Journal of Medicine and Philosophy, 41: 148–171, 2016
doi:10.1093/jmp/jhv064
Advance Access publication January 24, 2016

We Are Human Beings

ANDREW McGEE*
Queensland University of Technology, Brisbane, Queensland, Australia

*Address correspondence to: Andrew McGee, PhD, Australian Centre for Health Law Research, Faculty of Law, Queensland University of Technology, 2 George Street, Brisbane, Queensland 4001, Australia. E-mail: a.mcgee@qut.edu.au

In this paper, I examine Jeff McMahan's arguments for his claim that we are not human organisms, and the arguments of Derek Parfit to the same effect in a recent paper. McMahan uses these arguments to derive conclusions concerning the moral status of embryos and permanent vegetative state (PVS) patients. My claim will be that neither thinker has successfully shown that we are not human beings, and therefore these arguments do not establish the ethical conclusions that McMahan has sought to draw from the arguments in respect of the moral status of embryos and PVS patients.

Keywords: *brain, dicephalus, embryo, embryonic stem cell research, human being, human organism, McMahan, organ donation, Parfit, person, PVS, transplant*

I. INTRODUCTION

Jeff McMahan has claimed that we are not human organisms.[1] He advances two arguments for this claim, the aim of the arguments being to establish a number of moral consequences that, he believes, follow from it. One example of such a moral consequence concerns the permissibility of embryonic stem cell research. Many people believe we were once embryos. Although this belief can be disputed on other grounds, one important ground for disputing that belief is that we are not essentially human organisms at all. Assuming that it is wrong to kill "us" (he calls us persons [2002, 440]) but not wrong to kill organisms, then, if we are not human organisms, and the embryo is a human organism, it would follow[2] that it would not be wrong to destroy embryos for the purpose of embryonic stem cell research. A second example of a moral consequence McMahan draws from his two arguments concerns the permissibility, in some circumstances, of procuring organs from permanent vegetative state (PVS)

patients. McMahan claims that although the human organism remains alive when I fall into a state of PVS, I, by contrast, am dead (2002, 440). We would therefore not be killing persons if we killed the organism in order to procure its organs, or if we procured organs thereby causing the human organism to die. If it is wrong to kill persons, but not to kill organisms, then it would follow that it may be morally permissible to procure these organs (2002, 447ff). The two arguments McMahan advances for the claim that we are not human organisms have become influential, being invoked and applied very recently in a paper published in *Philosophy* by Derek Parfit (2012). Parfit made the second of the two arguments famous in *Reasons and Persons* (1984).

In this paper, I discuss and challenge the two arguments that McMahan advances in support of all these claims. My aim is to show that McMahan has not established that we are not human organisms, and so has not established the case he makes for the permissibility of embryonic stem cell research and organ procurement from PVS patients. If we are to explain why these practices are permissible, some other explanation must be found than that offered by McMahan. It is not my intention, however, to sketch any such alternative account here.

II. MCMAHAN'S TWO ARGUMENTS FOR THE CONCLUSION THAT WE ARE NOT HUMAN ORGANISMS

It is a characteristic feature of philosophical thinking since Socrates to put into question that which we take to be self-evident. Most philosophers consider this to be a virtue of the subject. Nothing should be taken for granted and left unquestioned. No belief or assumption should be immune to philosophical reflection. One such belief or assumption that most of us would think is self-evident is the belief that we are human organisms. McMahan's arguments are meant to put this belief or assumption into question. We are not human organisms, but the conscious part of that organism:

> A human organism is conscious only by virtue of having a conscious part. We are that part . . . The label I use to describe what we essentially are is "embodied mind." (McMahan, 2007, 186)

Explaining what he means by "embodied mind," McMahan states,

> According to the Embodied Mind Account, the criterion of personal identity is physical and minimal functional continuity of the brain. (McMahan, 2002, 69)

He also adds that, when referring to "the criterion of personal identity," his use of the word "brain" should be understood as shorthand for "those regions of the brain in which consciousness is realized" (67).

In his recent paper, "We Are Not Human Beings," Derek Parfit (2012), building on McMahan's arguments,[3] also denies that we are human beings. Indeed, not only are we not human beings, we are not even animals—we are

not an organism of any kind in both Parfit's and McMahan's view. Instead, Parfit says, we are the "embodied part" of human beings, and hence, of animals. Although my main focus will be on McMahan's arguments, I shall discuss some of Parfit's recent arguments when discussing McMahan's second argument below, since he endorses McMahan's account and describes it as the best form of what he calls "the embodied part view," the view that Parfit himself defends.

McMahan's First Argument

McMahan's first argument for the claim that we are not human organisms relies on cases of dicephalic twinning (as with the Hensel twins). This is an incomplete form of conjoined twinning "in which two heads, each with its own brain and its own separate mental life, *sit atop a single body*" (McMahan, 2007, 182, emphasis added).[4] McMahan points out that, in dicephalic twinning, there is very little duplication of organs below the neck and only one circulatory system, one metabolic system, one reproductive system, and one immune system. This leads him to conclude that, "[i]n these cases, there are two persons but only one human organism" (182).

It follows, he claims, that none of us is a human organism, for if we were, then there would have to be two organisms present in the case of the Hensel twins, rather than one, but there is only one (McMahan, 2007, 182). That being so, it follows *a fortiori* that nobody is killed when an embryo is destroyed (183, 186).[5]

However, McMahan's conclusion is too quick. His certainty that we would say that there is only one organism may come[6] from his conflating the organism with the trunk,[7] which the twins share, when he writes that "two heads sit atop one body." There is one trunk, but it does not follow that there is only one body, where "body" means *organism*, for of course the body in this sense *includes* rather than excludes the head. And there are two heads, not one.[8] The organism and its trunk are not identical; rather, the trunk is part of the organism. The fact that the organism, then, includes two heads makes it possible for us to regard the twins as two organisms conjoined, rather than one organism. Indeed, our very use of the terms "twins," "each," "the girls," etc. shows that we regard there to be two incompletely severed organisms here.

McMahan could reply that this is not so. He might insist instead that these terms show that there are two *persons* present, but not that there are two *organisms* present here. But this reply fails, for two reasons. First, it fails because it begs the question by *stipulating* that when we refer to the "twins," to "each," to "the girls," and to "Abigail and Brittany," we are only referring to *persons*, not living human beings. But nothing is ever proven by a mere stipulation or decree. Second, it fails because it ignores the complexity of the girls' situation. For a start, not all organs are shared. For instance, in

addition to the heads, although many organs are shared, some such as the heart, lungs, and stomach are not: they each have a heart; they have three lungs between them rather than sharing two, and each has a stomach.[9] In addition, each has her own spine. Also, one whole side (including arm and leg) is controlled by one of the twins, whereas the other is controlled by the other, and each does not have feeling on the other side. It is, of course, true that other organs are shared. But of itself, that is not a sufficient reason for concluding that there is only one organism and two persons; if we were inclined to regard the sharing of the organs as a reason for concluding that there is only one organism, there is no reason why we should not also be committed to regarding the sharing of the organs as a reason for concluding that there is one person. That we do not do so—that we treat them as girls rather than as one girl, that we give them two names, etc.—shows that we do not regard the fact that they are conjoined as decisive. As noted above, McMahan cannot appeal to the fact that there are separate heads here to conclude there are two persons but one body without begging the question.

