acquainted with the thing in question. So the kind of temporary blockage of self-monitoring channels envisaged by Anscombe is hardly decisive evidence against the proposal that singling oneself out first-personally requires self-acquaintance.

All this might be agreed. But how could any of this demonstrate that overlappers are incapable of thinking of themselves first-personally? Let us focus on the element of bodily self-monitoring. It would be immensely implausible to claim that an amputee who is under an illusion of his bodily extent, having a merely phantom left leg for example, is failing to self-monitor, and failing to meet the self-acquaintance condition on first person thought. Acquaintance with an object is surely consistent with a certain degree of

---

<sup>14</sup> Cassam 1997 focuses on bodily awareness in particular as foundational of first person thought, but there is no evident reason to give special weight to any one of these various channels of self-monitoring.

<sup>15</sup> Anscombe 1975: 31.

distortion or illusion in one's episodes of awareness of the object. But if this is so, then how could it be claimed that the overlapper who shares all of the parts of the humanoid but for the left leg fails to meet the self-acquaintance condition? Doesn't the overlapper, just like the amputee, have merely distorted bodily awareness of itself; but bodily self-awareness nonetheless?

Why does this seem plausible? In the next section we shall look at reasons for denying that overlappers have any subjective perspective whatsoever. But the current reply is not taking such a bold stance. It is accepting that the overlapper is in a sensory state similar to one's own. So it has bodily sensations similar to one's own. Since the overlapper's bodily sensations are precisely as reliably causally correlated with the state of its body parts as the bodily sensations of a phantom-limb hallucinating amputee are correlated with *its* body parts, it is tempting to think that the overlapper has bodily awareness of itself just in so far as the amputee has bodily awareness itself. The amputee surely does have bodily awareness of itself. Therefore the overlapper too must have bodily awareness of itself.

But this picture of bodily self-awareness is questionable. Reliably caused sensation is not in general sufficient for awareness. Suppose, as a result of stroke perhaps, I come to have a certain colorful visual sensation reliably causally correlated with the aroma of coffee in my environment. It does not follow that I am thereby visually aware of the aroma of coffee. I cannot now *see* the aroma of coffee. Why? A plausible answer is that the *function* of the human visual system is to select for attention things other than aromas. Even if the aroma reliably causes my visual sensations, the visual system is still malfunctioning. It is failing in its function of selecting one of its proper objects. That is why I am not seeing an aroma, but instead simply undergoing an unusually caused visual hallucination.<sup>16</sup>

So what is the function of body awareness? On one view its function is to select and monitor multiple body-parts, just as the function of vision is to select and monitor multiple objects in one's external environment. But this is not the only account of the function of body-awareness. One might instead hold that the function of the human

---

<sup>16</sup> See Davies 1983 for more on the role of function in distinguishing deviant and non-deviant reliable casual chains in perception. See also Burge 2010: part III on the importance of natural function in understanding perception.



somaesthetic system is to monitor the state of the whole organism housing the system. The object of awareness is thus the whole organism. Although various objects may be reliably causally correlated with one's body-sensations, the only object the human somaesthetic system presents to consciousness is the whole organism, and presents individual parts only to the extent that they are presented as parts of that whole entity.<sup>17</sup>

Is there a reason to prefer either of these views? There is, at least, a purely phenomenological reason to prefer the single-object view: apparent awareness of sensations in a pair of body parts is *eo ipso* awareness of sensations seeming falling within one's own single boundary. If the primary objects of body sense were just the individual body parts, then the phenomenology should in itself be neutral over whether a pair of body parts in which one felt sensation were parts of a single whole, or parts of distinct wholes. But the phenomenology does not seem to be neutral about this. Indeed if body-awareness did leave open which whole or wholes the various presented parts belonged to, it is hard to see how it could be the immediate source of first-person knowledge that it seems to be.

If instead we take the view that the function of bodily awareness is to make the whole organism present to consciousness and parts only insofar as they belong to that whole organism, then it is unsurprising that no further information could be needed to ensure that a pair of body parts of which one is aware belong to one rather than two wholes. The somaesthetic system would at best be malfunctioning if a part causing one's body sensations did not belong to the whole organism. One would not be perceiving the part in question, but simply undergoing a hallucination caused by that part. The somaesthetic system furnishes awareness of a part only if it belongs to the whole organism containing the somaesthetic system. Thus if one is aware of two body parts at all, then one is aware of two parts of the whole organism.

How does this bear upon the question of whether the overlappers are self-acquainted? An overlapper is a proper part of a whole organism. An amputee is not a proper part of a whole organism—it is itself a whole organism, albeit one shrunken. Both contain a human somaesthetic system, which, it has been argued, presents the whole organism to

---

<sup>17</sup> The distinction between these single-object and multiple-object views of body awareness, and the argument following, is indebted to Martin 1997.

consciousness, and parts only insofar as they seem to be parts of that whole. So, although the causal correlation of the overlapper's body sensations with the overlapper's parts is just like the causal correlation of the amputee's body sensations with the amputee's parts, the amputee's body sense gives it a distorted presentation of itself, whereas the overlapper's body sense gives it only a non-distorted presentation of the whole organism distinct from itself. Thus the overlapper is not self-aware through body sense, despite the fact that it has similar sensory states to something which is self-aware through body-sense. In contrast to the predicament of Anscombe's subject in the sensory deprivation tank, this is not a temporary gap in episodes of self-awareness consistent with a standing state of self-acquaintance. The overlapper never has been, and never will be, presented to itself in body sense.

