Identity and Freedom

Adam P. Taylor and David B. Hershenov

Introduction

It would certainly be unwelcome if one's preferred metaphysics of the person often makes free will and moral responsibility impossible. We contend that this is the fate of all the major materialist theories that understand human persons to be essentially thinking beings that physically overlap human animals. The source of the problem is that these accounts posit that there exist more than one entity possessing the same brain – the person and the animal. If the former can use the brain to think, so can the other. As a result, such theories are afflicted by *The Problem of Too Many Thinkers*.

Our focus is on an overlooked moral version of the Problem of Too Many Thinkers. If there is more than one overlapping thinker then there'll arise the problem that each cannot freely and responsibly act in a way that respects the exercise of the other's freedom. The trouble comes about because the overlapping thinkers can't simultaneously think and act on their interests and govern their lives in accordance with their values. And their interests and values will diverge due to their having different persistence conditions. We will show that there could be many occasions where only the person will give her free consent. The human animal overlapping the person won't freely consent to the same intention or action because he will either wrongly think that he is the person or will just be considering what is in the person's interests.

Our contention is that the only prominent materialist account to avoid such problems is the animalist theory that identifies the human person and the human animal and adopts a sparse ontology. We reach this conclusion in part because we accept a methodology in which ethical considerations and action-theoretic claims can weigh against certain metaphysical accounts of the person. So we aren't forced by methodological principles to claim that certain metaphysical conceptions of the person show that there is no free and responsible action; rather, we can plausibly

claim that practical considerations strongly suggest that those metaphysical approaches to personal identity are false.

One reason why we believe practical considerations should be included in the weighing of reasons in favor and against a metaphysical theory is that if there are moral truths then they must be consistent with metaphysical truths. If one adopts a metaphysic in which thinking beings overlap then one must reject certain seemingly obvious moral truths like we ought to respect the free choice and bodily integrity of beings like ourselves. We find it plausible that such considerations should tilt the scales against that metaphysics rather than show such core moral principles to be false. But even if one is an anti-realist about ethics and thus under no pressure to make metaphysical truths cohere with moral truths for there are not any of the latter, there are still true action-theoretic claims about free action that we think a metaphysic should accommodate. For instance, if a metaphysic makes it impossible for everyone intelligent enough to understand these sentences to also freely endorse and act upon their interests and values, then that too provides a reason to doubt the metaphysics rather than accept the impossibility of such creatures being free.

The Moral Problem of Too Many Thinkers

We believe that greater success in resolving *The Problem of Too Many Thinkers* is the closest there is to a criterion to choose between competing metaphysical theories of the person. If a theory implies that there exists more than one thinker under your clothes, then that is a major strike against the theory. This reason may not be strong enough on its own to warrant rejecting the theory outright, but it will greatly weigh against it, tilting the scales further in the direction of a view that avoids the problem.

Let's assume that persons are essentially thinking beings that are spatially coincident but numerically distinct from animals. The problem which arises is that if the person can think, then it would seem that the animal should also be able to think since it shares the same brain. Olson

highlights the epistemic problem that arises if both the person and the animal can think, then you have no reason to think you are the person rather than the animal. Either you or your co-located thinker will be wrong when you both say "I am essentially a person." The first person pronoun refers to each of the speakers and so one is falsely predicating personhood to itself. How can you be so sure that you are not the deluded animal making the erroneous self-ascription?

Noonan attempts to avoid the epistemic problem by endorsing what has become known as *Pronoun Revisionism*. Noonan suggests that to have thoughts about one's thoughts is not enough to make an entity a person, rather an individual must have the appropriate psychological persistence conditions. That is, the person goes out of existence if he loses certain psychological capabilities. So the thinking animal is never a referent of the personal pronoun "I" for the term doesn't pick out any entity thinking or uttering the word "I", rather it just refers to the person.