In addition to his apparent conflation of the body (in the sense of the whole organism including the head) and the trunk or torso (which excludes the head), McMahan may be pushed into concluding that there is only one organism (but two persons) because he refers to there being one biological mass,[10] and so he believes that they are not numerically distinct entities. In that regard, the criterion for a single substance as being one spatiotemporal continuant is taken by McMahan to be decisive. But even if we accepted that there is one biological mass here, it does not follow that we cannot refer to two, only partially severed organisms. Just as we can regard the squares in a bar of chocolate as numerically distinct squares of the one bar even before we break them off, so we can regard the twins as two organisms, even if we cannot regard them as two *separate* organisms.[11]

These points do not rule out the possibility of borderline cases—at least imaginary ones. The fact that the Hensel twins share sexual organs and that if one dies the other will die simultaneously are factors that McMahan considers relevant to his conclusion (McMahan, 2002, 36–37).[12] He also imagines a variation, which he calls a severe form of dicephalic twinning, where everything is shared below the neck (38–39). He takes this to be a clear-cut case of there being one organism, but two persons. But this conclusion is a *recommendation* on his part about how we should classify such an acute case, were it ever to become instantiated. We might, however, in such a case, refer to the entity as a human being with two heads. If, for instance, one head is not controlling one side—one arm and one leg—with the other controlling the other side (so that there has not been any need to learn to coordinate movements, as with the Hensel sisters), that might be a decisive reason for classifying the entity as a single organism with two heads. But for the same reason, we can say that there is a single person with two heads. There is no reason whatsoever to separate the concept of a person from the

concept of a human organism here. There is no compulsion to decide the issue in the way that McMahan assumes it would be decided. In any event, McMahan thinks that the advantage of his first argument over his second argument is that it is not a hypothetical example as is the brain transplant of his second argument (about which more below). But with the variation (where we are to *imagine* a severe but hitherto unknown case of dicephalic twinning), we are in the realm of hypotheticals, and so it is best now to turn to the second argument, which is a much stronger argument. Before doing so, however, it is worth stressing that, unlike his hypothetical severe case, the case of the Hensel sisters is not to my mind a borderline case where we might say that there is one person with two heads or two organisms. Still less is it the kind of case that warrants McMahan's claim that there is one organism but two persons. Nothing in the example shows that this way of classifying the Hensel sisters is remotely compelling. But even if it did—even if we accepted that we should classify the sisters as a single organism and two persons—this would not show that the rest of us are not human organisms. It would merely show that we have chosen to categorize the sisters differently from how we categorize the rest of us.[13]

If these considerations are correct, then McMahan's first argument does not show that we are not human organisms and consequently cannot show that we were never human embryos (or that PVS patients are dead). It is, in fact, merely a recommendation on his part about how the twins should be classified and, stemming from this, about how we should classify our relationship to our bodies, fuelled no doubt partly by his conflation of the trunk and the organism, and partly by his assumption that "organism" should be defined simply as one biological spatiotemporal continuant (which ignores the possibility that two organisms exist, only partially separated).

McMahan's Second Argument

McMahan also offers a second argument for his claim that we are not human organisms. Suppose you and your twin have an accident, and your body is destroyed but not your brain, but your twin's brain is destroyed, but not his body. The only way to save you is to transplant your brain into your twin's body. Suppose this is done. "Most of us," he claims, "believe that the person who then wakes up in that body is you. But if you were a human organism, you would now be the dead organism from which your brain was extracted" (McMahan, 2007, 182). It follows that we are not human organisms, according to McMahan.

However, the thought experiment—endorsed also by Peter Singer (2009, 160–1)—does not show that we are not human organisms. Instead, it redefines[14] the word "person" and the concept of personal identity[15] exclusively in terms of psychological continuity and brain identity, allowing bodily identity to drop out as a defining feature. As Wittgenstein pointed out:

> . . . the *ordinary* use of the word "person" is what one might call a composite use suitable under the ordinary circumstances. If I assume . . . that those circumstances are changed, the application of the term "person" . . . has thereby changed . . . (Wittgenstein, 1958, 62)

The intuitions to which McMahan refers ("most of us believe . . .") are therefore no more than a hunch about how we might redefine or refine our use of the word "person"—what it is to be a person and therewith what it is to be the same person—should such a science fiction scenario become a genuine possibility, with the resultant person who wakes up claiming to remember doing things that I did.[16] But our *current* concepts of person and personal identity include *both* bodily identity *and* psychological continuity. As Peter Hacker notes, if you woke up tomorrow and seemed to recognize all the places I know, and seemed to know all the people I know, that would not make you me (Hacker, 2007, 308).[17] We may, of course, say of someone after severe amnesia following a coma, "he is not the same person; he is not himself," but this can be paraphrased to mean "he is not behaving as he normally behaves" or that he is behaving "out of character," as we say. If the police knocked on our door and asked us to identify the severe amnesiac, we wouldn't say we were unsure if he was one and the same person we thought he was because he is behaving abnormally and we cannot see his brain. Physical appearance and identity of the human being, that is, of the organism, over time[18] is a criterion of identity under our *current* concepts, and, in cases of doubt, the police can rely on fingerprinting to determine the identity of the person concerned. And that shows that the identity of the organism over time is partly constitutive of what we mean by "the same person," "him," etc. But by suggesting that the identity of the organism over time is *already* irrelevant by appeal to the well-known brain transplant example, McMahan leaves this criterion out of his account, not realizing that he has subtly dropped one of our current criteria for the identity and identification of persons.

McMahan could respond that the criteria I have mentioned are criteria for the organism in which I (the "conscious part of my brain" on his view) am "housed," and so are criteria for *the same organism* rather than, strictly speaking, for the same *person*. Since the organism and the person "coincide" in normal circumstances, those criteria enable us to identify the same person. But it does not follow, McMahan could continue, that the person and the organism are identical, only that they coincide or, as he puts it, that one "houses" or "occupies"[19] the other. We can nonetheless separate the two, in principle. We can imagine, for example, that the severe amnesiac to which I referred above is an amnesiac not simply by reason of a medical condition but because he has had his brain replaced by a different brain, unknown to us. If we rely only on his physical appearance, then *of course* we would not believe him to be a different person. But we would be wrong. This response would, however, itself be appealing to the new concept of what it

is to be a person—namely, sufficient continuity of those parts of the brain in which consciousness is realized—and so does not answer my contention that McMahan is merely recommending a redefinition of our concept of person in the light of what "most people are inclined to say" in response to a mere thought experiment.

Could McMahan cheerfully accept this? He could, but only if *that* possibility (the transfer of my whole brain to the brainless body of my twin) was itself *realized* and so not only scientifically possible but practiced. If such operations began to become routine, that *may well indeed* bring about a shift in our concepts of what it is to be a person and the same person. He cannot cheerfully accept this, however, merely on the basis of the *thought* experiment. This is because there are *other* possibilities that would make us reach *different* views and would result in our concepts being adjusted in different ways. A well-known example is the single hemisphere transfer. In this case, unlike the previous case, only half of my brain is transplanted, with my other half remaining. Assuming I retain my capacity for consciousness (as can happen after a hemispherectomy, a radical surgical solution to severe epilepsy), and assuming that the brainless organism to which my other half is transplanted subsequently wakes up, the question is which one is now me? There is enough continuity of the brain in both cases for McMahan's criterion to be satisfied (enough parts of the brain in which consciousness "is realized"),[20] but our concepts of person and personal identity break down in this case. As David Wiggins has suggested,[21] we might instead regard me in this scenario as a universal with two instantiations (Wiggins, 2001, 227, 229–30). Expanding the logic of Wiggins's point, if it were possible for me to create cloned brains by somatic cell nuclear transfer and then program the resulting brains with my memories,[22] we could multiply the instantiations. I would then no longer *die* but could instead only *become extinct*. Here, we have a perspicuous case where our concepts of identity and difference break down—with a suggestion about how they would adjust in such a world where these possibilities became routine. But the adjustment is different in this case from that which would occur on the transfer of the whole brain to a brainless twin.

