

Can Ordinary Materialists be Autonomous?

Abstract

The traditional problem for materialists is to account for how matter could give rise to thought. But however materialists fill in the explanatory gap, there remains a further threat to their materialism due to the ironic possibility that there's more than one material thinking being overlapping each of them. This is 'The Problem of Too Many Thinkers.' Overlooked in discussions about this problem is the danger it poses to our moral lives. It isn't clear that the secular can offer a materialist account that makes autonomy possible. The materialist can accommodate what seem to be truths about respecting the freedom and autonomy of creatures like us only by accepting a very counterintuitive sparse ontology. Immaterial accounts of the person look good by comparison with the costs of materialism. However, those immaterialist theories that don't posit a divine maker of the soul suffer from certain metaphysical puzzles avoided by those who do claim God created one's soul. A soul that requires a divine maker strongly suggests that such immaterial beings were made for a purpose. Such a purposeful creation makes a theistic ethics seem far more plausible.

I. Introduction

We defend the rather ironic thesis that it is secular materialists rather than the religious who cannot provide an ethics that enables us to be autonomous. The irony is that a theistic ethics is often decried as heteronomous. However, we suspect that the most autonomy that one can hope for will involve acting in accordance with a divine plan or, in other words, functioning as we are designed to do. We'll say little about what that might involve. Instead, we'll concentrate upon a negative conclusion: why the materialist metaphysic assumed by nearly all secular theorists is unable to account for how creatures like us could be autonomous. This heteronomy is due to the mainstream materialist conception of persons suffering from *The Problem of Too Many Thinkers*.¹ If any material

¹ Shoemaker coined the phrase "The Problem of Too Many Minds" to describe the problem of the animal beings coincident with the person. We prefer "The Problem of Too Many Thinkers" in case two thinkers can share a mind in the way that one's elbow and arm can share one and the same bruise. Sydney Shoemaker and Galen Strawson, "Self and Body." *Proceedings of the Aristotelian Society, Supplementary Volumes* 73. Wiley (1999): 287–332.

entity can think, then there will be too many overlapping thinkers.² We understand *materialist* persons to be subjects of thought that are *ultimately* composed of only the kinds of things that physics studies – atoms, electrons, quarks, energy etc.³ So regardless of whether human persons are conceived of as identical to human animals (animalist theory) or as being distinct but sharing their matter with human animals (constitution theory) or as proper parts of animals (embodied mind theory), they will be material thinkers.⁴ Such entities will differ in persistence conditions, that is, the features that must be retained for them to continue to exist into the future.⁵ Their different persistence conditions will produce conflicting interests that they can't all autonomously pursue. Our autonomy wouldn't be threatened if we think in virtue of an immaterial soul. We'll argue that souls need a divine creator and that renders a theistic ethics as the most plausible moral guide.

² We will also be drawing upon Unger's 'Problem of the Many Mental' where thinking can overlap without having all their proper parts in common. Peter Unger, "The Mental Problems of the Many." In *Oxford Studies in Metaphysics* 1, ed. Dean Zimmerman (Oxford: Oxford University Press 2004), 195-222.

³ That is, even if there are parts of material thinkers such as brains that are not directly the topics of investigation for physics, they are still material because their parts are built up out of entities that physics studies. The division of a material being's parts into further parts will ultimately end with the entities of fundamental physics. Thus material thinkers will only have parts that are themselves ultimately composed of or identical to the entities that the fundamental physical science investigates.

⁴ Eric Olson provides the most thorough defender of animalist conception of persons in his 1997 *The Human Animal: Identity without Psychology*, Oxford University Press. Lynne Baker has produced the most extensive defense of the constitution theory of persons in her 2000. *Persons and Bodies: A Constitution View*, Cambridge University Press. Jeff McMahan had most forcefully defended the embodied mind theory of persons in his 2002 *The Ethics of Killing: Problems at the Margins of Life*, Oxford University Press.

⁵ An animal will persist as long as life processes continue, a person that is constituted by an animal will persist as long as it has certain capacities for thought.

There seems to be no denying that every thought or feeling has a thinker or feeler and every action has an agent. We can't make sense of the idea that experiences exist without experiencing subjects. As Strawson wrote "experience is necessarily experience for someone or something... Experience necessarily involves experiential 'what-it-is-likeness', and experiential what-it-is-likeness is necessarily what it-is-likeness for someone-or something."⁶ Pains can't float free of their subject. There must be a subject in pain for there to be pain. In fact, an enduring pain is more unwelcome if one believes that one will be the persisting subject of that continuous pain into the future. The prospect of enduring pain that will either not have a subject or a different subject than oneself in the future is not as disturbing, nor does it provide the same motivation to act, as the concern that one will be the subject of that lasting pain.

If I am a person in pain and there is a numerically distinct organism sharing my brain and perhaps all the rest of my matter, then that entity is also in pain. Likewise, for the head or brain or a cerebrum that is embedded within the person or animal in pain. This proliferation of overlapping entities in pain is counterintuitive in a way that positing smaller tables or clouds embedded in larger tables and clouds is not. It won't even be that absurd to discover that what we thought was our body was actually many overlapping bodies. When we are dealing with puzzles of composition and trying to determine what parts compose a larger entity we are not terribly bothered to discover that nearly all of the atoms composing one candidate cloud or body also compose a distinct but overlapping cloud or body. But overlapping experiencers and agents seems preposterous in ways a proliferation of tables and clouds and even human bodies does not.⁷

⁶ Galen Strawson, "What is the Relationship Between an Experience, the Subject of Experience, and the Content of the Experiences." *Philosophical Issues*. 13:1, 280.

⁷ Unger. Peter. 1980. "The Problem of the Many." *Midwest Studies in Philosophy*. 5:1. 411-468.