If thinking a first person thought requires self-acquaintance, and the overlappers are not self-acquainted, then the overlappers cannot think first person thoughts, however similar their sensory states are to our own. They are rather like unselfconscious animals, contained within us. This is the thought-theoretic epistemic solution to the sceptical challenge. The case of an overlapper is not a sceptical scenario after all. It is not after all a subjectively indistinguishable case in which one mistakenly thinks that one has human form. It is a case in which one has no first-person thoughts whatsoever.

Now, a full assessment of this thought-theoretic epistemic solution to the problem of thinking parts would need to assess whether parallel arguments can establish that channels of self-monitoring other than body-sense—such as the proprioceptive aspect of vision—fail in the case of overlappers to furnish them with the self-acquaintance required for first person thought. In each case the argument is likely to lean rather heavily on the notion of function, and it is a major question how best to understand these notions (although more will be said in the next section).

It would also need to be asked whether, even if overlappers cannot selectively self-monitor themselves as objects, mere introspective awareness of sensory consciousness might be enough to put the overlappers in a position to single themselves out first personally. Could they, for example, single themselves out as *the subject of this sensory field*, or *the subject causing this mental activity*? (Not if it is agreed that the sensory field and activity in question has many subjects.)

A full assessment of the thought-theoretic solution should also ask whether there is a sense in which this solution denies that overlapper cases meet even the ‘subjectivity condition’ (a) for being a sceptical scenario, let alone the ‘failure condition’ (b). For example, if we understand a case’s being a case in which things seem to one no different from this case as: a case in which it seems to one *that one is oneself* in such-and-such a situation, then, whatever sensory similarity there is to this case, if one cannot think first-personally then things cannot seem to one no different from this case.<sup>18</sup>

Although these are extremely interesting issues, I want to move on from this epistemic solution, for whatever happens it is likely to remain unsatisfactory in one obvious respect. It concedes that there is a multitude of subjects with sensory points of view—if not first-personal points of view—contained within my boundaries. For those interested in preserving our naïve conception of things this will not be a pleasant point at which to terminate inquiries. So it is time to look much more carefully at the reasons for supposing that overlappers of T-parts have any subjective states whatsoever.

### 3. Psychological Solutions

#### 3.1. Spatial Parts and Causal Parts

The empirical commonplace which generates the sceptical threat to the naïve thesis that we have human form is the fact that the local activity of the T-parts of a humanoid is enough to guarantee the presence of a subjective perspective similar to one’s own. The whole humanoid’s consciousness is secured by its bearing certain relations to these central T-parts. But—it might be argued—given that T-parts turn out to be smaller than the whole humanoid, many overlappers also bear these relations to the central T-parts. So the overlappers must be conscious if the humanoid is conscious.

But what are the relations which make this reasoning plausible? First consider the relation of *spatial containment*. In this respect the humanoid and overlappers are similarly

---

<sup>18</sup> See Peacocke (2008: 141) for argument that one could not seem to oneself to have the capacity for first-person thought without actually having the capacity for first-person thought.

related to the T-parts. The spatial region containing the T-parts is a sub-region of the spatial region containing the humanoid. And each overlapper bears the same relation to the T-parts: the spatial region containing the T-parts is a sub-region of the spatial region containing the overlapper.

But there is no plausibility to the idea that  $x$ 's spatial containment of T-parts is by itself sufficient for  $x$  to have a conscious perspective. If the city of London spatially contains a humanoid, which spatially contains T-parts, then the city of London also spatially contains T-parts. But the city of London is not for that reason conscious. A thing's spatial containment of T-parts is obviously not sufficient for it to have a conscious perspective.

What is the relevant difference, then, between the city of London and a thing which we do ordinarily take to be conscious, such as a humanoid? Here is *one* relevant difference. In the case of the humanoid, the activity of the T-parts it contains effects an extremely sophisticated causal co-ordination between the inputs and outputs of the system—between incoming stimuli and outgoing behaviour. The humanoid is in this sense *causally coordinated* by the T-parts it spatially contains. In contrast London is not causally coordinated by the T-parts of any particular humanoid it spatially contains. The T-parts of a humanoid, although spatially contained in London, do not have that kind of relation to the inputs and outputs of London (whatever they might be).