McMahan, of course, could concede this point but insist that in *both* scenarios (the single brain transfer and the half hemisphere transfer) we can still maintain that we are not human organisms, but merely the conscious part of one. But the point is that, in the second scenario, we are *not* the conscious part of one. Rather, we are a universal with multiple instantiations (even if only two such instantiations), and the concept of *death* no longer applies to us, but rather the concept of becoming *extinct*. What this shows is that we cannot rely on thought experiments alone—logical possibilities (assuming they do express logical possibilities)—as the basis for claiming to have discovered, as McMahan does, what it is we essentially are. For different logically possible scenarios call for different responses to

that question. In each case, the scenario requires a *different* radical adjustment in our concepts. So it is only when possibilities become more than logical—when some possibilities become actual—that our concepts and our self-understanding would really shift; they don't shift merely by thinking through *hypotheticals*; rather, we discover only the conditions under which our concepts *might* break down (Hacker, 2007, 309) and so might, if the scenario became realized, require decisions about how they should henceforth be altered and applied, together with different ethical conclusions about our obligations to one another.

I contend that these remarks are enough to respond to McMahan's second argument. However, McMahan might still insist that, in both of the alternative scenarios I have discussed, the arguments show the brain retains a central role. It is not as though we would reach similar conclusions if our *bodies* could duplicate but our brain did not. We could imagine, for instance, that our brain could control more than one body.[23] We might, as Wittgenstein once suggested, be able to feel toothache in someone else's tooth or we might be able to control their entire brainless body and force the body to do the shopping, cleaning, etc., while I (the body in which my brain is "housed") continue to read McMahan's *Ethics of Killing* in my armchair. In such a case, we *might* say that I have several bodies,[24] can be in more than one place at once, etc. But the converse does not seem to hold. We can imagine four brains attached to the one body, each "housed" in separate heads—a four-headed man. If each head communicates to the other, argues and debates with the other, do we not have four persons in one body?[25] This point returns us, though, to the first argument, and the criticisms I made in that context would then apply. In short, these further variations remain logical possibilities and it is not at all clear how our concepts would or should shift in these cases, were they to become regular or routine actualities. We might want to know, for example, exactly why there are four heads. It would be a relevant factor, for instance, if this resulted from incomplete embryo splitting or whether, instead, scientists genetically modified an embryo so that it could grow four heads rather than one head. If the latter, the incomplete severance of twins argument on which I relied in the previous section could not apply, and this would be a factor enabling us to decide which way to go in terms of how we now conceive identity. Once again, these points only illustrate that different options are open to us, and so illustrate the Wittgensteinian point that "personality hasn't got one legitimate heir only" (Wittgenstein, 1958, 62).

Although I believe that these points answer McMahan's second argument, the thought experiment of the single brain transplant is widely used in debates about personal identity and personhood and has been invoked once again more recently by Derek Parfit who famously deployed it in his earlier influential work, *Reasons and Persons*. It has also been invoked in the context of debates about death in organ donation (Khushf, 2010). I shall

therefore now devote more attention to this particular version of the argument and, in particular, to Derek Parfit's very recent arguments involving the transplantation of my cerebrum to another body.

Further Analysis of the Second Argument

I have argued, following Peter Hacker,[26] that the famous twin thought experiment only highlights indeterminate areas of our concept of "person" and "same person." These are areas in which application of these concepts is not clear because they do not concern the standard cases in which the words "person" or "same person" are normally applied in everyday life, but rather concern fanciful and far-fetched cases.[27] Our concepts are not designed to cover every conceivable case but function to enable communication in standard (and perhaps some periphery) cases that obtain in our actual world. The possibility of a single brain transfer from my body to the body of my twin is only one of several different possibilities that might all be extreme enough to result in a shift in concepts—and a different shift in each case. In the case of the single transfer, if the possibilities of transplantation McMahan imagines should ever become possible, and start to become widely practiced,[28] a decision about how the concept of "person" is to be applied might be necessary at that point.[29] It might be that we would decide to apply the terms "person," "same person," etc., by disregarding the criterion of bodily identity. But the concept of *person* would then have changed.[30] It would not therefore show, *pace* McMahan, that we, that is persons, are not, on the *current* definition of person, organisms, and so could not show that we were not once embryos, or that PVS patients are dead.

Essential and inessential criteria?

McMahan might respond by claiming that these points ignore the fact that, among the various criteria for the application of the term "person," some are essential, but others are merely accidental. Indeed, Derek Parfit has considered the fingerprint criterion and makes remarks that might seem to respond to the Wittgensteinian objection. In the case of a conflict between that criterion and the brain and psychological continuity criterion, it seems obvious that the latter is essential, whereas the former is not—in other words, the former is more a *symptom* of identity than a *criterion* of it. For example, suppose a surgeon could remold the fingertips of someone's fingers, and did so. In such a case, Parfit says, we would still maintain that the person would continue to exist, with the same brain and psychology, though with different fingerprints (Parfit, 2012, 10).[31] But would we say the same thing if someone's brain were removed, and he were kept alive on life support? Parfit contends that we would not (Parfit, 2012, 10). If correct, then Parfit could say that, on our *current* concept of "person," the person whose fingerprints are changed remains one and the same person—indeed, this fact is

already recorded in how I have just expressed this very point ("the person whose fingerprints are changed"). By contrast, on our current concept of "person" we would not say that the person who has had a full brain transplant remains the same person—indeed, for that very reason, we would not say "the person who has had a brain transplant" at all, for that assumes that one can remain the same person with a brain transplant, but the transplant example is meant precisely to show that this is not the case. Does this not show that our criteria of identity *already* record what is essential and what is accidental to being a person? If this were not the case, then why is it that so many of us feel that the brain is the one organ in our body that we would not want to have replaced via a transplant? A kidney or heart transplant does not raise any issues of identity, but surely even if we hesitate or are unsure, the very prospect of needing a brain transplant would raise the question of whether I would be the one who survives the transplant or whether or not it is actually someone else who inherits my body. If so, it might be concluded that the Wittgensteinian points just made would fail, because they would have ignored the fact that some criteria for the application of "same person" seem essential, whereas others do not.

In reply to this point, we can begin by noting that some of the intuitions relied on by McMahan and Parfit in the thought experiment seem to work because it is the body of my *twin* to which my brain is transferred, and only the immediate circumstances of the transplant are described. Neither thinker countenances the possibility that we might regard the resultant being as a *hybrid*, that is, as a different person from either A or B. But what if the brain of an unrelated female was removed and my brain was transplanted to her female body? Is the resulting person still me? It would be controversial to say that our sex is not an essential feature of our identity, of who we are.[32] Could we not classify the resultant person as a hybrid of me and the female whose brain has been discarded? Is the resultant person a male trapped in a woman's body, or is he a female?

Parfit, at least, imagines his head being attached to Bernard Williams's body, and so offers a non-twin example. However, he believes that our analysis would nonetheless remain the same. But the way he sets up his thought experiment partly forces us to reach the conclusion he wants us to draw. Parfit invites us to imagine that we know both Parfit and Williams personally. Imagine that Parfit's head is transplanted onto Williams's body, and that we visit the resultant person in the postoperative recovery room (Parfit, 2012, 9). Seeing only Parfit's head on the pillow—the body to which it is attached being covered by sheets—we hear Parfit talking and assume that it is indeed Parfit who is talking. Parfit contends that, even once the sheets are lifted and we see that Parfit's head has been attached to Williams's body, we would not conclude that we were *not*, after all, talking with Parfit. But the description of the circumstances makes that conclusion compelling. It leaves out other factors that are relevant, and, more importantly, it ignores

the possibility that we might come to revise that intuition and persist in a state of uncertainty, sometimes thinking that the resultant being is Parfit, and sometimes withdrawing that belief.