The problem is not just an increase of the entities having experiences but the moral quandaries that result. These are in part due to overlapping thinkers giving rise to some epistemic problems. If you think you're a very different kind of entity than you are - with a nature, persistence conditions and interests distinct from those that you really possess - then you'll often fail to act fully autonomously. An initial example without overlap may help. If you think you are an immortal angel but you are really an elderly mortal, then your actions won't often be fully autonomous because you have confused yourself with a different kind of entity that has a very different life span and interests. You can't control your life and live in accordance with your values if your self-conception is so riddled with falsehoods.⁸ Other errors may not affect your autonomy if they don't involve a difference in persistence conditions and interests. For example, if you think you are a compound of a soul and a body and you are actually just a soul controlling a body that is not literally amongst your parts, your autonomy may not be affected since the chronology of your earthly persistence on either account is the same and the same disembodied afterlife awaits you. A more significant threat to your autonomy will be if you are a thinking being that overlaps another thinking being and wrong about which one you are. If you think you are an essentially self-conscious being and you are instead an

⁸ In making this point, we are assuming that only a certain degree of *self*-knowledge is necessary for autonomous, i.e. *self*-determining action. We are not denying that one can be autonomous when acting against one's interests especially when one does so in accordance with one's values and when there is no malfunctioning or bypassing of one's cognitive machinery. Autonomy is no guarantee of maximizing well-being. It may even be, as a referee suggests, that one can act autonomously in pursuit of one's interests while wrongly thinking those interests are imposed but they aren't. As long as those interests and resulting intentions are not bypassing the subject's proper mental functioning as they would be in the case of manipulation (hypnotism, neural implants etc.) they could be autonomous and *owned* by the subject, even if the agent is wrong about whether he is autonomous. Just as believing one is autonomous is no guarantee that one is, thinking that one is not autonomous is no guarantee that one is not. But if one thinks that one is someone else then that would seem to fail to meet a minimal self-ownership condition for self-determination.

animal that could survive the loss of self-consciousness, then there will be times where your actions are not autonomous because you are deciding how to act upon the basis of essential traits, persistence conditions, and goods that belong to someone else and are distinct from yours own.

Moreover, even if you don't make such errors about yourself but just decide what to do without considering your own interests *as your own*, your actions will often be heteronomous. That is, you may know that there is a spatially coincident organism and person and don't wrongly claim to be a person when you are an organism. There is still a problem if you and the person decide what to do on the basis of the person's interests or vice versa. That is you both think it would be in the interest of your person to do such and such and then an intention is formed to do such and such. The animal was not thinking of those interests as its own animal interests when the intention was formed to do such and such. This will become clearer below when we discuss different entities the first person pronoun can take.

Our contention is that this threat to autonomy is unavoidable for the materialist who believes that human persons overlap but are distinct from human animals. So if one is a materialist, then a necessary condition to avoid the problem is to accept the animalist identification of the animal and person. Otherwise the problem will arise regardless of how one construes the overlap of the numerically distinct person and the animal. They may be spatially coincident and consist of all the same atoms. Or Lynne Baker may be correct that the person is constituted by the animal akin to how the statue is constituted by the lump of clay.⁹ When the lump is arranged a certain way by the sculptor, the statue emerges. Analogously, when the brain of the organism develops in a certain manner, the person arises. Nor will it matter if the materialist thinks the person is roughly just a brain-size part of the organism as McMahan and Parfit believe.¹⁰ The problems will also arise if one

⁹ Lynne Baker. 2000. *Op. cit.*

¹⁰ Jeff McMahan. *Op. cit.* Derek Parfit, "We Are Not Human Beings." *Philosophy* 87, no. 1 (2012): 5-28.

believes with Hudson that persons and organisms are perduring four-dimensional objects that share some but not all their temporal parts.¹¹ On such views, persons and organisms are more like events, extended in time. The person would consist of the self-conscious stages of the organism while the organism would include mindless states as well as the self-conscious stages.

The problem arises because the person and the animal possess the same brain and thus it would seem that if one can think so can the other. If the first-person pronoun has a Kaplanian character and refers to each user of it, then either the animal will be in error when he and the overlapping person assert “I am essentially a person” or the person will be wrong when he and his animal claim “I am essentially an animal.” It won’t help if the referent of the first-person pronoun switches in different contexts from the person to the animal, or if the first-person pronoun is always used by the animal and person to refer to the person as Noonan conjectured. Either way, one of the two overlapping beings will be deciding upon particular actions in ignorance of his kind and persistence conditions and consequently often his interests. We contend that such widespread heteronomy can be avoided only by adopting a very counterintuitive sparse materialist ontology or a more plausible soul theory.

We will describe a number of cases in which both the animal and the person wouldn’t autonomously endorse the same action. These include the staple of the personal identity thought experiment literature - the brain transplant. Imagine the case of a cerebrum transplant undertaken to avoid the increasing burdens of a deteriorating body. It may be in the person’s interest to get a new body but not in the animal’s interest to be left behind mindless or destroyed. Less futuristically, imagine there’s a risky Alzheimer’s drug that would more likely kill the person just diagnosed as being in the early stages of the disease than prevent the disease from worsening. The essentially self-conscious person has little to lose for he soon ceases to exist either from the disease progressing and

¹¹ Hud Hudson, *A Materialist Metaphysics of the Human Person* (Ithaca: Cornell University Press, 2001).

destroying his reflective capacities or the lethal side effects of the drug. But the animal, on the other hand, could survive without being self-conscious in a childlike state of mere sentience and thus has more at risk in taking that drug. It is easy to generate other autonomy undermining conflicts between persons and animals due to their dying at different times. We'll sketch scenarios that reveal their divergent attitudes about paying for long-term care, the appropriate time to transplant their vital organs, or when to bury their bodies, grieve, inherit, and remarry. Advanced directives dealing with any of them will not be autonomously endorsed by all who sign them.

We'll next show that moral claims about respecting autonomy can only be preserved by the materialist through adopting the most counterintuitive sparse ontology where the only existing composites are either organisms or that part of the brain directly involved in the production of thought. Both views not only leave us without the vast majority of the objects of folk ontology and the special sciences, but are at odds with how we individuate thinkers in a number of cases, real and imagined.