The principal reason that an overlapper might seem to be a much more serious candidate for having a conscious perspective is that, unlike London, an overlapper spatially contains T-parts that *do* causally coordinate the inputs and outputs of the overlapper, and in a way that exactly mirrors their causal coordination of the inputs and outputs of the whole humanoid. There is of course a certain difference between the overlapper and an ordinary whole humanoid. The inputs into the boundary of some of the overlappers will be more proximal to the T-parts than the inputs into the larger humanoid, and the outputs exiting the boundary of some of the overlappers will be more proximal to the T-parts than the outputs from the whole humanoid. But it is hard to believe that this difference could in itself make a difference to whether an overlapper has a conscious perspective. After all the local activity of the T-parts of an amputee causally coordinate

between inputs and outputs which are more proximal to the T-parts than those of a humanoid of normal form. But an amputee is still conscious.

If it is enough for  $x$  to be a conscious subject that  $x$  has T-parts which causally coordinate in a sufficiently sophisticated way between the inputs of  $x$  and the outputs of  $x$ , then we have to admit that an overlapper is a conscious subject. But is it enough? I want to suggest a line of thought according to which, while a humanoid and an overlapper match in respect of their spatial and causal relations to the T-parts they overlap, something's bearing spatial and causal relations to T-parts is *not* in fact sufficient for it to be a conscious subject. In order to begin to see how, we can consider an analogy.

### 3.2. Maximality and Functional Parts

What is an *automobile*? Here is a first attempt at an answer:  $x$  is an automobile iff  $x$  contains parts—principally wheels coupled to an internal engine—which together have the causal capacity to move  $x$  around as a whole. Now, here is a *prima facie* problem for this first attempt at an answer. Select for attention the thing which is the whole automobile minus its four doors: call this thing the *sub-car*. The problem is that the sub-car is also an  $x$  such that  $x$  contains parts, wheels and internal engine, which cause  $x$  as a whole to move around. For whenever the wheels and engine move the car around, they inevitably move the sub-car around as well.

Does that mean that we have to accept that what we naively took to be the case of a single automobile driving around is really a case of a multitude of automobiles nestled within one another? After all, there is an indefinite number of large undetached parts of a car that lack only 'peripheral' parts—that overlap the wheels and engine, thus each containing parts which move it around.

In reaction to this sort of question some contemporary metaphysicians are drawn to the claim that a property like the property of being an automobile is a *maximal* property: no large proper part of an automobile is itself an automobile.<sup>19</sup> If we conjoin the first

---

<sup>19</sup> Sider 2003. Burke 2004.

attempt at saying what an automobile is with the maximality clause: ‘... and  $x$  is not itself a large proper part of an automobile’ then we will get the intuitive result that there are no additional automobiles on the scene. The sub-car will not count as an automobile in addition to the large automobile.

But in addition to a certain lack of independent motivation, the problem with this maximality requirement is that it seems not to be true. We can quite easily imagine a kind of automobile which has a quick-escape ‘mini-car’ built into its structure, designed to detach and drive away in emergencies. That kind of automobile *does* have an automobile as a large proper part. Of course automobiles do not usually have smaller automobiles as large proper parts. But it seems not to be a necessary truth about automobiles that this must be so. Some kinds of automobiles do, and some kinds of automobile don’t, have automobiles as large proper parts.

So how can we diagnose the difference between the sub-car and the mini-car? Each is an  $x$  such that  $x$  has parts, wheels and internal engine, which have the causal capacity to move  $x$  around as a whole. A natural answer is that the mini-car is *designed and built* in just that way. The engine and wheels of the whole which is the mini-car were put together *to* have the joint capacity to move the mini-car around. The parts of the sub-car were not put together to move the sub-car around. They will cause that to happen, but only as a side effect of their assigned function, which is to move the whole car around.

This suggests a way of improving on the first attempt at saying what an automobile is, and without attaching an implausible maximality clause:  $x$  is an automobile iff  $x$  contains parts, wheels and engine, which have the *function* of moving  $x$  around as a whole. The reason that a sub-car is not a car is not that no large proper part of a car can itself be car; a mini-car is a large proper part of a car. The reason that a sub-car is not a car is that the parts of the sub-car do not have the function of moving the sub-car around, although they have the causal capacity to move the sub-car around. They were put together in order to move the whole car around. They are spatial and causal parts of the automobile and sub-car alike, but they are *functional* parts of the automobile alone.

Now we can return to the case of the putative conscious overlappers of the humanoid. Just as a maximality requirement on being an automobile would exclude the various

additional sub-cars, a ‘maximality’ clause on possession of a conscious perspective would exclude the overlappers from being conscious subjects in addition to the humanoid. But again, a maximality requirement on being a conscious subject is no more plausible than a maximality requirement on being an automobile.

Consider the following example. Although the matter is of course controversial, no philosopher has ever given a convincing reason to believe that a sufficiently fast and well-organized ‘Chinese nation’ of interacting human beings could not amount to a system with a conscious perspective of its own. Even if in the end they do find a reason, it is hardly likely to concern the mere *size* of one of its component people. For suppose that just one of the members of the Chinese nation swells up to colossal size, so that he becomes a spatially large proper part of the whole system. As long as this growth did not disturb the fast and integrated functioning of the system as a whole, it is no less plausible that the system as a whole would be conscious. Why should size matter? In that situation a large proper part of a conscious subject would itself be a conscious subject.<sup>20</sup> The moral, again, seems to be this: even if some kinds of conscious system do not have conscious subjects as large proper parts, others might well do so. The nature of consciousness *as such* seems not to settle the matter.

