ORGANISMS, BRAINS AND THEIR PARTS
UB PHILOSOPHY OF BIOLOGY CONFERENCE

David B. Hershenov

I. Introduction

The brain has been described as the organ of thought. In the 18th century, Pierre Cabanis notoriously claimed that "*The brain secretes thought as the liver secretes bile*." For some reason, the 19th century materialist Karl Vogt believed the point needed to be made even more emphatically so he declared: "*The brain secretes thought as the stomach secretes gastric juice, the liver bile, and the kidneys urine*." Countless neuroscientists make claims like the mind is the brain, or the mind is the functioning brain, or the mind arises from the functioning brain. Fairly typical is a quote like the following: "Even though it is common knowledge, it never ceases to amaze me that all the richness of our mental life – all our feelings, our emotions, our thoughts, our ambitions, our love lives, our religious sentiments and even what each of us regards as his or her own intimate private self – is simply the activities of these little specks of jelly in our heads, in our brains.There is nothing else." (Ramachandram - *A Brief Tour of Human Consciousness*

If the brain produces thought, the question then arises: "Does the brain think?" It's difficult to see why it should not be able to think. But if human organisms think *and* their brains think, given that they are not identical, are there then two thinkers where we had assumed just one? How can any of us be sure that we are the thinking organism and not the thinking brain? Was this paper co-authored? Or are these pseudo problems, a result of harmless sophistry from a recently tenured academic philosopher with perhaps too much free time who might be better served by a greater teaching load and more committee work? Alas, I do not believe that the problem of a thinking brain is trivial sophistry nor "language gone on a holiday." In fact, I am a bit worried that the solution to the problem will necessitate a radical revision of commonsense ontology and scientific anatomy. After surveying the relevant alternatives and finding them wanting, I will tentatively and rather reluctantly endorse an eliminativist stance towards any organ that could be a candidate for being a thinker and any proper part of the organism that includes that organ.

II. Thinking "In Virtue" of the Brain

A common response to the problem of there allegedly being a thinking brain embedded within the thinking human organism is to claim there aren't two thinkers but that the human organism thinks *in virtue* of its brain. Yet what exactly is this *in virtue* relation and how does it help reduce the number of thinkers? It is often illustrated by analogy: it is said that we think in virtue of our brain just as we see in virtue of our eyes. Unlike our touching the floor in virtue of our feet literally touching the floor, it is not the case that eyes *literally* see. You and your eyes don't make three seeing things. However, this alleged analogy of seeing in virtue of the eyes won't rid us of an unwanted thinker. The reason why the virtues of the "in virtue" relation are less than they might initially appear is that the brain may be able to think in a self-sufficient way that distinguishes it from eyes that can't see by themselves. Or at least philosophers in their thought experiments seem to accept such a possibility and noted neuroscientists back them up. Hilary Putnam writes of brains in vats and the neuroscientist Antonio Damasio supports the possibility of such brains thinking, with some qualifications. Sydney Shoemaker hypothesizes brains being transplanted and thought continuing throughout the transplant procedures. Eric Olson, Peter van Inwagen and Trenton Merricks write of organisms being pared down to the size of their brain, and the two most prominent neuroscientists in the debate over the brain death criterion, James Bernat and Alan Shewmon, maintain that there could be brain-sized organisms. They hypothesize that thought could continue even if the rest of one's organs are removed from one's brain. Without artificial support, the thought would be very short-term, but our philosophers and neuroscientists believe the requisite support can be provided.

Some in the audience might suggest that the thinker or subject of thought in such bizarre cases is not the brain but the composite of the brain and supporting machinery. Thus the brain is not a thinker but something else thinks in virtue of contributions by the brain and other entities. Denying that the composite is the thinking entity is easier for those who maintain (like van Inwagen, Olson,

Merricks and Bernat) that the organism could become reduced to the size of a brain and yet remain a living organism. We generally are not tempted to say that a normal-sized human organism connected to high tech equipment in the hospital ICU has such equipment as parts or itself is part of a larger thinking object composed of it and the inorganic equipment. We don't say this even in extreme cases of Guillain-Barre syndrome which mimics brain death except for detectable cerebral activity. It is only the hospitalized organism that thinks, though its life depends upon entities external to it. Likewise, a brain generically dependent on the vat's machinery would still count as capable of thinking. It would be the subject of thought, not some cyborg composed of it and the high-tech machinery. Perhaps if the human organism can't be reduced in size to that of the brain and still considered an organism, or if this can't be considered a case of two organisms resulting from fission, as Shewmon and I have suggested, then a stronger case can be made for the brain not being capable of thought but merely being a vital part of a thinking being whether that thing be a full size human being or a vat/brain composite. (Incidentally, if this is so, then epistemologists should change their skeptical question to "how do we know that we are not brain/vat compounds"?)