First, assume, as Parfit later invites us to do, that only the cerebrum is transplanted. Parfit maintains that our conclusion would remain the same if only the cerebrum is transplanted onto Williams's brain stem, on the ground that "it is our cerebrum on which all of our distinctive mental activity depends" (Parfit, 2012, 11). Assuming that this claim is accurate,[33] then the same conclusion is said to follow, namely, that we would conclude that the person who wakes up is Parfit. Yet the resultant person, with Parfit's cerebrum, would not look like Parfit, and certainly would not sound like him. This may give us pause before we assume we are talking with Parfit—even if Williams's body remains covered by the sheets. Second, assume that the transplant happened when they were much younger. I might know that Williams is a brilliant sprinter and was close, at one stage, to turning professional. Parfit, on the other hand, was an awful sprinter, so much so that he and Williams used to joke about the fact that this was the one competition at which Williams could give Parfit a good beating. When the resultant being who wakes up following the operation has suitably convalesced and starts running like Williams did, I might now hesitate in assuming that I had been talking to Parfit on that day. I might now come to believe that "Parfit" is actually a *hybrid*. Further, conditions in Williams's body that Parfit never had may make "Parfit" irritated and hot tempered, like Williams was. Parfit, however, was never hot tempered in this way. Would it be irrational for me to hesitate, here, to continue with my assumption that I was talking that day in the postoperative theater with Parfit, and not some hybrid? Once these other factors are introduced, I think we are likely to be less certain—to change our minds and to conclude that we do not really know whether the resultant person is Parfit or not: we might instead decide that this is some heir of Parfit, closely related, but nonetheless distinct. Changes in personality, the acquisition of different interests that are aligned to those Williams had, such as sprinting, may lead me to this conclusion. To be clear, the point of these other examples is not to suggest that we definitely *would not* say that the resulting person waking up is Parfit. Rather, it is only to suggest that saying so is not the *only* option available, that it is not inevitable, or already dictated by, our *current* concept of personal identity.

Parfit might respond by saying that we are certainly not tempted, here, to call the resultant person *Williams*. It is one thing to be uncertain about whether the resultant person is *Parfit*, but quite another to be uncertain about whether it is Parfit or *Williams*. None of the above examples—the sprinting, the irritability that were traits of Williams—makes us consider the possibility that it might actually be Williams who wakes up. This is true. It is not my contention that we would hesitate about whether the resultant being could be Williams rather than Parfit. My contention is only that we might

decide that the resultant person is neither Parfit nor Williams. But I do not need that stronger contention to argue that we are human organisms, rather than brains or embodied brains. If the weaker possibility is open, then it follows that we might still consider that we are human organisms, even if the science fiction scenario becomes possible.

Given these points, the Wittgensteinian claim that the term "personality hasn't got one legitimate heir only" (Wittgenstein, 1958, 62) still stands: we can vary the facts, as we have done with these different thought experiments, to show that different ways of applying the term "person" under those various circumstances become possible. Parfit himself concedes this, when recognizing that our intuitions about what we would say in the case of a *whole* brain transplant take a completely different turn if, instead of imagining a whole brain transplant, we imagine *one half* of my brain going to one body, and *the other half* going to another, different body (Parfit, 1984). The transitivity of identity rules out my being identical to another person, so which resultant person is me? We cannot say—but as noted above, following Wiggins, our concept of what it is for "me" to be a person may shift so that I am regarded as a universal with two instantiations here. Yet if only one half of the brain is transplanted, and the other half is destroyed, we seem more tempted to say that the resultant person is me. These problems might themselves form good grounds for concluding, as I have suggested above, that the resultant person is *neither* the person from whose body the brain is transplanted, *nor* the person to whose body the brain is transplanted. On such a view, the problem caused by the transitivity of identity would not arise. Once again, however, only a *decision* will settle the issue—and which decision we make will depend on the possibilities that become standard.

A questionable empirical assumption—ignoring the vertical plasticity of the brain stem

One final point is worth making. As noted above, Parfit's latest work involves imagining only the cerebrum removed to another person's body—the brain stem being left intact. Parfit says that, on this experiment, if our brain stem continued to maintain the functioning of our heart, lung, and most other organs, the human animal would continue to exist, though in an unconscious, vegetative state, or coma (Parfit, 2012, 11). Yet, Parfit says, "it is our cerebrum on which all of our distinctive mental activity depends" (11). This, he thinks, would make us say that the person goes with the cerebrum onto the brain stem and body of the organism to which the cerebrum is transplanted, notwithstanding that the organism from which the cerebrum is taken remains alive, supported artificially (11). This, he believes, proves that we are not human organisms—for the organism would be left behind, while Parfit himself goes with his brain.

However, this account is based on assumptions that have been challenged scientifically. Franklin Miller and Robert Truog have criticized McMahan's

belief that consciousness and mental activity are realized in the outer layer of the cerebrum (Miller and Truog, 2012, 88), on the basis that there is evidence that consciousness does not require the function of the cerebral hemispheres. If correct, these criticisms would also apply to Parfit's claims. The evidence in question has been reported by Bjorn Merker (2007) and Shewmon, Holmes, and Byrne (1999).[34] Evidence presented in these papers of children born without a cortex nevertheless being conscious leads Merker to suggest that "the brainstem mechanisms are integral to the constitution of the conscious state" (Merker, 2007, 63). In these hydranencephalic children, the brain stem acquires some of the functions of the cerebral cortex, including consciousness (Shewmon, Holmes, and Byrne, 1999). Indeed, Merker suggests that the fact that bilateral cortical damage will typically result in PVS "does not, however, allow us to make an equation between cortical function and consciousness, because such damage also inevitably disrupts numerous *brainstem* mechanisms normally in receipt of cortical input" (Merker, 2007, 65, italics added). This suggests that a person with an undamaged brain stem might in principle be able to regain some conscious awareness,[35] and these cases are precisely the cases that are relevant to the transplant examples used by Parfit and McMahan. If that were so, the case of the cerebrum transplant would then become akin to the case of the transplantation of one of my hemispheres to one awaiting body, with one half retained in my body, which, as we have seen, yields no determinate answer as to which of the two beings I am. Indeed, in this case, the option that Parfit (whose brain stem acquires some consciousness and whose cerebrum is transplanted to Williams's waiting decerebrate body) should be classified as a universal with two instantiations is not available either. For if Parfit's cerebrum is placed on Williams's brain stem after Williams's brain stem has regained some consciousness via "vertical plasticity" (Shewmon, Holmes, and Byrne, 1999, 371; whereby the brain stem acquires some of the functions that would normally require the existence of the cerebral cortex),[36] we can only say, once again, that the resulting organism is a hybrid of Williams and Parfit. But what if the brain stem left behind in Parfit's body by Parfit's cerebrum should regain some consciousness at the same time? For example, Parfit might also regain consciousness, in spite of the removal of his upper brain, by means of "vertical plasticity." We would then be much less inclined to say that the person who wakes up when Parfit's upper brain is transplanted onto Williams's brain stem is Parfit. In other words, the strongest intuition that Parfit and McMahan rely on to ground their claim that we are not organisms could in fact turn out to be the weakest. If these suggestions are correct, our "intuition" that Parfit would go with his brain, such that Parfit is the one who wakes up, attached to Williams's body, would now not be so great at all.