Now since a functional version of the criterion for an automobile did better than the original causal criterion with a maximality clause added, the obvious next question is this: can a functional version of the criterion for a conscious subject do better than the causal criterion with a maximality clause added? The functional version is as follows:  $x$  is a conscious subject iff  $x$  has parts whose *function* it is to causally coordinate in a complex way the inputs of  $x$  and the outputs of  $x$ . Just as the functional criterion for an automobile gives a principled way of solving the ‘too-many-cars’ problem, this functional criterion promises a principled psychological solution to problem of thinking parts. We might hope to distinguish by its means a thinking large proper part such as the Chinese Colossus from a non-thinking large proper part such as an overlapper. The T-parts of the large proper part of the Chinese Nation have the function of coordinating the inputs of that large proper part. But the T-parts of the overlapper, while they causally coordinate the inputs and outputs of the overlapper, do not have the function of coordinating the inputs and outputs of the overlapper. They have the function of

---

<sup>20</sup> Thanks to Ian Phillips for this ‘Chinese Colossus’ example.

coordinating the inputs and outputs of the whole organism only. The T-parts are spatial and causal parts of the humanoid and overlapper alike, but they are functional parts of the humanoid only.

### 3.3. Natural Function and The Mind

This approach, it seems to me, is extremely promising. Its detailed development will be a substantial project. But it is worth indicating in a preliminary why it has motivation quite independent of the personal identity debate.

First note that the analogy with the treatment of automobiles seems immediately to run up against the point that while the purposes, aims, and intentions of a designer can isolate, from among the various causal effects of a part within a system, the function of the part within a system, we cannot seriously appeal to the goals of a designer in distinguishing between the functions and mere effects of parts of a human organism or its overlappers.

But fortunately there has been a great deal of progress made in understanding naturalistically acceptable notions of function. For instance, aetiological theories analyse the function of a part as the effect of the part which explains the current existence of the part in the system.<sup>21</sup> There are also non-etiological theories of function, such as ‘systems-analysis’ theories, which identify the function of a part in terms of its mention in the analysis of a certain capacity of the whole system. A salient such capacity of an organism is its capacity to perpetuate itself as self-organized system over time.<sup>22</sup>

Still, one might be sceptical, not so much about the prospects for a naturalistic account of function, but about its relevance to the case at hand. Should function be mentioned at all in a criterion for distinguishing conscious and non-conscious systems?

---

<sup>21</sup> See Wright 1973, Millikan 1984, Godfrey-Smith 1994.

<sup>22</sup> Cummins 1975 is a seminal statement of the systems-approach. ‘Natural design’ theories of function claim advantages of both kinds of theory. See Kitcher 1993, Krohs 2009. We should also note that a philosopher might take the view that function is an unproblematic part of nature and yet reducible in none of these ways. See McDowell 1994, Burge 2010.



A positive answer to this question is a theme of many of the most serious attempts to understand the mental representation in naturalistically acceptable terms. A first attempt at a naturalized theory of mental representation might appeal to law-like causal correlations between the representing system and the represented feature. For example, one might take the view that a system represents *F*s just in case it has states which are related by causal inputs and causal outputs to *F*s. However, there are a number of well-known reasons to be dissatisfied with a simple causal picture of this sort. First it seems that there should be a distinction, among the many and various causal correlates of a system, between those features which are represented by the system, and those which are not. Second, a simple causal picture is not obviously able to cope with the possibility of misrepresentation, a characteristic of genuine intentional mental representation as opposed to mere causal indication.

We can illustrate these points with a tree. This is a system which has parts that ‘causally coordinate between its inputs and outputs’ in the following sense: as the years pass, its trunk produces tree rings which keep tally with the passing years. But the tree is not *representing* the passing of the years. For a start there is no reason to suppose that it would be representing any of the following in particular: solar rotations, springs, warm spells, or oscillations in the height of the sun in the sky. It is causally correlated with all of these features. And suppose that the tree began to output rings only every other year: there could be no real sense in which it the tree could now be *misrepresenting* the years elapsed. It would simply have begun to causally indicate two-year intervals instead of one-year intervals.