We then argue that divinely created souls can avoid certain metaphysical puzzles that plague theories of non-divinely created souls that emerge on their own and don't require God to pair them to bodies. We conclude that if the sole solution to these problems is a divinely created soul, then it's likely that human souls were made for a purpose and thus an ethics devised by their creator will be the correct moral guide.

II. Unger's 'Problem of the Thinking Many'

The materialist typically assumes a person consists of atoms but when we consider the microscopic arrangements of those atoms there seem to be many equally good candidates for being the aggregate of atoms that compose the person. If one candidate aggregate composes a thinking being, yourself, why doesn't another with just a few more or less atoms also compose a thinking

being? This is Unger's 'Problem of the Thinking Many'.¹² One version of this problem stresses that there would seem to be a countless number of conscious beings feeling pain and pleasure and other qualia where you are. Unger calls this the 'Problem of the Experiential Many'. Of course, one could just accept that there are countless beings experiencing phenomena where you are but that is extremely counterintuitive.

Accepting many thinking beings overlapping oneself would also render something akin to libertarian free will impossible since that involves a person being able to act differently from how others act at the same time; but the overlapping thinkers will never diverge from each other. Unger claims that even more disconcerting than the 'Problem of the Experiential Many' is what he calls 'The Problem of Too Many Choosers'. Hudson calls it 'The Problem of the Choosing Brothers' since all of the beings overlapping him would be his genetic twins.¹³ Unger and Hudson argue that the problem raises doubts about our being free and what they say can be extended to autonomy. Differences between free will and autonomy are not a major concern for us here. As John Fischer has noted, the free will and autonomy literature often runs enough in tandem for our purposes.¹⁴ We are mostly interested in control over one's actions and life and if someone would have grounds for complaint that their autonomy was not being respected. We are less focused on questions about who is responsible for what they do to others. However, we provide a framework where overlapping thinkers will, at times, not both be responsible for what they do to others.

Unger and Hudson aren't just arguing that the laws of nature and the past don't determine one's present and future behavior. It's not enough to have the ability to have chosen otherwise. If

¹² Peter Unger. "The Mental Problems of the Many." In *Oxford Studies in Metaphysics* 1, ed. Dean Zimmerman (Oxford: Oxford University Press 2004), 195-222.

¹³ Hud Hudson, *A Materialist Metaphysics of the Human Person* (Ithaca: Cornell University Press, 2001).

¹⁴ Personal correspondence

one person shares the same brain structures with many overlapping persons, it doesn't seem as if any one of the many thinkers can choose differently from any of the others even if the laws and past doesn't determine the choice. Each person's choice is still "determined" to be in lockstep with that of the others. Unger thus adds to the classical libertarian idea that history and laws don't determine your choices, that if you are truly free, then your choice won't be determined by anyone else's. So if free will or autonomy means your choice is independent of the choices of others, then the inability of the embedded thinkers to choose differently from those that they overlap means they all lack freedom. It defies the odds to think each overlapping being was free to do what he wanted and freely chose the same course of action. But if you are one of many overlapping thinkers, you all will always be choosing the same, even if that choice isn't determined as a result of the initial conditions and deterministic laws. None of you could ever choose otherwise. How do we know that? Well, it doesn't show up in any of our actions. Your body always, or nearly always, speaks and moves as you intend. So if you and your brothers are autonomous in the sense that your actions are all independent of others' choices, then countless thinkers overlapping you are participating in a cosmic miracle of choosing exactly as you do. The implausibility of this seems to Unger to clearly show the falsehood of any materialist conception of the person.

Soul theories not only prevent countless thinkers where we would like just one, but they can make more sense of how we can act freely than the libertarian materialist. The latter claims that indeterminate laws provide the necessary leeway for free or autonomous actions. Such approaches still limit our choices to being in accordance with probabilistic laws. But one might want to decide to go right more than say 50% of the time. (We are assuming an analogue at the level of macro agency to the probability that a particle might go through the right slit rather than the left.) A libertarian doesn't want one's conduct limited by probabilistic laws. So this means our freedom will involve violating laws. It is hard to explicate our idea, but it seems to make more sense for a nonmaterial soul

to violate physical law than a material entity. It seems to us that something nonphysical can be a better source of a violation of a physical law than can a physical entity. If an entity is outside of the world of physical objects, it is less mysterious that it could intervene in a way that violates physical law than an object already within the physical world. We can't explicate this intuition any further, though it is perhaps worth mentioning our experience is that others share it.¹⁵

Now the above arguments render materialism or physicalism¹⁶ unattractive only if one is already persuaded that free will, autonomy and moral responsibility require that individuals are capable of choosing differently from others.¹⁷ However, our self-conception as free, autonomous and responsible may not need such a capacity. Many a philosopher accepts a compatibilist conception of free will or autonomy or a semi-compatibilist account of responsibility. But we will see in section III that even compatibilists will not be able to easily maintain their favored conception of freedom and autonomy and a commitment to materialism. And regardless of the implications for free will and autonomy, they should also be bothered by sharing a brain with countless other experiencers.

III. Why the Standard Materialist can't Account for Our Autonomy

Advocates of a neo-Lockean approach to personal identity maintain that persons go out of existence when they lose their capacity for self-consciousness. They acknowledge that human animals, on the other hand, can persist without being self-conscious and instead go out of existence when life processes irreversibly cease. Our contention is that both the overlapping animal and the

¹⁵ This intuition about lawbreaking is shared by the other philosophers in our free will and moral responsibility reading group "Blameless Buffalo?" – Steve Kershner, John Keller, Robert Kelly, David Limbaugh, and Yishai Cohen - though none of them share our sympathy for soul theory. Peter Unger has defended something akin to this in his "Free Will and Scientificalism" *Philosophy and Phenomenological Research* 65 (1):1-25.

¹⁶ We are not distinguishing physicalism from materialism.