Again, these considerations illustrate the point that changes in the facts lead us to different conclusions concerning what we would say, that is, about how we might project the concept of "person" into these new contexts. The

core of our current concept, however, remains unchanged by the imaginary scenarios, because we are not in a situation where we must make a decision. Consequently, although it is possible that we might, one day, define ourselves as something other than human organisms, this does not show that we are not human organisms on our *current* definition of person.[37]

III. COULD MCMAHAN'S POINT BE MADE INDEPENDENTLY OF THE TWO ARGUMENTS JUST CRITICIZED?

I have been claiming that the two arguments McMahan advances in favor of the claim that we are not human beings are not successful. I have also claimed that Derek Parfit's more recent arguments for the same conclusion are likewise not successful. If my claims are correct, then the ethical conclusions they believe to follow from their arguments will not in fact be established.

It might be objected that, notwithstanding my criticisms of the arguments just discussed, it is nonetheless plausible to regard scientists as having discovered that we are our brains or regions of our brains in which "consciousness" is realized. Could McMahan's claims that we are not human organisms, but "the conscious part" of human organisms, stand independently of the two arguments I have criticized? Without the brain, we cannot be conscious, and so we cannot have a personality at all. It seems that the brain is essential to our identity in the way that other organs of the body are not.

This is, of course, true, but it does not show that we *are* "the conscious part of the brain," but merely that those regions of the brain to which McMahan refers are the enabling conditions for the *organism* to be conscious.[38] As Hacker notes, although it is true that without a brain we cannot think, it is also true that without a brain we cannot walk, but nobody is tempted to say that it is really the region of the brain that walks (Hacker, 2007, 307). Similarly, it is the *organism* that is conscious, not the brain or part of the brain. But couldn't one insist that the organism is conscious via the brain's being conscious so that, strictly speaking, it is only the brain that is conscious rather than the organism, the organism only being conscious in a derivative sense? If we talk of human beings being conscious, it might be said, that is just a shorthand way of referring to their brain or the relevant parts of their brain. Just as the human being clasps the cup only by its hand doing so, so the human being thinks only by its brain doing so.

However, in the case of the hand, things are really the other way round. The hand clasps only because the human organism moves it and clasps objects with it, so to speak of the hand clasping is really only a piece of metonymic substitution. The hand is not the actor or agent, but the human being. So that analogy would be misleading.

Behavioral Criteria for Ascribing Being Conscious, Unconscious, Seeing, Hearing, Speaking, etc., to a Being

If it were really *the brain* that is conscious, sees, hears, and speaks, then we would need to know what the criteria are for the brain's[39] doing these things, as opposed to the organism's doing them.[40] The organism is not a property of the brain but is a substance and so a subject of predication.[41] If we say of the *brain* that it sees, we need to know with what it sees, but once we introduce the other organs to solve this problem, we have reintroduced the organism.[42] When we speak of you or me doing things, such as thinking, reflecting on a problem, listening to the lecture, etc., there are behavioral criteria[43] on which I rely to know what you are doing. For example, you can tell me what you think of a certain problem, and I can know from that not only what you think of it, but how deeply you have thought about it. I rely on what you say, *out of your mouth*, literally. Similarly, I can tell when you are deep in thought, or if you have heard a funny line or appreciated a fine piece of art or poetry, for there are criteria I can go on. Of course, these won't always be satisfied, and pretense is possible. But pretending to think is parasitic on thinking in the sense that you would still have to engage in behavior that makes it *look as though* you are thinking.[44]

What are the criteria that we would go on if we were to apply such predicates as "conscious" and "conscious of" to brains? Another way of asking the same question is this: how do we know that we haven't merely discovered that the brain is an enabling condition for *our* doing these things, that is, for the human being doing them?

Could we answer this question by saying that the brain uses the organs to see? But this would be to adopt the metaphor that thinkers in favor of the brain as the proper subject of predication reject, of the captain steering the ship.[45] It would be like saying that the engine uses the body of the car to move.[46] Just as the body of a car is not a puppet for the engine, so the human organism is not a puppet for the brain.

Another attempt could be made to respond to these claims. These claims rely on Wittgenstein's well-known dictum that only of a living human being and what resembles (behaves like) one do we say that it has sensations, it sees, hears, is blind, is deaf, is conscious or unconscious (Wittgenstein, 1953, para 281). But could we not imagine that my brain is taken from my body and temporarily survives in a jar until it is placed in the body of a different human organism, while retaining consciousness throughout? Here, Parfit could say, we do not need to imagine violating the well-known Wittgensteinian dictum when entertaining this possibility. We can avoid ignoring the point that behavioral criteria are a precondition for the ascription of these properties to us in the following way: we can say that the resultant being can *now* tell us that he remembers being told what would happen to him just prior to the brain being removed and remembers suddenly being able to speak and

feel and see again once the brain was "hooked up" to the new organism. The behavioral criteria here, Parfit might say, are satisfied by the *subsequent* organism (with my brain now transplanted) *speaking and telling us* that he retained consciousness. We can even vary the example to make it less objectionable to someone persuaded by the Wittgensteinian dictum, by imagining that I was given an anesthetic prior to the operation and on waking up in the new body I state that I remember being given the anesthetic "when I was in the old body."[47] In such cases, Parfit may contend, there is no problem whatsoever in saying that, in the interim period, the brain was conscious (on the first version of the scenario) or unconscious (on the second version of the scenario). And if, indeed, it is true to say that only of a being that *can* be conscious does it make sense to say that it is *un*conscious, then on either of these scenarios we have a case where we can truly say of the brain that that brain was I, and, as that brain, I was conscious or unconscious while I waited for the transfer to the new organism. Since this makes sense (so it would be argued), and since my identity is preserved throughout on this scenario, it seems to follow that I am my brain.

In reply, I believe it would be wrong to say in such a case that I am a brain, or was my brain at least during that period the brain was kept alive in the jar. If we indulge the fantasy, we *might* instead say that I was reduced entirely to the state of having only potentialities, as I was when I was an embryo. But this no more shows that I *am* the brain *now* than does the fact that, on some views, I was once an embryo show that I am *now* essentially an embryo. Furthermore, without the organism, it is highly doubtful that the brain could really be conscious of anything—all the perceptual capacities would be absent, for instance. Even if we accept "it can think" on the behavioral criteria that might on the above argument be suggested by Parfit, this does not show that it is the brain that thinks (and note that "it can think" is already a grammatical revision). It only shows that because the human being whose brain was removed could think, the brain, detached from the human being, *might* be able to exercise in a residual way some of the capacities, to some extent, that the human being could exercise, for a short period of time thereafter.

I contend that, in order to say that it is really our brain doing all these things, we would have to be able to imagine "teaching a brain"[48] from birth to do all the things the human being can do—it is not sufficient to start with the human being first, and then extract the human being away by these thought experiments while tacitly relying on the capacities for whose development and exercise the behavioral repertoire of the *human being* is a precondition. Since I could not have developed to be the person I am without being the organism that I am, I cannot be identified with my brain. It is true that I could not have developed to be the person I am without my brain either, but that is why I am the whole human organism, including brain and body. So peeling off the body, as it were, *after* the features that give me my

distinctive personhood and personality have developed, does not suffice to show that I am a brain, notwithstanding that I might survive, in rudimentary form, in the experiment just discussed.

Is the Pronoun "I" Systematically Ambiguous?

Parfit has very recently argued that the personal pronoun "I" may be systematically ambiguous, referring at times to what he calls "Outer I," the human organism (as when I say I am sunburnt all over), and at other times to what he calls "Inner I," my conscious part (as when I say "I have been thinking about an abstract problem") (Parfit, 2012, 21–24). This might seem to resolve the logical problems that otherwise flow from identifying us with regions of the brain. For, on this view, "I can see" would mean "Outer I can see," whereas "I am thinking" means "Inner I am thinking."