So what can be said in favor of the brain-size organism? I will let Olson argue the case though similar discussions can be found in Shewmon and Bernat's more biologically detailed papers. Olson writes in the *Human Animal* (pp. 133-34):

> *In fact, it seems likely that the head would behave as a living organism if it were attached to a life-support system. The head would be able to regulate its metabolic rate and wake-sleep cycle. It would retain its muscle tone (even if no consciousness were present). Its pupils would open according to the amount of light hitting the retina and the lens of the eye would focus. And so on…Part of what makes something a living organism, I suggest, is its capacity to coordinate and regulate its metabolic and other vital functions. A living organism may be prevented from carrying out those functions…the instruction that my brainstem sends out to the rest of my organs may not arrive at their intended destinations, or they may nay not have the effects they are designed to have ("intended" and "designed" by evolution, not*

consciously.) Nevertheless, the control and coordination mechanisms are intact…consider what would happen if we turned off the artificial life support systems in each case. If we switched off the heart-lung machine that the detached head needs to survive, the head would behave like a dying organism. It would continue to regulate its internal metabolic activities in its characteristic way until the "control centers" shut down owing to lack of oxygen. After a few minutes death, with its characteristic symptoms – lack of coordinated activity, decay and so forth – would occur.

III. Other Undetached Thinking Parts

Even if a detached but functioning brain is an impossible philosopher's fantasy – that has also carried away some neuroscientists who perhaps should have known better - a related problem of an embedded thinking entity remains for it is undeniable that a human organism can survive the loss of many of its parts. Whatever is the smallest thinking remainder is something that earlier would seem to have been an embedded object and thus would pose the same kind of problems as did the thinking brain. So the problem will arise with arbitrary undetached parts, even if it doesn't with the brain. Moreover, it isn't that easy to dismiss such parts from our ontology. To borrow Olson's example, if there exist such things as hands, then one would expect that there also exist hand complements. Since these hand complements are composed of every part of the human being but their hands, they should be able to think. And they shouldn't lose the capacity to cognize merely by having something attached to them. Nevertheless, the puzzle is most interesting with the brain or head since these are not always considered *arbitrary* undetached parts but well established parts, respectively, of anatomy and folk ontology. So I will continue to suppose that a pared down brain can think.

IV. Maximal Thinkers

Some philosophers (such as Michael Burke) dismiss the problem of embedded thinkers by claiming it is part of the very idea of thinking that no thinker can be a proper part of another thinker. First of all, understanding "maximal" in this way fails to actually guarantee that the human organism

thinks. It only entails that if the organism did think, then it would not have a thinking part. So it doesn't rule out that a part thinks – such as the brain - and the whole does not. Of course, one can tinker with the definition of "maximal thinker" in order to make it more likely that the organism is the maximal thinker by claiming that the maximal thinker is a thinking being that has neither proper parts that think nor is a proper part of anything that thinks or fails to think. This approach would be on firmer ground if we could be confident that organisms are not parts of anything (a society with collective intentions, eco-system, universe etc.), nor could come to be a part of a thing if we were attached to something in a way that satisfies a plausible compositional principle. Unfortunately, this redefining of "maximal" seems to just be stipulating away the problem. I doubt that any explanation is forthcoming for why something can think but later loses the ability to do so merely by being attached to something else and becoming a part of that larger object. So if there could be thought when all of a person's parts but the brain are removed, why couldn't the brain think when it earlier was attached to a person's other parts, or later when similar parts are restored?