So what is missing for genuine mental representation? An influential answer is that the relevant parts of the tree, although they in fact causally coordinate its inputs and outputs, do not have the *function* of causally coordinating its inputs and outputs.<sup>23</sup> The causal coordination system is a mere byproduct of its disposition to springtime growth spurts. It is not in the interests of the tree to indicate the passing of the years in this way; the indication activity is not playing a role in maintaining the organism’s structure over time; the tree does not possess the system as a causal result of the fitness conferred upon the tree’s forebears; nor as a result of any process of reinforcement learning within the organism. And so on for any reasonable naturalistic account of function. In contrast,

---

<sup>23</sup> Cf. Dretske 1988, 1995, Millikan 1984. Papineau 1984.

when it is the *function* of parts of a system to coordinate between its inputs and outputs in that way, as it is with the cognitive parts of a human organism, then we may speak of the system's genuinely mentally representing some feature. For then we may understand the possibility of the system on occasion failing in its function, as misrepresenting the presence of a feature which is absent. And we may single out the genuinely represented feature from among the various causal correlates of the system, as a feature it is the system's function to represent.

So this standard naturalistic approach to mental representation suggests that a system is a mentally endowed just in case it has parts whose *function* is to causally coordinate its inputs and outputs in a sophisticated way. Thus we arrive at the functional version of the criterion for a conscious thinker, which paralleled the functional version of the criterion for an automobile.

In this light, the task for a psychological solution to the problem of thinking parts will be to argue that the T-parts do not have the function of coordinating the inputs and outputs of the overlappers, although they do have the function of coordinating the inputs and outputs of the whole organism. I can only sketch a couple of ways in such an argument might be developed.

We might think of an undetached part as a mere aggregate of matter which fills a sub-region of the organism at a time. On that understanding an undetached part can—and soon will—persist in scattered form given the metabolic processes of the organism. And suppose we take the view that the function of a part in a system is an effect which contributes to the persistence of the system over time. Now, the T-parts currently contained within an overlapper causally coordinate its inputs and outputs, just as they causally coordinate the inputs and outputs of the organism. But that coordination makes no contribution whatsoever to the persistence of the overlapper over time: the overlapper will persist whatever happens. So the T-parts do not have the function within the overlapper of coordinating its inputs and outputs. So the overlapper is not a conscious subject by the functional criterion.

But it might be more satisfying to avoid basing an argument on the premiss that overlappers are mere aggregates of matter which rapidly scatter and do not change their

constitution over time. Whatever one decides about unfamiliar objects like ‘leg complements’ or ‘undetached upper halves’ it is surely more natural to view an overlapper such as an undetached head as an object which may persist and retain its form through changes in its matter, and which—like whole human beings—gets destroyed when its matter is scattered. On this way of looking at the overlapper’s nature, it is hard to see why the workings of the T-parts don’t causally contribute to the persistence of the overlapper just as much as they do the persistence of the humanoid.

However, even on this metaphysical conception of an overlapper, one can still make the case that the T-parts have the function of coordinating the humanoid only. For given that it is the human organism and not the undetached head that is a member of a reproducing population, it is the past success in coordinating the inputs and outputs of ancestors of the organism that causally explains the current presence of the T-parts in the organism. Success in coordinating the inputs and outputs of the ‘ancestors’ of the undetached head is causally screened off as an explanation of the current presence of the T-parts in the organism. Why? Suppose, counterfactually, that ancestral T-parts had remained successful in coordinating the organism but had become unsuccessful in coordinating the undetached head—perhaps the undetached head degenerated over time into a mere twitching appendix of the organism. The T-parts would still be present in the organism today, for the coordinated organism is a reproductively fit organism propagating its traits. But the reverse is not true: if somehow T-parts had stopped successfully coordinating the organism, while continuing to coordinate the undetached head alone—perhaps the efferent nerve signals randomizing only upon exiting the boundaries of the undetached head—then the T-parts would not be present today. The organism would have been rendered utterly spasmodic and reproductively unfit. So although the T-parts causally coordinate the undetached head and causally coordinate the whole organism, it is the latter activity that explains why they are there at all. Organism coordination is their function. So by the functional criterion for a conscious subject it is the organism and not the undetached head which is the conscious subject.

This is only a sketch of an argument. The present aim is really just to make the point that undetached parts and whole organisms do *not* stand in all the same relations to the T-parts they contain. And moreover, from the perspective of leading naturalistic

approaches to mental representation, functional differences of this sort make a mental difference. So there is a principled way forward here for a psychological solution.

#### 4. Remnant Persons

But must an argument for the claim that organisms contain conscious thinking parts depend upon the general presupposition that overlappers and organisms stand in relevantly similar relations to T-parts? Perhaps not: there is an argument abroad in the recent personal identity literature that aims to pressure animalists in particular into accepting the existence of at least *one* conscious proper part of each organism. This ‘remnant person’ argument does not seem to depend on any general presupposition that overlappers are related to T-parts in the right way for consciousness. If animalists are to resist this route to the sceptical possibility of a conscious proper part, then they need to meet this argument in some other way. I shall finish the paper by explaining how.

##### 4.1 A Dilemma For Animalists

Johnston, recently joined by Parfit, uses the following scenario to force upon animalists a problem of thinking parts.<sup>24</sup>

Let us assume that a human organism can be whittled down to something which is too minimal to count as the original organism, but which is nevertheless nourished and intact enough to sustain a conscious perspective of the sort one has now.<sup>25</sup> We can visualize such a thing as a ‘cerebrum-in-a-vat’. Let us follow Johnston in calling such a thing a *remnant person*.