One immediate difficulty, however, with this claim is that, as Wittgenstein pointed out, a word has a meaning we have given it (Wittgenstein, 1958, 27–28), and we cannot milk any more out of it than we put in. In order for the word to be ambiguous, there must therefore already be a recognized meaning that the word has been given that would, for example, be recorded in standard dictionaries such as the Oxford English Dictionary. Otherwise, Parfit's claim is no more than a recommendation to *give* the word a *new* meaning, one it currently does not bear, thereby *creating* the ambiguity whose existence he purports to be reporting.[49] This is, I think, a significant problem for his theory. But suppose we accept his theory as a recommendation instead. The difficulties for his position are not solved. For we must now ask: what are the criteria for stating that "part of my brain, or Inner I, is thinking about an abstract philosophical problem"? If they are our normal criteria—for example, A's head is cupped in his hands and he is frowning with concentration as he sits in silence thinking about it—the recommendation reduces to no more than one for a change of *label* ("I think" being replaced with the label "Inner I thinks"). No different *fact* is thereby recorded at all.[50] For Parfit's argument to count as more than a merely verbal procedure, there must be something, some criterion, over and above our normal criteria for the application of "he's thinking intently about this," for the change he recommends to be genuine—some way we can meaningfully ascribe the thinking to the relevant part of the brain, but not to the whole animal or organism. And this is extremely difficult to cite. If Parfit cites, for example, electrical impulses, or other activity in the brain such as might appear through imaging techniques such as PET or fMRI, this still would not show that it is the brain (or relevant region of it) that thinks. For, as Bennett and Hacker have pointed out, what is observed via PET or fMRI is the brain activity of the *subject* performing the tasks that the subject is invited to perform (we do not observe the *brain* performing those tasks, for "perform" can only be used of the *subject*, on the basis of the subject's behavioral criteria) (Bennett and

Hacker, 2003, 83). So it is the human being that we are witnessing engage in inferring, thinking, seeing, hearing, etc., not the brain—in looking at the image produced by the scan, we are merely looking at the brain activity that occurs *when* the *human being* is thinking, etc. To discover the *brain* doing things, which would not amount merely to the enabling conditions of an *organism* doing these things, we would surely need to see the brain doing something while the human being whose brain it is is *not* doing those things. But no such criteria are forthcoming. But without the criteria, the recommendation is one merely of a change of label (from talk of the human being doing x to the brain doing x).

Could we not say, however, that the brain activity that we witness via PET or fMRI *is* thinking, seeing, hearing, etc., and that these imaging techniques are another way of having access to someone's thinking, etc.? If so, have we not shown what it means to say that it is really the brain that thinks, sees, and hears? This temptation is especially acute when we consider that we might rely on activity in the brain in the case of minimally conscious patients as the only way of telling whether they are in a state of consciousness or not. But Bennett and Hacker (2003, 84) have argued that the correlation here is inductive, rather than criterial, which means that we already need to know what, for example, thinking is—what counts as thinking—in order then to correlate the activity in the brain with (what we call) thinking (84). A logical consequence of this point is that we do not discover what thinking is—how would we know what to look for if we did not already know what thinking is?—but rather we discover what makes thinking possible, that is, the enabling conditions that make it possible for the organism to exercise the powers or capacities to think and reflect, etc.

This is confirmed by a very simple logical point that Bennett and Hacker make in the context of the use of these scanners to observe patterns in the brain when the subject is in pain: if evidence from the imaging techniques suggests that the patient is in pain, but the patient sincerely insists that he or she is not, the evidence from the imaging techniques is defeated (Bennett and Hacker, 2003, 83). We apply our concept of pain—with its associated behavioral criteria (here, the sincere insistence that he or she is not in pain)—to determine whether the imaging is accurate, rather than the reverse. The same points would apply if scientists started to claim that such techniques could decipher the content of our thoughts. We already presuppose our concept of thinking as it applies to the organism as a whole when we undertake these experiments, and, at most, we are inductively correlating the patterns we see via a scanner with the thinking that the *subject* is doing. If the subject sincerely says that he is not thinking, then the evidence on the scanner is defeated. The subject's own sincere expression of whether he or she is thinking, and what he or she is thinking about, is the final court of appeal. But if it were the brain that thinks and feels, as shown through the scanner, then this would not be the case. Instead, we would have to say

that the subject is wrong, and that the brain is thinking, or thinking about a particular thing (a trip to London, say), even when the subject denies this. But we do not do this.[51] Our not doing so tacitly shows that we do not really ascribe thinking to the brain, but rather to the organism. To ascribe it to the brain, instead, would change what we mean by thinking, because it would change the criteria for its application. No reason, however, has been disclosed for us to change our way of speaking, and to start referring to the brain as the real thinker, as opposed to the organism.

One final problem with identifying what Parfit calls "Inner I" with the so-called conscious part of my brain is this. The expression "I am conscious" would be a tautology, for it would mean "my conscious part is conscious." And *whose* conscious part is my conscious part? What does the word "my" mean in "my conscious part"? We cannot say it means the conscious part of me, where "me" refers to the organism, for the organism does not have a conscious part. Rather, the brain does so, and the brain has been distinguished from the organism by both McMahan and Parfit (it being an organ not an organism). But we cannot say that it means the conscious part of me, where "me" refers to the conscious part of the brain, for that would yield the nonsensical: "my conscious part's conscious part." Spelling out the logico-grammatical consequences of any such recommendation in this way is likely to make us less inclined to accept Parfit's proposal.

IV. CONCLUSION

McMahan's claim that we are not human organisms appears to question the obvious. But because it is a philosophical claim, the fact that it questions the obvious is no objection, for, traditionally, questioning the obvious has been considered a virtue of the subject. Nonetheless, once the arguments of McMahan for the conclusion that we are not human organisms, but merely the conscious part of the organism, are subjected to critical scrutiny, it can be seen, in my view, that they are found wanting. The claim that we are *not* human beings arises principally from a misconstrual of the implications, for our self-understanding, of the possibility that our world could be different in a number of alternative ways—each of which might mean that we would adopt different concepts of personhood and personal identity for that particular world. The resultant self-understanding would be different in each case. If I am right, then those possibilities do not have any bearing on our self-understanding in *this* world, and we can continue to classify ourselves as human beings in our world. Accordingly, McMahan has not shown that we are not human beings in our world, and his arguments therefore cannot be used in support of the claim that, since the embryo is an organism, but we are not, we are not killed when an embryo is killed. For the same reasons, they cannot support the further claim that stem cell research should be

permitted. Other arguments for that conclusion would need to be relied on, but I shall not pursue those arguments in this paper. Neither can McMahan's arguments support the conclusion that PVS patients are dead, even though the PVS organism is alive. This has implications for McMahan's endorsement of the upper brain criterion of death and his ethical arguments concerning the use of PVS patients and other patients whose upper brain has died, as potential sources of organs. McMahan claims that in PVS, "cerebral death has occurred, but the brainstem remains intact and functional." "Thus," he continues, "although the organism remains alive, the person has died or ceased to exist" (McMahan, 2002, 440). He then claims that "we should assign the living organism in a PVS much the same status as we now assign a dead human organism" (447). I believe that my arguments above invalidate this claim to the extent that it is based on McMahan's two arguments and therefore on his view that we are already defined as the conscious part of the organism. To that extent, the upper brain criterion for declaring death is not, in my view, a valid criterion for death. In my view, the use of such patients would be a clear violation of the dead donor rule. In this paper, I have tried to expose what I believe to be the faulty metaphysical thinking that has led McMahan to these ethical views.