V. Little People

Thus if our brain can think, and our organism can think, there appear to be one too many thinkers. Listeners may no longer be confident that this problem could be "resolved" by denying that one's brain has the capacity for producing thought by itself and thus being a subject of thought. There are two other options that have been pursued in the literature. Some have argued that one actually doesn't possess a brain; others have identified themselves with their brains. The strategy of claiming to be brainless is likely to be met with derisive quips that the advocate of the view has uttered a self-verifying proposition. But sarcasm and incredulous stares are not the fate of just the eliminativist. Similar treatment will be bestowed upon those who identify themselves with their brain and claim that persons are but a small part of the human organism. At least, the eliminativist about brains doesn't claim we are only a few inches and pounds and have never been directly observed nor touched. While it is admittedly good news that the identification approach guarantees that none of us

are really, strictly obese or ugly, it is a shame that this has to be achieved by our neither having bodies nor faces.

Strange as the identification strategy may initially seem, there are prominent philosophers who defend it (Hud Hudson, Michael Tye, Jeff McMahan, Ingmar Persson, Roland Puccetti, and Roderick Chisholm). Some claim we are 3D brains or 3D parts of such brains, others claim we are 4D brains. Some have done so to avoid the extra thinker, others for different reasons. 1) McMahan and Persson describe two-headed conjoined twins, real and fictional. The thought realized by each of the two brains would be inaccessible to the other so there is a Neo-Lockean pull to say that there are two distinct persons rather than one person cut off from herself. And since there is only one organism in such scenarios, organisms can't be persons on pain of violating the transitivity of identity. Split brains and dissociative disorders also motivate some to claim we are just part of an organism on the basis of the same Neo-Lockean reasons about psychological unity. 2) A second reason philosophers might be attracted to this strange downsizing view is that they don't see how the brain or certain parts of it could be denied the capacity to think. The brain is seen in some sense as the minimal thinker, nothing else is needed or is *directly involved* in the production of thought, and this makes it the best candidate for being the thinker of our thought. And as Chisholm rightly insists, if anything strictly thinks one's thoughts, one should be identical to that entity rather than one that derivatively thinks in virtue of something else strictly thinking. 3) The claim that we are our brains receives further support from the brain transplants and body switching thought experiments that are so popular in the metaphysics of personal identity literature. 4) A fourth reason to believe we are our brains is that one believes mental states are brain states. That makes it sound like the brain would be the possessor of those states and thus the subject of any thoughts.

Although I am attracted to the prospect of neither being strictly bald nor having a beer belly, I am otherwise not a fan of the identification of the brain and person. There are a number of problems such as doing justice to our intuitive account of the persistence conditions of both brains and persons

if we identify them, and also of trying to cash out the idea of being *directly involved* in thought. But I just want to point out that there still looms the problem of the thinking organism. It is hard to believe that there are organisms possessing our brains that don't think. McMahan seeks to downplay the organism thinking by comparing it to a car that is only derivatively noisy because its horn is noisy. Just as there aren't two noisemakers, so there really wouldn't be two thoughtmakers. But matters are not as innocuous when thought is involved. Imagine that the brain-sized person insisted "I am essentially a person." The organism thinking a similar thought in virtue of possessing the same brain would have derivatively said something false.

Eric Olson points out that there would also be an *epistemic problem*. The thinking organism doesn't know whether it is the person or the organism if it thinks qualitatively the same thoughts with the same organ as the much smaller person. An additional puzzle concerns why the organism isn't a person since it is self-conscious and apparently can think about the past and the future in the way that the Lockean tradition maintains is characteristic of personhood. Locke famously wrote *"We must consider what person stands for: which, I think, is a thinking intelligent being, that has reason and reflection, and consider it self as it self, the same thinking thing in different times and places."* The organism has the capacities of the person but just lacks the persistence conditions of persons. Olson labels this the *personhood problem*.

There is an additional difficulty with the claim that the brain is strictly the thinking being and the organism only derivatively thinks. If Olson, van Inwagen and Merricks are right that the human organism could live pared down to the size of the brain, then it would be composed of every part of the brain and nothing else. It is quite odd that it would *then* only be derivatively a thinker. This queerness could be reinforced if it were metaphysically possible for a functioning brain to be made first and then head, neck, trunk and other appendages added later. It is very hard to explain why at the early stage only the brain would be strictly the thinker. It seems that there would be two bodies strictly thinking the qualitatively same thoughts. If one believes that the minimal thinking being is

just part of the brain and thus smaller than the maimed brain-size organism, one can just reproduce the problem with the brain and its minimal thinking part rather than the organism and the brain. The brain can survive the loss of some of its parts and become reduced to the size of the minimal thinker. These two cases presents us with the familiar puzzle of spatially coincident objects: two things physically the same and with the same relations to the environment, somehow end up with different modal qualities, and if an unwanted extra thinker is to be avoided, an account must be given of why one of the two physically indistinguishable entities can think.