---

<sup>24</sup> Johnston 2007. Parfit 2012.

<sup>25</sup> It will be assumed for the sake of argument that this is possible. But it must be said, it is not at all obvious that such a small parcel of tissue, even if it sustained a conscious perspective, would sustain a conscious perspective anything like ours. Our T-parts might be larger than anything which is too small to count as a severely pruned organism.

Where did the remnant person come from? Johnston, Parfit—and Olson too<sup>26</sup>—find it incredible that one could bring a new conscious entity into being simply by cutting away surrounding ‘tissue that was not functioning to suppress consciousness’ (Johnston 2007: 47). But if the remnant person is not a new entity but a thing which was present all along, then the question arises: how did it relate to the animal before the animal was destroyed by the operation? The natural answer is that it was some undetached portion of the animal—the undetached cerebrum perhaps—only now unclothed and exposed.

But a further question now generates a difficult dilemma: was this undetached thing conscious *before* the whittling operation? If the undetached thing was conscious before the operation, then—since this could happen to any one of us—animalists have to concede that each of us contains a distinct conscious being within our boundaries right now. Thus the sceptical challenge arises: how could one possibly know that one is the whole animal rather than this conscious part of the animal?

On the other hand if the thing was not conscious before the operation, then, since it was conscious after the operation, it has to be concluded that removing matter from outside the boundaries of the contained entity has somehow bestowed consciousness upon the entity. But that does seem incredible. It seems incredible to suppose that a merely relational change, a ‘mere Cambridge’ change, could turn something from an unconscious being into a conscious being.

How should animalists react to this dilemma? They might grasp the first horn, and perhaps try to pursue evidential-externalist, or thought-theoretic, epistemic solutions of the sort considered earlier, in an effort to show that one may still know that one is the whole animal despite the presence of the conscious part. Or they might grasp the second horn, and somehow come to terms with the consequence that a mere environmental change in an entity can change it from unconscious to conscious.<sup>27</sup>

## 4.2 Creation and Exposure

---

<sup>26</sup> Olson forthcoming.

<sup>27</sup> The radically externalist perspective explored in Madden 2011a in effect grasps the second horn.

Fortunately there is a far better solution available to animalists, which is simply to reject the presupposition upon which the dilemma is based. The conscious entity resulting from the whittling operation really is a *new* entity, brought into existence by the operation. Is there any reason not to take this option?

There is, I think, a tendency among philosophers to visualize the living organism along the lines of a familiar kind of wooden block ‘anatomical toy’—as if it were a collection of interlocking wooden blocks, any one of which may be freely removed and returned. The casual discussion of brains-in-vats and brain-transplants reinforces this imagery. On this anatomy toy picture, the whittling-down operation which results in the remnant person looks metaphysically not much different from taking a pack of closely nestling snooker balls, and then potting them one by one, until finally only a single snooker ball is left exposed, an entity which undoubtedly pre-existed the process. On this way of thinking of the remnant person the difficult question then arises of whether the pre-existing entity was, or was not, conscious before the operation.

But this is not the way animalists should think about the remnant person operation. Conceding that the remnant person is not the original animal shrunk down does not leave as the only, or even the most plausible option, the view that it is a pre-existing object, which has merely been *exposed* by the operation. A more plausible view is that it is an entity *created* by the operation. After all, there is nothing in itself mysterious about the possibility of having created an entity by ‘subtraction’—by the removal of matter external to the matter which comes to compose the new entity. When the sculptor creates a wooden statuette from a block of wood he really *creates* something. The entity standing at the end of the sculptor’s labours is not an entity that was there hidden in the wood all along, only now exposed. Likewise, the filet mignon in my frying pan is not an entity that was once riding around inside the living cow. It is an entity the butcher has created, by skillfully carving from the carcass of the cow. Animalists should say something similar about the remnant person: the surgeons have carved a new entity.

Is this creation verdict much less plausible if the remnant operation involved the surgeons prising apart natural membrane boundaries between the cerebrum and the rest of the brain, carving along these pre-existing anatomical planes to some extent? No. The skillful sculptor or butcher will also do their best to carve along the grain of the

wood, or between muscle fascia—prising apart strata in the wood or meat. These are nonetheless violent processes of entity creation. No doubt there will be an unclear boundary between cases of entity creation and entity exposure. But is there any reason why animalists should not view the remnant person operation as a clear case of the former?

#### 4.3. Continued Participation

There is, one might think, a special reason for an exposure verdict over a creation verdict in the remnant case. It is a reason which is absent, or at least much less salient, in the statuette and filet cases. The reason is this. In neither the case of the statuette created from the lump of wood, nor the filet carved from the beef carcass, is there quite so obviously something that the statuette, or filet, is *now doing*, which the lump of wood, or whole carcass, was also doing before the operation. In stark contrast, the remnant person is now consciously thinking, just as the original organism was consciously thinking. Moreover the remnant is doing so in virtue of local processes in the cerebral tissue which we can imagine in principle to have gone on completely undisturbed throughout the operation, processes sufficient at each intervening time for the presence of a conscious thinker at that time.