ACKNOWLEDGMENTS

I should like to thank Dr. Peter Hacker for commenting on a previous draft of this paper and for saving me from some errors. Any errors that remain are entirely my own. I should also like to thank an anonymous reviewer at the *Journal of Medicine and Philosophy* for helpful comments. Finally, I hope that my criticisms of McMahan and Parfit do not conceal my admiration and respect for their work, as hopefully is evidenced by the extent to which I have grappled with their treatment of the issues in this paper.

NOTES

1. In this paper, "human being" and "human organism" will be used interchangeably.
2. I leave aside here the more esoteric possibility that we are souls (ensouling, or being attached to, the embryo from conception), which McMahan (2007) discusses and dismisses. For reasons of space, I cannot examine that possibility in this paper, but it seems to me that before we can ask whether the claim that we are souls is true or false, we should first need to know what counts as "being a soul" and how a soul might be "attached" to, or "ensoul" the embryo. These questions concern questions of meaning because, if the soul is conceived of as a nonphysical entity, it is not clear what it would mean for a "nonphysical" entity to be "attached" to a physical entity. For discussion of these and similar issues, see Rundle (2004).
3. The argument represents a modification of his earlier view in *Reasons and Persons*. I will not pursue the subtleties of those differences here.
4. Parfit cites this argument and endorses it in Parfit (2012, 14, 17).
5. Note, once again, the qualification in note 2, which I leave aside here.
6. It may also come from his taking the criterion of "spatiotemporal continuant" to be decisive, for there is only one spatiotemporal continuant, one biological mass, in such a case. See Campbell and McMahan (2010). I discuss this point below.
7. Or at least with the body below the neck. But "organism" does not refer to the body below the neck, but to the whole entity, and so includes the head. As such, it would be nonsense to say that two

heads sit atop a single body where "body" means *organism*—that only makes sense if the word "body" refers to the torso or trunk.

8. Derek Parfit may very recently have made a similar mistake when suggesting that "we are *embodied* heads" (Parfit, 2012, 17). The head is *part* of the body, in the sense of organism, and so we can no more be *embodied* heads than we can be embodied *bodies*. At most, we could be embodied brains—and I shall discuss this suggestion below.

9. McMahan (2002, 36); Campbell and McMahan (2010, 286).

10. See note 6. He also refers to "a single biological life" that supports "the existence and thus the lives of two distinct persons" (McMahan, 2002, 37). But "single biological life" is compatible with there being two impartially severed organisms present. Here, McMahan trades on an ambiguity between "single spatiotemporal continuant" and "single organism."

11. Alternatively, we can simply say that they are numerically distinct in one sense (we can count the squares) but not in another (they are part of one bar). Just as we leave this issue open and don't consider there to be a problem in referring to distinct squares and one bar, so we can refer to distinct, incompletely severed organisms but one biological mass.

12. For the possibility of drawing a conclusion different from the one McMahan draws from these two considerations, see Mulhall (2002).

13. Whether we would make the particular choice McMahan believes we should would depend on the coherence or otherwise of the proposal with the rest of our conceptual scheme.

14. In making this point, I have relied on Hacker (2007, 306) and Glock and Hyman (1994), who discuss the famous thought experiment in the context of a discussion of Bernard Williams's criticisms of P. F. Strawson's *Individuals*. The following arguments attempt to develop the point.

15. It is sometimes claimed that the concept of personhood and the concept of personal identity are separate issues. We can ignore this claim for now. On the views being here considered, if I stop being a person, it is not strictly I who do so, but only "my organism." My identity is therefore tightly bound up with what I am on these views. On these accounts, I can't become a different human being because I am not a human being; I am a brain (more precisely, I am those parts of the brain in which consciousness is realized). And I cannot become a different brain on these views. There is, of course, a distinction between qualitative and numerical identity: the possibility of A becoming a different creature (say turning from a human being into a frog) or even of A becoming a different thing altogether (Lot's wife becoming a pillar of salt, see Broyles [1985, 59]). We can imagine one spatiotemporal continuant transforming from one kind of thing into another kind of thing and we can say of *that* spatiotemporal continuant that it was so transformed. If we do, we can say that that thing is now something else, meaning only that it is qualitatively different. In this sense, the identity question does come apart from the nature of the thing question. But there are conceptual limits to such possibilities—it is doubtful, *contra* Broyles, that there is one underlying thing that remains "self-identical" in the case of Lot's wife transforming into a pillar of salt, of which the human being and the salt are mere phases. When wine is transformed into vinegar, the wine no longer exists. But wine is a substance, not a phase of some other substance (thanks to Peter Hacker for this example).

16. Note that if the resultant person did not claim to be me and had none of my memories, etc., we would be far less tempted to claim that the resultant person is me.

17. We shall see later that one reason for this is that there would be two people claiming to be the same person, and the transitivity of identity rules out two people being one and the same person. Some philosophers think that if I died immediately before the twin claimed to be me, we might *in that case* claim that the twin *is* me. Parfit's psychological continuity account might lead to this view, though it is worth noting that his claim that psychological continuity is not identity (identity being psychological continuity *plus* the no-branching requirement), means that Parfit is not himself committed to this view. This distinction between psychological continuity *without* the no-branching requirement and psychological continuity *with* the no-branching requirement is what enables Parfit to claim that identity is not what matters and also enables him to bypass the transitivity of identity problem. These moves are all, however, redefinitions of our concepts of personal identity and personhood based on imaginary hypothetical possibilities that, were they to occur, would result in a breakdown and shift of our current concepts of personhood and personal identity. It is far from clear what implications these *imaginary* cases really have for "what matters," given the world we actually live in.

18. I owe this way of expressing the point to Peter Hacker.

19. McMahan speaks of organisms in PVS as "unoccupied" (McMahan, 2002, 443).

20. I here adopt only McMahan's own way of expressing this point—I don't myself endorse the view that we are "those parts of the brain in which consciousness is realized."

21. I owe this point, and the reference to Wiggins, to Peter Hacker.

22. I assume that the possibilities I am entertaining here are meaningful, for the sake of argument. But it is doubtful that "programming the brain with memories" is a meaningful proposition: a picture of programming computers is transferred to the brain while all the differences are dropped to make the picture seem compelling. For formidable criticism of the view that we can transfer memories, see Schechtman (1990).

23. This way of imagining the next scenario is misleading, because it implies that the brain controls our bodies, but this is not so. It makes no sense to speak of the brain controlling the body.

24. Note that it might in *that* case then make sense to speak of possessing a body in the way that, in our *current* conceptual scheme, it does not. As Hacker notes, under our current conceptual scheme, to have a body is not a relationship of possession. Rather, talk of having a body is merely a way of referring to the corporeal characteristics of the human organism. See Hacker (2007, 269–84).

25. Note how our grammar has already shifted in this imagined case. In our grammar, it makes no sense to say that heads communicate with each other. We have to imagine a case for such a combination of words to acquire a sense.

26. Hacker (2007, 308) makes this suggestion.

27. The point derives from Wittgenstein (1958, 62).

28. Such a scenario would require more than transplantation between twins, of course.

29. See Hacker (2007, 310) for a comment to this effect in discussion of the thought experiment generally.

30. See Glock and Hyman (1994) from whom I have taken this point. It derives, however, from Wittgenstein (1958).

31. Where is the line drawn between a symptom and a criterion in the case of bodily continuity?, Parfit might ask. We can imagine starting with the fingerprints and progressively removing the rest of the body, leaving only the head.

32. Gender reassignment surgery might seem to cast doubt on this claim. The same person, after all, was male and now is female, or vice versa. But of course, in such cases, one and the same body is changed, and it remains numerically identical to itself throughout. That is not so, however, in the transplant example.

33. There are empirical grounds for contesting the claim that I discuss later.

34. The papers of Merker (2007) and Shewmon, Holmes, and Byrne (1999) are briefly discussed by Miller and Truog (2012) in support of their criticism of the higher brain standard of death.