VI. Noonan's Pronoun Revisionism

Harold Noonan doesn't think the abundance of thinkers is an insurmountable problem. He offers the three dimensionalist confronting spatially coincident and embedded thinkers an account of linguistic revision designed to serve his favored four dimensionalist approach which allegedly avoids the above problems. He separates the thinkers of our thoughts from the referent of "I." The idea is that the first person pronoun doesn't refer to all those who think or utter it but only to the person who is distinguished by his persistence conditions. So neither a brain that is part of the organism, nor an organism spatially coincident with the person, nor a perduring organism overlapping a perduring person, nor a stage that is part of a four-dimensional worm refers to itself. The first person pronoun "I" refers not to the thinker of the thought as we English speakers may claim to have been taught, but to the individual that is a person. Noonan insists that it is an a priori truth that the reference of the first person pronoun is a person and that metaphysical arguments can be made to show that the person is the entity that has the identity or persistence conditions of a thinking being rather than those of an essentially living being. As a result he insists that the organism, brain and the person can think truthfully and knowingly when they think "I am the person" because the referent of all the thinkers is the person.

I don't think Noonan should have offered a solution based on just changing our understanding of *how we speak*. Rather he should make a claim about *how we think*, or if you like,

about the language of thought. His claim shouldn't be just that the word "I" refers only to the person, but that the organism doesn't have the capacity to refer to itself in a way that is characteristic of a self-conscious person. It is the nature of the person to be capable of thinking about thoughts that it knows to be its own. The organism cannot think about its thoughts in this way. And the same can be said of the brain. So it isn't that a certain meaning of "I" in Kaplan's sense of character has been learned rather than a different one, but that the organism and the brain are psychologically incapable of reflecting on their own thoughts as their own. If the difference between the organism and the person's referential abilities were just a matter of conventional linguistic usage then we could just as well stipulate that "I" refers to the organism, and the organism could then refer to itself in a way the person couldn't. However, such a move should merely take away the person's ability to refer to itself by that word rather than to self-consciously think about itself. And while such a stipulation would allow the organism to refer to itself by "I," it should not bestow upon the organism the capacity to think that it is identical to the referent of "I." Otherwise, the epistemic problems return.

So Noonan's idea should be one about the nature of thought and not just about the referents of our natural language. While the organism and brain have second-order thoughts, thoughts about thoughts, they are not thoughts about themselves qua organism or qua brain. They have a very odd sort of direct, immediate, non-inferential idea of what someone other than themselves is thinking, but they don't have self-knowledge. Organisms, as well as brains, can think about themselves in the way John Perry's sugar spilling shopper can by some description. There may even be a sense in which they think about their own thoughts since they are co-generating the person's thoughts and aware of those thoughts. But they aren't aware of those thoughts as thoughts of their own. They can't conceive of themselves as thinking beings, as entities with thoughts about themselves. Thus there is no personhood problem as envisioned by Olson: two or more entities with the very *same* mental capacities but only one is a person due to its distinct persistence conditions. And since the first

person pronoun refers to the person, there is no epistemic problem as such questions can't be raised by the non-persons.

What we should demand of such an account is an explanation of why the organism or the brain can't refer to themselves. I doubt that a compelling answer is forthcoming. Why should the organism be unable to do what the person can if they are physically the same and causally related to the same environment? If persons learn that the first person pronoun "I" refers to themselves, why don't organisms learn the same lessons and thus refer to themselves? Likewise for the thinking brain: If the person or organism can be pared down to the size of the brain, why can't the brain self-consciously think like the person or the organism at that time? And even when the organism or person hasn't been reduced to the size of the brain, there is still quite a mystery about what the brain is missing that would prevent it from having the same mental capacities. It would seem to share the relevant physical makeup of the organism or person and would be causally related to the same environment, though a few steps removed. So while pronoun revisionism, at least when supported by a language of thought construal, can solve the *personhood problem* and prevent the *epistemic problem* from arising, it does leave a mystery.