Why does this matter? It matters because the following *continued participation principle* looks plausible: if  $x$  is the thing which is  $\phi$ -ing at a later time, and the local processes which suffice for something to be  $\phi$ -ing have gone on undisturbed since the earlier time, then  $x$  must have been  $\phi$ -ing at the earlier time.<sup>28</sup>

Applying the continued participation principle to the remnant person case, we must conclude that since a thing distinct from the original animal is thinking after the operation, and the local cerebral tissue processes which suffice for something to be consciously thinking have gone on undisturbed since before the operation, the thing distinct from the original animal must have been consciously thinking before the operation. Animalists are thus faced with the sceptical conclusion that the operation just had the effect of exposing a rival thinker.

---

<sup>28</sup> Something like this principle seems to be implicit in Johnston's presentation of the remnant person problem (2007: 46-7) and perhaps also in his much earlier anti-animalist argument, based on the metaphor that 'a person cannot be outlived by (what once was) his own mind' (1989: 77).

However, while the continued participation principle may seem fairly plausible at first sight, it has counter-examples in cases of entity-creation that are exact, and I think illuminating, analogues of the remnant person operation.

Suppose that a bush is blossoming. There are petals unfurling on just one stem of the bush. We then engage in the process of ‘plant cutting’. This is a horticultural technique of propagating plants by asexual means. Rather than growing a new plant from seed, one creates a new plant by cutting off a small amount of the parent plant. Suitably hydrated and nourished this thing will grow roots and come to flourish in the normal kind of way. It is a genetic clone of the parent plant but it is a new plant nonetheless.<sup>29</sup>

Now, suppose that we happen to take the cutting from the stem of the bush on which the petals are unfurling. Then after the operation the newly created plant will be blossoming in just the same way as the parent plant was earlier, and it will be doing so in virtue of local processes in the tissue of the unfurling petals, processes which have continued undisturbed throughout. But the thing which is blossoming after the operation is not the same as the thing which was blossoming before the operation. It is a plant which did not exist before the operation.

So the continued participation principle is simply false. One has brought a brand new ‘blossomer’ into existence, by separating it away from plant tissue ‘that was not functioning to suppress blossoming’. There is nothing conceptually puzzling about this kind of creative practice. Under these conditions a newly created entity can continue the same kind of activity as its distinct parent, even when the local tissue basis of that activity goes on undisturbed. This, I submit, is exactly how animalists should conceive of the remnant person operation. It is an example of entity creation of the same general kind as plant cutting.<sup>30</sup>

---

<sup>29</sup> I am most grateful to Elena Cagnoli Fieconi for this case. The procedure was well known to Aristotle—and was studied in detail by his successor Theophrastus, often credited as the father of botany.

<sup>30</sup> I shall leave unsettled the question of whether the remnant person deserves to be called a new ‘animal’. That term is surely polysemous: between (a) an individual of a certain genetic zoological kind, and (b) a fairly well developed example of such an individual. It would obviously be odd to call the thing an animal in sense (b)—just as it would be odd to call an early zygote an animal in that sense. In parallel it would, initially, be odd to call a brand new plant cutting in a glass of water a *bush* (as it would be odd to call a newly sprouting seed a bush). But the cutting, or sprout,



## 5. Conclusion

This paper has introduced some new ways of defending the naïve thesis that one has human form against the sceptical threat posed by the empirical commonplace that the local processes in only relatively small parts of the humanoid are needed to guarantee a subjective perspective. None of these solutions takes the usual, and quite unwelcome, animalist step of denying the very existence of undetached parts.

Among epistemic solutions, those of the evidential-externalist and linguistic variety are unlikely to be satisfying. A thought-theoretic analogue of a linguistic solution has been shown to be worthy of further investigation. More satisfying still would be a psychological solution. I have indicated why a functional approach to the conscious subject promises a much more principled solution than an appeal to maximality. Finally, I turned to the remnant person dilemma, an independent argument which threatens to force animalists in particular into accepting the existence of at least one conscious undetached part of the humanoid. We have just seen how the recognition of perfectly ordinary cases of entity creation paves the way for animalists to reject the key presupposition of that dilemma.

Our conclusion is this: the commonplace empirical fact that consciousness is guaranteed to be present by the activity of only relatively small parts of the humanoid is not yet sufficient reason—for animalists or anyone else—to doubt the naïve thesis that one has human form.

Dr Rory Madden  
Department of Philosophy  
University College London  
Gower Street  
London WC1E 6BT

---

is still a new individual of a certain bush-species. In parallel it might be argued that the remnant person is a new animal in sense (a) although of course unlikely ever to make it to animal-hood in sense (b).