¹⁷ We will see a threat to even compatibilist freedom or autonomy later in the paper but it isn't due to Unger's 'Problem of the Thinking Many'.

person can think and that their interests will often diverge as a result of their persistence conditions differing. A consequence of this is that they can't both autonomously endorse the same actions.

Consider cerebrum transplants undertaken to avoid the increasing burdens of a deteriorating body – or the transhumanist vision of inorganic replacements of living body's parts undertaken for motives similar to those of transplant. It may be in the person's interest to get a new body but it is not in the animal's interest to be left behind mindless in the transplant case or destroyed with the inorganic part replacements. (If the brain can think, then undergoing the transplant but not the inorganic part replacements will be in its interest.) Less futuristic, consider a risky Alzheimer's drug that would more likely kill the person in the early stages of the disease rather than prevent the disease from worsening. The essentially self-conscious person has little to lose for it soon ceases to exist either from the disease progressing if not the drug's lethal side effects. But the animal could survive without being self-conscious in a childlike state of mere sentience and so risks more.

Consider a person making an advanced directive that will be in effect once the human animal possessing just minimal sentience comes to exist without overlapping or constituting the person any longer. The person may have preferred her resources not being exhausted on a being that she is not identical to but the animal could very well benefit from a pleasant but cognitively diminished existence. The person is making an advanced directive that will be in effect once the human animal comes to possess just minimal sentience due to disease or injury and exists and no longer overlaps or constitutes the person. The person is deciding the future of a being with whom she is numerically distinct. The person may have preferred her resources not being exhausted on a being that she is not identical to but the animal could very well benefit from a pleasant but cognitively diminished existence.

Other examples will involve disposal of the body or its parts. Some religions demand speedy funerals that cannot be satisfied if the person and animal die at different times. The person may be

extinguished while the animal lingers. In the transhumanist future, inorganic part replacement or cerebrum transplants might mean that the animal is extinguished and the person survives. Thus the person and the animal cannot always both be buried immediately after their deaths. Attempting to do so will in the non-futuristic cases cause the death of individual that has initially survived the destruction of the being that overlaps it. A similar problem arise for transplanting the organs of the deceased. Multiple vital organ transplants of the deceased person will bring about the death of the human animal. Waiting for the death of the animal, which will require either meeting the whole brain death or the criterion of circulatory respiratory cessation, will mean that it might be too late to direct some organs to loved one's or others then in need.

Moreover, there is no guarantee that person and animal will agree about the appropriate timing of other so called "death behavior". When should the spouse begin to mourn? Should the spouse mourn twice? When could spouses begin to date or remarry, assuming that they so care to. And when do they and others in the family inherit? It might not seem like the person and animal will disagree for they both give lip service to the time being their death. But they may die at different times. So readers can see that a commitment to "until death do us part" is not as helpful a guide as one might have thought. Perhaps the animal will believe that it should be allowed to die when its brain is irreversibly non-cognitive; if so, then the animal's death will soon follow the person's and the autonomy of both can be satisfied. But if they both believe that their own death is the appropriate time and are opposed to hastening death, there is no such autonomy respecting solution.

We shall discuss below in some detail the possibility that the person is just a proper part of the organism. The view we will consider maintains that persons may be brain size and their body is not literally a part of themselves. What we want to do now is show how this leads to a space based rather than temporal divergence of interests. The small person may have a different attitude to her body than if she literally was identical to her body. An autonomy conflict could arise in the classic

lumpectomy vs. mastectomy debate in cancer care. If a woman's breast were literally a part of her, it might be harder to remove it than if she did not stand in such a relation to it. Such conflicts might seem far-fetched at first thought. The typical organism and person never believe that they are distinct and stand in a different relation to their body and so would have initially the same attitude to their body. But readers should imagine their different reactions to parts of their body if they were to find out what they thought was literally part of them wasn't. Imagine that what you thought was a part of yourself was really a prosthetic beyond your own boundaries. Your concern for that object may well lessen, at least after a certain time.

There are what we will call 'metaphysically acceptable' and 'metaphysically unacceptable' conflicts of autonomy. The above conflicts between coinciding animals and persons differ from the metaphysically acceptable conflicts that you and I could have over politics. We can simultaneously endorse, intend and pursue conflicting ends but that isn't the case for coincident animal and person. You and I also have a realm (our respective bodies) in which we are autonomous, you can autonomously take the risky experimental drug that I can autonomously refuse. You can donate your organs while I can keep mine. You can choose a less invasive breast operation while I opt for a more invasive one. But such different displays of autonomy are denied to the coinciding person and animal – or the overlapping person and brain.

We don't think such choices could be considered free for *both* the animal and the person on any of the leading accounts of freedom and autonomy. It doesn't matter if such accounts stress the endorsement of desires that we act on by higher order desires or values.¹⁸ Nor will it help to emphasize the history of how those higher order attitudes arose or to insist upon choices meshing

¹⁸ Gerald Dworkin, "Acting Freely." *Noûs* 4 (1970): 367-83; Harry Frankfurt, *The Importance of What We Care About* (Cambridge University Press 1988); Gary Watson, "Free Agency." *Journal of Philosophy*, 72 (1975): 205-20.

with long term plans.¹⁹ Autonomy and freedom are not spared by insisting upon or requiring reason-responsiveness and a mechanism that is sensitive to reasons.²⁰ Nor will they be saved by requiring that the agent exercise his causal power to choose between alternatives regardless of antecedent circumstances.²¹ 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.

IV. The Irrelevance of Self-Referential Mechanisms to the Problem of Autonomy

Olson and most others stress the epistemic aspects of the Problem of Too Many Thinkers.²² We contend the major problem is an ethical one and that will remain even if Olson's epistemic problem is dealt with perhaps by a theory like that of Harold Noonan.²³ Noonan attempts to avoid the epistemic problem by endorsing what has become known as *Pronoun Revisionism*. Noonan

¹⁹ John Martin Fischer, *My Way: Essays on Moral Responsibility* (Oxford: Oxford University Press 2006); Susan Wolf, *Freedom within Reason* (New York: Oxford University Press 1990); Michael Bratman, "Planning Agency, Autonomous Agency." In *Personal Autonomy: New Essays on Personal Autonomy and Its Role in Contemporary Moral Philosophy*, ed. J. Taylor (Cambridge: Cambridge University Press 2005) 33-57; John Martin Fischer and M. Ravizza, *Perspectives on Moral Responsibility* (Ithaca: Cornell University Press 1993).