Some in the audience might insist that it is just a brute fact that the person has the ability to think and the brain and organism do not, assuming that the organism is distinct from the person. But this is an unattractive move. "Going brute" would be more appealing if we were wondering why material simples like electrons can't think but immaterial simples like souls allegedly can. With simples, there is nothing to explain. Appeals are just made to the nature of things. There are not parts with properties and standing in relations to other parts that can explain the macro phenomena in question. But since organisms and brains are complex composite objects, one would think the difference in their cognitive capabilities should have an explanation in virtue of the properties of their parts.

VII. Coming to Terms with Being Brainless

Running out of options, the audience might want to join me in seriously considering eliminativism now that a look of puzzlement has perhaps replaced the incredulous stare. This is van Inwagen, Merricks and Olson's solution to the problem of too many thinkers - and the one that I accept though reluctantly. They eliminate brains from their ontology, identify persons with organisms, and reject the metaphysics of temporal parts. In fact, they eliminate all undetached parts that are large to enough to subsume the brain and apparently be capable of thought such as hand complements. Independent support for this approach is provided by van Inwagen's compositional principle that *X*s compose a *Y* iff they are caught up in a life. This is not as arbitrary (and disjunctive) a compositional principle would have to be to include cerebrums, heads and hand complements.

Olson, van Inwagen and Merricks suggest that we speak of atoms arranged brain-wise that don't compose anything where others refer to brains composed of atoms. (These are philosophical atoms, simples, not the composite atoms of the physicist.) Ordinary object language gets paraphrased in a way in which is in some sense close enough to true and can serve our purposes. Falsely claiming that there are brains when there are only atoms arranged brain-wise, is a lot different from falsely claiming that there are any unicorns for which no satisfactory paraphrase is available.

VIII. Tu Quoque?

There are various costs of paraphrase but I don't have the time to go into them here. Perhaps more worrisome for the Olson and van Inwagen-style theorist advocating a sparse ontology of just organisms and simples is the possibility of conjoined twins that share a cerebrum but no other organs (or rather atoms arranged organ-wise.) There was an actual case of conjoined fetal twins which shared a cerebrum but not a brain stem nor any other vital organs involved in the life processes thought to individuate organisms. Since they each had their own brainstems controlling different life processes there would have been distinct organisms on Olson and van Inwagen's accounts of organism individuation. But each organism would be the subject of thought, thinking thoughts that

appear to be qualitatively alike. Although the just described twins died when late term fetuses, perhaps then capable of only minimal sentience, if they had lived, the shared cerebrum might have brought with it the problem of too many thinkers and the epistemic bewilderment that Olson argues his opponents suffer. While such cases of conjoined twins are rare, we can conceive of a possible world in which they are not uncommon. And claims about the nature of human persons and human organisms and their relationship should apply in every metaphysically possible world. We don't want an account of the relationship of organisms and their brains to work just for our actual world.

So audience members might be thinking that if even the Olson/van Inwagen-inspired ontology must accept an extra thinker then they and their opponents are in the same leaking boat and thus there is no reason to advocate a sparse ontology. (Doubts about de re vagueness might also give this result.) We should just accept thinking persons, thinking organisms and thinking parts and rely on pronoun revisionism to take the sting off this. However, I think rival accounts (except for immaterialism – of some varieties of Cartesian and emergentist dualism) fare even worse in this cerebrum sharing case. My materialist rivals must posit more thinkers in the conjoined twin case as well as in normal scenarios. Furthermore, the van Inwagen-Olson ontology doesn't get stuck with embedded thinkers nor is there any violation of supervenience and a modal mystery as there are in cases of spatially coincident thinkers, so it doesn't share that weakness with most of its theoretical rivals. Thus I will bite the bullet and just accept the fact that I am unable to account for any self-referential thought in such cerebrum sharing twins. So in the case of the two organisms sharing a cerebrum, there would either be a failure of reference or, if reference succeeded, it would not be accompanied by the ability to ever know, when connected, which twin one was.

My hunch, though not necessarily my hope, is that the atoms arranged brain-wise in the audience will not be able to make any contribution to a better solution here or in any of the other scenarios of extra spatially coincident, embedded or overlapping thinkers.