## References

- Anscombe, Elizabeth (1975) The First Person. In *Collected Philosophical Papers, Vol. 2*. Minneapolis: University of Minnesota Press.
- Ayers, Michael. (1991) *Locke*, Vol. 2. London: Routledge.
- Baker, Lynne Rudder. (2000) *Persons and Bodies: A Constitution View*. Cambridge: CUP.
- Bennett, Karen and Dean Zimmerman (eds.) (2010) *Oxford Studies in Metaphysics Volume 6*
- Burge, Tyler (2010) *Origins of Objectivity*. Oxford: OUP.
- Burke, Michael (2004) 'Dion, Theon, and the many-thinkers problem'. *Analysis* 64: 242-50.
- Cassam, Quassim (1997) *Self and World*. Oxford: OUP.
- Cummins, Robert (1975) Functional Analysis. *Journal of Philosophy* 72: 741-765.
- Davies, Martin (1983) Function in Perception. *Australasian Journal of Philosophy* 61: 409-426.
- Dretske, Fred (1988) *Explaining Behaviour*. Cambridge, MA: MIT.
- (1995) *Naturalizing the Mind*. Cambridge, MA: MIT.
- Kearns, Stephen and Magidor, Ofra. (2008) Epistemicism about Vagueness and Meta-linguistic Safety. *Philosophical Perspectives* 22: 277-304.

- Giaquinto, Marcus (2001) 'Knowing Numbers'. *Journal of Philosophy* 98: 5-18.
- Godfrey-Smith, Peter (1994) 'A Modern History Theory of Functions' *Nous* 28: 344-62.
- Johnston, Mark (1989) *Human Beings*.
- (2007) "Human Beings" Revisited. In Zimmerman (ed.) 2007: 33-74.
- Jones, Nicholas K. (2010). *Too Many Cats: The Problem of the Many and the Metaphysics of Vagueness*. (PhD Thesis, Birkbeck, University of London.)
- Kitcher, Philip (1993). Function and design. *Midwest Studies in Philosophy*, 18, 379–397.
- Kripke, Saul. (2008). Frege's Theory of Sense and Reference: Some Exegetical Notes. *Theoria* 74: 181-218.
- (2011) The First Person. In his *Philosophical Troubles*. Oxford: OUP.
- Krohs, Ulrich (2009) Functions as based on a concept of general design. *Synthese* 166: 69-89.
- Loux, Michael J. and Dean Zimmerman (eds.) (2003). *Oxford Handbook of Metaphysics*. Oxford: OUP.
- Madden, Rory (2010) Could A Brain in a Vat Self-Refer? *European Journal of Philosophy*
- (2011a) Brain Transplants and Externalism in Bennett and Zimmerman (eds.) 2010: 287-316
- (2011b) Intention and The Self. *Proceedings of the Aristotelian Society* 111: 327-351.
- Martin, M.G.F. (1997) Self-Observation. *European Journal of Philosophy* 5: 119-140

- McDowell, John. (1994) *Mind and World*. Harvard.
- Merricks, T. (2003) *Objects and Persons*. New York: OUP.
- Millikan, R. (1984) *Language, Thought and Other Biological Categories*. Cambridge, MA: Bradford Books/MIT.
- Olson, Eric. (1995) Why I Have No Hands. *Theoria* 61: 182-197
- (1997a) *The Human Animal*. New York: Oxford University Press.
- (1997b) Dion's Foot. *Journal of Philosophy* 94: 260-265.
- (2007) *What Are We?* New York: OUP.
- (forthcoming) Animalism and the Remnant-Person Problem. In J. Gonçalves (ed.) *Metaphysics of the Self*
- Papineau, David (1984) Representation and Explanation. *Philosophy of Science*, 51: 550–72.
- Parfit, Derek (2012) We Are Not Human Beings. *Philosophy* 87: 5-28.
- Peacocke, Christopher (2008) *Truly Understood*. Oxford: OUP.
- Putnam, Hilary. (1981) *Reason, Truth, and History*. Cambridge: CUP.
- Shoemaker, Sydney (1999) Self, body, and coincidence. *Proceedings of the Aristotelian Society*, supp. vol. 73: 287–306.
- Sider, Ted. (2003) Maximality and Microphysical Supervenience. *Philosophical and Phenomenological Research* 66: 139-149.
- Snowdon, Paul (1990) Persons, animals, and ourselves. In Christopher Gill, ed. *The Person and the Human Mind*. Oxford: Clarendon.

Unger, Peter (1980) The Problem of the Many. *Midwest Studies in Philosophy* 5: 411-67

Van Inwagen, Peter (1981) The Doctrine of Arbitrary Undetached Parts. *Pacific Philosophical Quarterly* 62: 123-57.

———(1990) *Material Beings*. Ithaca, NY: Cornell University Press.

Williamson, Timothy (2000) *Knowledge and Its Limits*. Oxford: OUP.

Wright, Larry (1973) Functions. *Philosophical Review* 82: 139-168.

Zimmerman, Dean. (2003) Material People. In Loux, Michael J. and Dean Zimmerman (eds.) 2003.

———(ed.) (2007) *Oxford Studies in Metaphysics Volume 3*.