²⁰ John Martin Fischer and M. Ravizza, *Perspectives on Moral Responsibility* (Ithaca: Cornell University Press 1993).

²¹ Clarke, Randolph, "Toward a credible agent-causal account of free will." *Noûs* 27 (2) (1993): 191-203. Timothy O'Connor, "Agent causation." In *Agents, Causes, and Events: Essays on Indeterminism and Free Will*, ed. Timothy O'Connor (New York: Oxford University Press) 61-79.

²² But see Buford and Brueckner for an epistemological argument that challenges Olson's claim that there is an epistemic problem keeping the person from knowing that she is a person rather than the coincident animal. Anthony Brueckner and Christopher T. Buford. "Contextualism, SSI and the Factivity problem." *Analysis* (2009): 431-438.

²³ Harold W. Noonan, *Personal Identity* (London: Routledge 1989).

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. "I," after all, is a *personal* pronoun. That it refers to a person is analytic, i.e., a truth given by the meaning of the word. Yet it remains open to object that "person" is a phase sortal and is applied to beings that are not essentially persons who could cease to instantiate personhood without ceasing to exist, just as a student can cease to be a student without ceasing to exist. This is the view of animalists who maintain that the typical human animal is not always a person, not being one when mentally immature or cognitively impaired. They hold that the animal is identical to the person, but only contingently instantiates the property of personhood. Their rivals, the defenders of various psychological approaches to personal identity, believe the person is an entity numerically distinct from the animal it overlaps. Noonan thinks that the latter are right and that philosophical arguments, perhaps ones relying heavily upon thought experiments, will indicate that we persons have essentially psychological persistence conditions and thus are distinct from animals.

Pronoun revisionism means that the animal wouldn't know of his interests under the mode of presentation required for autonomy. But even if a first person pronoun "Schmy" was introduced for the animal and person to refer to the animal similarly to the direct way that Noonan claims they both use "I" to refer to the person, they couldn't *simultaneously* endorse a preference or intention. If the person announces that "I will take the risky Alzheimer's drug" and then acts on that intention, the overlapping human animal cannot proclaim "Schmy will not take the risky Alzheimer's drug" and have his autonomy respected by the overlapping person's actions.

Nor will it help matters if the reference of “I” switches in contexts in a manner similar to the demonstrative ‘*this*’. One might point to the spatially coincident sculpture and lump and say “*This* used to be in a quarry” and mean to be referring to the lump and not the statue. Or one might point in the same direction and say “*This* came into existence when George Segal carved it last May” and be referring to the statue and not the spatially coincident lump. However, such contextual shifts will not enable the non-referred to thinker to act autonomously. It would just mean that sometimes the person’s autonomy is accommodated and on other occasions the animal’s autonomy is accommodated. *Either way, one of the two doesn’t exercise control over its life qua animal or qua person, and thus fails to act in accordance with its own values and interests.*

V. Preserving Autonomy by appeal to Sparse Animalist Identity

Identifying the human person and organism will preserve autonomy for then there will not be divergent persistence conditions and conflicting interests. However, there are considerable costs to advocating that the materialist identify persons with organisms that are essentially alive, i.e., our persistence conditions are those of animal and have nothing to do with our psychological capacities continuing. If persons are animals essentially and only contingently thinkers, we have to give up the intuitive responses to the thought experiments – transplants and inorganic replacements. We aren’t moved in the former and we don’t survive in the latter. Moreover, our mental lives become insignificant ontologically. We are not to be distinguished from those animals that can’t think nor those that possess minimal thought. We all would belong to the same fundamental kind, thinker.

Furthermore, if animals are identical to persons, then it would also mean that we must interpret the conjoined twins in the case of one animal with two heads to turn out to be just a single person cut off from himself. Consider the Hensel twins, or a more extreme version, where there appears to be a just a single organism but one with two cerebra whose contents are as inaccessible as my thoughts are to you. It is very difficult to consider those two thinkers as one person cut off from

herself. But if the organism is not a person in this situation, instead is a part, why think that the person was identical to the organism in the typical case, which likely includes yourself? If you are tempted to say there are two kinds of human persons, one identical to persons and the other a proper part of organisms, then you will have a problem handling the later grafting or growing of a head on your body. The presence of that head shouldn't shrink you to just a small part of your organism. But it is hard to believe that you remain identical to your organism but the additional head is just a proper part of it. This is especially difficult to sustain if the two cerebra are related to the rest of the body in parallel ways.²⁴

One doesn't want to identify animals and persons and claim that their persistence conditions are those of persons. This would mean that removing a cerebrum can destroy an animal even though it looks like life processes are uninterrupted. And it has an additional odd biological consequence that the transplant of the cerebrum into another skull will destroy the organism that is already there. We typically don't think receiving other organs can destroy the recipient. But if the mind is essential and determines our persistence conditions, then if an animal acquires a new brain and mind, it will go out of existence. If it doesn't, we will have two animals after the transplant in the same place, the one that was already there and the one that just arrived with the transplant.

There are shared costs of identifying persons and animals regardless of whether the animal/person has psychological or biological persistence conditions. We will be compelled to accept a sparse ontology devoid of brains or other candidate thinkers. *If the brain can think, then the problem of too many thinkers will return even if persons are identical to organisms. Moreover, to avoid Unger's Problem of the Mental Many, we must accept precise boundaries (epistemicism) that render us one of the many equally good candidate thinkers.* And we must not only have precise boundary, but deny the existence of an entity one atom bigger, or accept an explosion of person-like thinkers.