However, even if Noonan is right about the animal's use of the personal pronoun, this will not be enough to mitigate the ethical problems. While we'll draw mostly upon bioethical examples involving what is known as informed consent, readers can easily imagine similar scenarios in other domains that would likewise undermine free will. If there are non-persons such as human animals that can't refer to themselves with the first-person pronoun, then how can they be said to freely agree to any immediate treatment or make provisions for their future with say a living will? While we don't have a favored theory of free will to expound, it would seem safe to say that one couldn't be free if unable to reflect upon one's interests, desires, values, intentions as *one's* own, and then act on the basis of the reasons they provide. If we assume pronoun revisionism, a problem is that the animal is thinking about the person's interests as the person does, for the animal refers to the person when it uses first person pronouns to entertain thoughts of the following types: "*I* would prefer such and such a course of treatment for it increases *my* well-being" or "I would prefer to forgo all treatments so I can lead the remainder of my life in accordance with my values." Our worry is that

the animal might have interests and values that are not the same as those of the person because of their different persistent conditions. Thus the animal's choosing to act on the basis of what the person has reason to do cannot be understood to be a free and responsible action of the animal who didn't reflect upon the action qua animal, i.e., did not think of himself as an animal engaged in that action. A similar lesson can be drawn in the absence of pronoun revisionism if the animal uses the first person pronoun to self-refer but is ignorant of his kind membership, wrongly thinking he is the person. He will be choosing actions on the basis of the person's interests and values and so the action cannot be said to emerge from him in a way characteristic of freedom.

We'll provide examples below in which overlapping thinkers won't be self-governing on any of the leading theories of free will. On psychological accounts like those of Noonan, Shoemaker, Parfit, Baker and Hudson, persons go out of existence with the loss of certain sophisticated psychological capacities while human animals can survive the loss of such capacities, existing in impaired mental states. We contend that the animal might have an interest in continuing to exist in a childlike state after the rational, self-conscious person has ceased to exist as a result of injury or stroke. Conversely, persons might have interests that are not strictly those of our animals. A conflict can be generated if there is an experimental drug that may prevent the further decline into Alzheimer's disease, but will far more likely kill its user. The person, who inevitably goes out of existence with the loss of self-consciousness, might think she has nothing to lose since either the disease or the drug's unwanted side effect will end her existence. However, it may be in the interest of the human animal not to take the drug since it could survive with the minimal sentience of late stage Alzheimer's disease.

We don't think such a choice could be considered free for *both* the animal and the person on any of the leading accounts of freedom. It doesn't matter if such accounts stress the endorsement of desires that we act on by higher order desires or values (Dworkin, Frankfurt, Watson), emphasize

the history of how those higher order attitudes arose (Wolf), insist upon choices meshing with long term plans (Bratman), require a reason responsiveness and a mechanism that is sensitive to reasons (Fischer and Ravizza), or insist that the agent exercise his causal power to choose between alternatives regardless of antecedent circumstances (Clark, O'Connor). The overlapping thinkers in the above and below scenarios could consider in succession that they were the person and then the animal, the result being that if they first each thought they were an animal they would endorse different acts, be alienated from different parts of a shared history, have divergent long term plans, be sensitive to different reasons, and exercise agent causation differently from how they would if they thought they were the person.

The Alzheimer's drug isn't the only scenario where free will cannot be exercised by overlapping thinkers. Conflicts between the person and the animal could prevent them from both *freely* endorsing the same advanced directive. For instance, the person may not want his resources to be spent on sustaining an organism with dementia with whom it is not identical. That person would have written a very different advanced directive if he had thought he was the animal. Or the person might leave directions to try an extremely dangerous experimental treatment if his Alzheimer's Disease progresses to a certain state. But the treatment would be contrary to the animal's interests. So the advanced directive written by the animal and the person while the animal thought it was the person or only considered the latter's interests will not do justice to the animal's freedom.

We can generate other infringements due to the animal and person's different interests due to their different persistence conditions. Assume the person and the animal both support donating organs at their deaths but not before. Let's add that they even believe they are morally obligated to engage in *directed donation* and bestow organs upon an ailing friend or relative after they die. However, the possibility of the animal and person's deaths occurring at different times could prevent the full realization of their seemingly shared value. The person is essentially a thinking being and the animal

is essentially a living being and so the criteria for their deaths would diverge. The problem is that when the animal dies after its respiration and circulation have irreversibly ceased, less of its organs may be viable for transplant than if organs were taken when just consciousness was lost irreversibly with the onset of a persistent vegetative state. And in the case of directed donation, the person might not be able to donate upon her death because the organism is still alive and doesn't pass away until long after the needy friend or relative dies.