35. Even if this should not prove to be so with adults, the cases of hydranencephaly should give us pause, and we can imaginatively vary the findings concerning the vertical plasticity of the brain stem to generate problems for Parfit's own imagined cases (see below).

36. This directly parallels what is currently known to happen in the case of single hemispheres, which take on some of the capacities of the one destroyed.

37. This point extends beyond the mere organism/person contrast. Assume chimpanzees are not persons. Are chimpanzees organisms or merely the conscious part of "their" organism? McMahan would say that they are the conscious part of their organism. The same issues would apply if we conducted similar thought experiments in the case of chimpanzees.

38. It is often questioned whether a PVS patient is a person rather than merely a living organism. This can be answered by a question: is a television set that has lost its picture a television set, or merely an electrical device?

39. Henceforth "brain" should be understood, as McMahan himself says, as shorthand for the relevant regions of the brain that McMahan understands as the "conscious part" that, according to McMahan, we are.

40. See Bennett and Hacker (2003, 445): "There are no criteria for whether the brain is conscious of this or that, only inductive correlations between brain states and the animal's being conscious, unconscious, or transitively conscious of this or that feature of its environment, or in one or another state of consciousness."

41. See Hacker (2007, 53) who also cites Wiggins (1995, 227) in support of this claim.

42. A similar point applies to the claim made by Parfit (2012, 12), that the brain might survive outside the body. Presumably, to survive, it would continue needing a blood supply, and once this is provided, we introduce things other than the brain, so it is really the brain *and* these other things that are *jointly* the enabling conditions for consciousness to continue. And once that step is taken, we can say that it is still not the *brain*, that is, the *organ*, that is conscious, but the entity consisting of the brain plus

that which supplies blood to it. Would this not at least show we are not organisms? It would only show that we might then *redefine* ourselves perhaps as something that is the *heir* of the human organism. But it wouldn't show that we are essentially brains, or the part of the brain in which consciousness is realized.

43. For the following points about behavioral criteria, I am once again indebted to the work of Peter Hacker (2007). For readers concerned that this account amounts to a form of behaviorism, see Hacker's rebuttal (Hacker, 1996, 254).

44. John Searle, in debate with Bennett and Hacker (2003), has pointed out that it is possible to be conscious and not show it. Bennett and Hacker concede this. An animal does not have to exhibit such behavior in order *to be* conscious. But they insist that *the concept* of consciousness is bound up with the behavioral grounds for ascribing it to a being (Bennett et al., 2007, 135). Only a being *capable* of showing consciousness is capable of *not* showing it. Hence, when a being is conscious but does not show it, it might be *acting like* a stone, but it is not *equivalent* to a stone, for stones are not capable of showing consciousness (and so are incapable of *not* showing it either). Similarly, those with locked-in syndrome *did* at one stage show it, and so are the kinds of beings of whom it makes sense to say that they are *not* capable of showing it—something has gone wrong, which is why they can no longer show it (except via the movements of the eyelids, which are still behavioral grounds for the expression of their consciousness and thought).

45. Parfit (2012, 20), for example, expressly rejects the captain in the ship analogy.

46. Cf. Hacker (2007, 306) who claims that to say that it is the brain that thinks, sees, hears, etc., is like saying that it is the jet engine that flies, rather than the aeroplane.

47. But note once again how much grammar has now shifted: we are now speaking of my "being in" an old body, and the like. These ways of speaking are excluded by our current rules. It makes no sense to say that I am "in" a body, though I can feel comfortable in my skin.

48. It is not at all clear, however, what could *count* as teaching a *brain* as opposed to the human being anything.

49. Parfit would respond to this contention by insisting that the word is *already* ambiguous: when I refer to being sunburnt all over, I am talking about my body. When I refer to thinking about a problem, by contrast, I am referring to me, whatever I may be. This reply, however, begs the question. As Hacker has argued, when I say that I am sunburnt all over, I am speaking of myself—this living human being. When I say that my body is sunburnt all over, I am speaking of one of my somatic characteristics.

50. Hacker has noted that what we should say is that, since these are criteria for a human being's thinking, they could only be used as criteria for the human being's brain to be in whatever state it is when the human being is thinking—and only *that* would be what it would then mean to say that my brain is thinking (personal communication).

51. It is also worth bearing in mind that scientific research has shown that pattern activation in the brain "that attends speech is largely indifferent to the content of what is said or meant, and to whether it is spoken or merely mentally rehearsed" (Nachev and Hacker, 2010, 70). This paper cites studies by Nachev, Kennard, and Husain, and by Picard and Strick. This research tends to contravene the otherwise natural intuition that there must be some one-to-one correlation between activity in the brain and the behavior that we go on when we ascribe the exercise of capacities such as thinking and speaking to the human being.

REFERENCES

Bennett, M., D. Dennett, P. Hacker, and J. Searle. 2007. *Neuroscience and Philosophy*. New York: Columbia University Press.

Bennett, M. and P. Hacker. 2003. *Philosophical Foundations of Neuroscience*. Oxford: Blackwell.

Broyles, J. 1985. Wittgenstein on personal identity: Some second thoughts. *Philosophical Investigations* 9:56–65.

Campbell, T. and J. McMahan. 2010. Animalism and the varieties of conjoined twinning. *Theoretical Medicine and Bioethics* 31:285–301.

Glock, H. J. and J. Hyman. 1994. Persons and their bodies. *Philosophical Investigations* 17:365–79.

Hacker, P. 1996. *Wittgenstein's Place in Twentieth Century Analytic Philosophy*. Oxford: Blackwell.

——. 2007. *Human Nature: The Categorial Framework*. Oxford: Wiley-Blackwell.

Khushf, G. 2010. A matter of respect: A defence of the dead donor rule and of "whole brain" criterion for determining death. *Journal of Medicine and Philosophy* 35:330–64.

McMahan, J. 2002. *The Ethics of Killing*. New York: Oxford University Press.

——. 2007. Killing embryos for stem cell research. *Metaphilosophy* 38:170–89.

Merker, B. 2007. Consciousness without a cerebral cortex: A challenge for neuroscience and medicine. *Behavioral and Brain Sciences* 30:63–81.

Miller, F. and R. Truog. 2012. *Death, Dying, and Organ Transplantation*. New York: Oxford University Press.

Mulhall, S. 2002. Fearful thoughts. *London Review of Books* 24:16–18.

Nachev, P. and P. Hacker. 2010. Covert cognition in the persistent vegetative state. *Progress in Neurobiology* 91:68–76.

Parfit, D. 1984. *Reasons and Persons*. Oxford: Oxford University Press.

——. 2012. We are not human beings. *Philosophy* 87:5–28.

Rundle, B. 2004. *Why There is Something Rather than Nothing*. Oxford: Oxford University Press.

Schechtman, M. 1990. Personhood and personal identity. *Journal of Philosophy* 87:71–92.

Shewmon, A., G. Holmes, and P. Byrne. 1999. Consciousness in congenitally decorticate children: Vegetative state as a self-fulfilling prophecy. *Developmental Medicine and Child Neurology* 41:364–74.

Singer, P. 2009. Reply to Don Marquis. In *Peter Singer Under Fire*, ed. J. Schaler, 153–63. Chicago: Open Court.

Wiggins, D. 1995. Substance. In *Philosophy - A Guide Through the Subject*, ed. A. Grayling, 214–49. Oxford: Oxford University Press.

——. 2001. *Sameness and Substance Renewed*. Cambridge: Cambridge University Press.

Wittgenstein, L. 1953. *Philosophical Investigations*. Oxford: Blackwell.

——. 1958. *Blue and Brown Books*. Oxford: Blackwell.