²⁴ The Hensel twins couldn't engage in certain movements unless they both wanted to.

Accounts of vagueness that understand vagueness as semantic indecision are unattractive because they seem to posit many thinkers. Precisifications assume the existence of many overlapping entities. As Lewis said about the Australian Outback, there are countless number of candidates, we just never bothered to say and shouldn't try to say which one is THE Outback. So precisifying which of the many candidates if the person will leave us with too many thinkers. As Unger noted, the Problem of the Mental Many isn't really a vagueness one.²⁵ Even if our boundary was precise and known, there are too many other candidate thinkers. As long as there are atoms in the brain that are redundant and not necessary for thinking, there could be a thinking entity one atom smaller than another thinker. So a sparse ontology is needed to deny that there are other entities that have such a redundant atom as a part.

It won't help to think there is just one entity there with worldly vague boundaries. If our parts admit of worldly or ontic vagueness, then we would have to tolerate vague existence and even vague identity. To see this, imagine replacing too much of a person either with parts that are too large or with many smaller parts in too quick a manner. Such too large or too speedy part replacement leads to the original person being replaced by a duplicate. A small part replacement, on the other hand, will just mean the same person briefly became smaller when a part was removed and then was restored to his original size with the assimilation of a new part. A part of in-between size will leave it vague whether the original person survives. However, there will be clearly a person and the end of the replacement of the middle-size part. But to avoid the original person being vaguely identical to the person after the replacement, there will have to be a sharp cut off point. So there will be a last atom of the person that can't be replaced on the pain of vague identity. And that last atom can be attached to the person to different degrees. So there will have to be a precise manner of

²⁵ Peter Unger. "The Mental Problems of the Many." In *Oxford Studies in Metaphysics* 1, ed. Dean Zimmerman (Oxford: Oxford University Press 2004), pg 196.

attachment to again avoid vague identity. If there is ontic vagueness whether that crucial atom is attached to the person, then it will be vague if that person persists.²⁶

If one is not convinced by the logical objections of Evans and Salmon against vague identity, there are considerations from the philosophy of mind. Try to make sense of being indeterminately identical to another being and sort of sharing their pain. We can't make any sense of that idea. So if worldly or ontic vagueness of parts or existence leads to an objectionable vague identity, then *modus tollens* means that there isn't any worldly part or existential vagueness. So there is no avoiding Unger's puzzle by positing one and only one vaguely bounded thinker.

We shall see later that *a soul theory can allow there are countless overlapping entities without multiplying thinkers*. Since none of them thinks, their existence is strange but hardly as objectionable as if they could all think. One problem with overlapping thinkers is that the loss of parts can mean many spatially coincident entities consisting of the same parts. To avoid this, it might be claimed that nothing can lose any parts (mereological essentialism).²⁷ But this makes the boundaries of the material person quite strange. One either becomes slightly scattered when exhaling and losing some particles or one goes out of existence if parts no longer cohere. But we thinkers surely are neither such scattered nor short-lived entities. If we are souls we have the right duration. Such a soul theory can even allow that entities don't gain or lose parts without suffering the consequence that we are short lived entities or exist scattered for most of our existence.²⁸

²⁶ See David Hershenov, "Vague Existence Implies Vague Identity." In *Vague Objects and Vague Identity*, eds. Ken Akiba and Ali Abasnezhad (Netherlands: Springer 2014), 283-303.

²⁷ Eric T. Olson, "The paradox of increase." *The Monist* (2006): 390-417.

²⁸ See Eric Olson (pg. 156) or Dean Zimmerman (pg. 495): Eric T. Olson, *What Are We?: A Study in Personal Ontology* (New York: Oxford University Press, 2007). Dean Zimmerman, "Material People." In *Oxford Handbook of Metaphysics*, eds. by Michael J. Loux and Dean Zimmerman (Oxford: Oxford University Press, 2003), pp. 491-526.

VI. Preserving Autonomy by appeal to Sparse Psychological Account of Identity

It might be thought that we don't have to accept a sparse ontology for we can just claim that we are persons spatially located within organisms and only persons genuinely think. McMahan and lately Parfit²⁹ have argued that we are proper parts of organisms consisting of those parts directly involved with the production of thought. Toes and hands aren't so involved, though they are caught up in life processes and thus part of an animal. Perhaps McMahan and Parfit's best strategy is to claim the organism doesn't really think, only a brain size part of the organism truly produces thought, just as the car is not really noisy, only its horn or engine is. However, we suspect that the cerebrum that allegedly produces thought doesn't meet the traditional independence criterion for being a substance. A cerebrum can't think on its own independently of the body or a substitute for the life processes like the philosophers' famous brain supporting vat. We maintain that the cerebrum doesn't think while in the vat, rather, if there is thought going on, the thinker is the composite of the cerebrum and the vat. Unable to think on its own, the (undetached) cerebrum needs a good part of the organism for thought to be produced. But it might not need all of it (hair, nails, toes etc.) However, whatever parts of the organism are irrelevant to thought production, their complement is probably large enough that the organism could survive a reduction in size to that earlier large proper part. It would then be thinking non-derivatively if the person is. And we don't see why it would cease to think non-derivatively if the amputated parts were restored. Moreover, it isn't clear that derivative thought isn't really thought. It is typically believed that the organism digests in virtue of its digestive system non-derivatively digesting but we don't then claim that the organism doesn't really engage in digestion. So why then deny that derivatively thinking is really thinking? So the too many thinkers isn't avoided by claiming the person is a spatial part of the organism.

²⁹ Derek Parfit, "We Are Not Human Beings." *Philosophy* 87, no. 1 (2012): 5-28.

It thus seems that the brain-size view of McMahan and Parfit will require that the world doesn't include any organisms that could be reduced in size to that of the person. That means a sparse ontology which just has thinkers as the only composite or a very gerrymandered sparse ontology that allows things besides thinkers but nothing that can be reduced to the size of a thinker. The latter is rather unprincipled so we will concentrate on the former special composition answer.