It is important to realize that these types of conflicts aren't the standard conflicts of interests between free parties where say a government health official doesn't allow the person to have the experimental drug that he covets, or a court rules that a health insurance company needn't provide payment for the expensive treatment that the patient is interested in, or a doctor refuses to undertake the risky procedure that the patient wants. Each of these individuals can freely formulate an intention and act upon it even if someone else later prevents their action from producing the desired results: the acquisition of the drug requested, the petitioned for payment, the provision of the sought after procedure. Rather, it is impossible for the overlapping animal and person to *simultaneously* freely endorse an intention that *both* then act upon. Nor do they each have free control over a personal realm, their body. While you can choose to take a risky experimental drug that I can refuse to take, the spatially located person and animal cannot each make and act on their own choice. If one takes the drug, the other does so as well. If the person donates multiple organs upon his death with the loss of the appropriate mental capacities, the animal will be killed when his vital organs are taken. Conflicts like these make it impossible to respect the bodily integrity of both. So we are not presenting just another instance of the typical problem where someone's freely endorsed intentions and acts are foiled by the freely produced preferences and deeds of others in the society – and without any rights being violated.

We don't think one can escape this by appealing to Parfits' famous claim that identity isn't what matters, our prudential-like concern being only with psychological connections to a future person regardless of whether we are one and the same person or distinct persons. If Parfit were right, the overlapping human animal and person would have the same interests and thus would not choose differently. Parfit bases his claim on cases like those involving Adam's cerebrum fissioning and both cerebral hemispheres transplanted into different bodies B and C. Adam would have survived if just one of the hemispheres was successfully transplanted, the other destroyed upon removal. So Parfit reasons that having these two equally good psychological successors is as good as ordinary survival (no fission and no transplants). The claim that identity doesn't matter depends upon Parfit holding an account of identity involving a uniqueness clause. Parfit's criterion for personal identity across time is that it consists of i) the appropriate psychological relation R and ii) being uniquely the possessor of such relations. Since this *uniqueness clause*, aka no *branching clause*, is trivial and satisfied by what is extrinsic to us, Parfit insists it can't be what matters to us. So it must be the other component of the personal identity criterion, the psychological relation R, that matters to us. This led Parfit to his famous conclusion that identity doesn't matter.

One reason for our skepticism about Parfit's thesis is that it doesn't mesh with our reactions to torture and death following a great change in our psychology due to a stroke. It doesn't seem that we now will view the later torment and death as being less bad since we are less psychologically connected to the being after the stroke than we would be if there had been no damaging stroke.

Secondly, Parfit's claim that identity doesn't matter depends upon his interpretation of a fission scenario that violates the rationale behind *the only x and y rule.* That rule does not allow that our identity in the future can be determined by whether there are two or more equally good candidates as there would be in cases of fission. The rule restricts questions of whether x is identical to y to the internal relationship between x and y, the existence of a z being irrelevant. The rationale

for the rule is that there should not be unexplained existences where entities owe their existence to other beings despite the absence of a causal connection between them (Hawley, 2005). The problem with Parfit's cerebral fission and transplant case is that the person in body B would not be there if it wasn't for the existence of the person in body C likewise being psychologically continuous with Adam. So the person in Body B owes his existence to the person in body C, and vice versa, but there are no causal connections between person in body B and the person in body C despite the existence of each playing a role in the creation or sustaining of the other. So if unexplained existences are to be avoided, then the criterion for identity should not involve a uniqueness rule and psychological relation R. But it is only the extrinsic and trivial features of the uniqueness clause that leads Parfit to the conclusion that only psychological relations matter. If he is not allowed to introduce the uniqueness rule into the account of identity, then fission can't show that identity doesn't matter, merely psychological relation R is of importance. So Parfit's thesis can't save autonomy in the above cases by giving the overlapping thinkers the same interests.

Conclusion

Thus if you're a materialist and care about freedom and responsibility, then you'd better identify yourself with your animal. So this gives us an additional and rather weighty reason to put on the metaphysical scale, perhaps tilting it in favor of the view that we are animals. The animal is the person. And there aren't any other thinkers overlapping the animal.

If you don't believe that moral or action-theoretic considerations should be given any weight when considering rival metaphysics, then we suggest that you come up with a radically new ethics. It will be an ethics that downplays satisfying free choice, and autonomous control over one's body due to the recognition of the divergent interests and values of overlapping entities. The new ethics will recommend some sort of compromise between the interests and values of those individuals now being counted by the latest metaphysical census, no doubt abandoning many of our currently

established rights in the process. But that is the topic for another paper, or rather book, and one that we hope no one ever has to write.