A sparse ontology of organisms better fits into the natural world than a sparse ontology of thinkers without organisms. Ontogenetically and phylogenetically it makes sense to posit the existence of mindless organisms that develop or evolve mental properties than to have no such entities and suddenly consciousness emerges and composites arise mid-pregnancy in the case of an individual thinker, or millennia ago in the case of thinking species. There is a nice evolutionary story of organisms arising from mindless to minded creatures. One can imagine minds providing certain survival and reproductive benefits. So the animal fits in nicely as a natural kind into evolutionary theory. The embodied mind sparsist doesn't fit as well. There are no animals that develop mental properties but thinkers just pop into existence when atoms are arranged brain-wise. Although souls are unnatural entities, they can fit better in some ways in the evolutionary picture than the sparse ontology. They enable the organism to reproduce and survive more effectively than mindless organisms. This might seem more evident with an emergent soul than a divinely created one but it is compatible with both. No laws about organisms need to be paraphrased away by the soul theorist, unlike the sparse ontologist

Substances have typically been construed as capable, in some sense, of existing independently. So instantiated properties (tropes) are not substances. Likewise, proper parts that can't survive removal from a larger entity can't be substances. Animals meet such an independence criterion but it isn't clear that cerebra do. A cerebrum won't be involved in thought production unless it has a body, brain stem or substitute. There is no reason to think a detached cerebrum can

think when it is removed and prior to be successfully transplanted, even if it will be involved in thought production post-transplant. So a cerebrum is not as good of a candidate as the animal or soul to be a thinker. Even an undamaged cerebrum won't think without the brainstem functions, according to experts on brain death.

Even if the cerebrum is a thinker rather than part of a thinker, and does pose a *Remnant Problem* for the animal, why doesn't the cerebrum-size thinker suffer its own version of a remnant problem when a large part of it is removed? The *Remnant Problem* for the animalist is that is the cerebrum can think when removed, then it should have been able to think earlier when embodied. That would bring a too many thinkers problem if the animal was a thinker. The opponents of animalism rightly ask why would attaching something external to the cerebrum prevent it from thinking? But the same problem exists for the embodied mind account of McMahan and Parfit. If a large part of the cerebrum can think post removal, then why couldn't it think earlier when embodied in the cerebrum? So the *Remnant Problem* generalizes to any theory that has composites that can become smaller. Only an immaterialist avoids it. Likewise, for the already discussed Problem of the Thinking Many. So we should judge the soul theory to have the advantage.

Cerebra don't maintain themselves the way animals and organisms do, which is one reason to think they are not substances and perhaps not natural kinds. The animal or organism takes in matter, metabolizes it, builds up itself, removes waste products, uses some parts of the body to defend and repair others etc. The cerebrum does little or none of this. It is utterly dependent upon the organism for its parts, energy, part removal, repairs etc. That suggests it lacks the independence constitutive of a substance. Perhaps McMahan and Parfit can appeal to functions within the cerebra for assimilating and removing particles that were brought to its boundary by the "atoms arranged animal-wise." But we suspect the lack of self-standing, autonomous, self-maintenance makes the

cerebrum too dependent to be a substance or natural kind and thus not as good a candidate for being the person as the animal.

A soul, on the other hand, is independent of the body in a way the cerebrum does not. Its existence doesn't depend upon the organism for its origins or its continuing existence. On some formulations the soul may emerge from the brain's functions, but not on the account we will defend later. Nor does it depend upon the organism for its sustenance, repairs, and replacements as does the cerebrum. It is the organism's life processes that enable the cerebrum to obtain parts, maintain them and replace them. Although it is true on the theory we will later defend that the soul's thought is neurologically dependent and would not think without the brain unless there was provided a substitute such as the philosophers storied brain in a vat. But this dependence is not in the form of a part to a whole nor does it existentially depend upon the brain. The soul is the person, on some construals. So a soul seems like a better candidate than the cerebrum for being a substance.

While embodied mind theorists have intuitive advantages over animalism in the transplant scenarios, they fare worse here than do the soul theories. This is evident in cases of a splitting of the cerebral hemispheres and a double transplant. The typical response is that there are two new persons. However, if only the left hemisphere was successfully transplanted, the right hemisphere destroyed in the process of dividing and removing the two hemispheres, the original person would be identical to the person with the left hemisphere. And if the right hemisphere had been split off from the left hemisphere and successfully transplanted while the left was destroyed, then the original person would be identical to the person with the right hemisphere. It is commonly held that a person can survive the loss of a hemisphere due to stroke or cancer etc.

However, Parfit and McMahan's appeal to brain size persons runs afoul of 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.³⁰ 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. If there was no person in the C body that was psychologically continuous with Adam, then Adam would be the person in the B body after the transplant. And if there had been no person psychologically continuous with Adam in the B body, then Adam would have been the person in the C body after the transplant. But if there is a person in the C body psychologically continuous with Adam, and a person with the B body that is psychologically continuous to Adam, then that not only prevents Adam from continuing to exist, but it also *noncausally* brings it about that a different person inhabits the B body than would have been the case if the C body person didn't exist. Likewise, it noncausally brings it about that a different person inhabits the C body than would have been the case if there was no one in the B body psychologically continuous with Adam.³¹

³⁰ Katherine Hawley, "Fission, Fusion and Intrinsic Facts", *Philosophy and Phenomenological Research* 71 (3) (2005): 602-621.

³¹ There is also a supervenience rationale for *the only x and y rule*. The physical relationship between Adam and the person in the B body is the same in the cases in which Adam is the person in the B body and when Adam is not identical to the person in the B body. One would think that identity differences in the two cases should track some physical differences between the relata but they don't in Parfit's interpretation of fission scenarios. See Harold Noonan, *Personal Identity* 2nd Edition, (London: Routledge Press, 2003).

Souls don't have any fission puzzles. The soul is simple and can't be split and its location and persistence doesn't depend upon what is happening elsewhere. It may be that the person's soul is "found" connected to the left hemisphere in a successful double transplant, but this has nothing to do with the presence or absence of the other hemisphere.

Having explained why unexplained existences are to be avoided, we can also prevent the materialist from trying to save autonomy by claiming that the overlapping thinkers have the same interests. Such reads may think Parfit has shown *that identity doesn't matter* and so the person and the animal will have the same interests. But Parfit only reaches this conclusion by providing a criterion for identity that involves 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. Parfit claims identity consists of relation R and a non-branching rule. Since the latter is trivial and extrinsic, he concludes that identity doesn't matter. But no theory of identity should have such a clause built into it. Thus Parfit can't obtain his conclusion that identity doesn't matter.³²

VI. Soul Theories May Begin to Look like the Sole Solution

Soul theories avoid 'The Problem of Too Many Thinkers' without adopting a sparse ontology for they maintain that there's just one soul attached to each group of overlapping entities. So the autonomy threatening moral conflicts do not arise. Brains need not be eliminated for they aren't sufficient for thought production. It is not just brains but their larger parts and whatever fully

³² We also find the conclusion counterintuitive in cases of brain damage that reduce one to infant like states devoid of psychological continuity to the pre-injury state. Our concern for ourselves is not reduced, or at least not dramatically, as Parfit's conclusion suggests. See the thought experiments of Unger and Williams in which people are told they will be tortured after their brains are wiped clean of their idiosyncratic psychologies.

material entities they all are embedded within (the heads, upper halves, and right hand complements of the organisms etc.) that aren't capable of thought. Soul theories also deliver the intuitive response in the thought experiments. The same soul remains related to the transplanted brain and the inorganic replacement matter. And soul theories don't have to consider the two-headed human animal to be just a single person cut off from himself. There is a different immaterial person related to each brain. So our mental capacities, our most valued and distinguishing traits, don't have to be construed as ontologically insignificant, i.e., having nothing to do with our persistence conditions

Zimmerman's emergent dualist solution is to have just one soul emerging from all the overlapping brain-candidates. The emergentist account is problematic because all the equally good overlapping physical candidates should be able to produce different souls. But Zimmerman argues that they overdetermine the same soul. This view strikes me as dubious. Moreover, the materialist could help herself to her own version of overdeterminism.³³ She might claim that all the candidate aggregates vaguely constitute the same person. The vagueness would be in the constitution relation, not the identity relation.³⁴

The emergent dualist view seems even more suspect when we consider non-overlapping parts of the brain that could each produce consciousness. If the split brain can give rise to thoughts cut off from each other, then transplanting those cerebral hemispheres into two disjoint bodies should do the same. Unlike the case of the split brain in one body, the two disjoint bodies would seem to be distinct agents. (The split brain in a single body rarely produces any frustration of agency that one would expect if there really were two persons connected to a single body.) But prior to the fissioning and transplant, the hemispheres can't produce emergent souls or there will be two people

³³ Will Bynoe and Nicholas Jones, "Solitude without Souls", *Philosophia* 41 (2012), 109-125.

³⁴ Lynne Baker, *The Metaphysics of Everyday Life: An Essay in Practical Reason* (Cambridge: Cambridge University Press, 2007).

where we want just one. So the presence of an adjacent cerebral hemisphere must somehow serve to nullify each hemisphere from giving rise to a soul. Yet this nullification must not prevent the soul that is overdetermined by the many overlapping brain candidates! It is very difficult to fathom this qualified nullification on top of the overdetermination that Zimmerman posits. It seems much more plausible for God to bestow the soul on the overlapping candidates and add any needed souls with the transplantation of the hemispheres.

Secular philosophers pushed towards dualism might opt for Unger's view of particular souls being disposed to be connected with overlapping bodies or brains. Unger is not bothered by one soul uniting with many overlapping material entities. He doesn't mind many countless overlapping bodies in your chair, just not many thinkers. He thinks there is just one soul propensitized to interact with the overlapping brain-like structures. This means there is haecceistic causation, individuals causally paired with other individuals but not in virtue of their being instances of a type. Contrast this with the conduct of particles like electrons or protons. They would attract or repel each member of the kind in question. A particular electron X doesn't interact differently with some protons than it does with others. Likewise for other electrons. It interacts in the same manner with every instance of that kind. Unger's theory of the propensitized soul doesn't do this but is causally connected in a haecceistic manner, interacting with just a particular body, not a duplicate of that body. This is very strange but it does enable Unger's account to avoid recourse to the overdetermination needed by Zimmerman. But Unger's account will suffer the same nullification problem that plagued Zimmerman's emergent dualism. It seems that parts of your brain could support thought if they appeared in different skulls of different bodies after fission and transplantation, so there would have to be souls disposed to interact with them. But why would matter beyond their boundaries block or extinguish their souls prior to fission or after a fusion scenario?

The Unger-inspired soul theorist has a time-indexed Ship of Theseus-like problem that the emergent dualist avoids. Assume that all the matter that first composed your body gets gradually replaced and is later reassembled elsewhere. Supposedly there will be a soul propensitized to the reassembled matter. But that means that the dispositions of the respective souls involve “memories” of a sort, or rather are time-indexed in a way hard to fathom. The soul of the reassembled matter would have waited its turn, not becoming involved with the same matter earlier. The soul of the reassembled matter isn’t your soul, which is still co-located with your replacement matter body.

The divinely created souls of Augustine and Descartes avoid the part replacement difficulties and the problems of overdetermination, fission, and fusion. God pairs the soul and the overlapping brain structures at the outset. If there is a pair of disjoint bodily structures, each with a cerebral hemisphere of a fissioned brain, God ensouls each disjoint mass so there is just one agent for each series of overlapping bodies. Likewise, God ensouls the body of the reassembled matter.

VII. Conclusion: A Theistic Ethics

If autonomy requires a soul and the best account of a soul are theistic ones, then a theistic ethics is looking like a very good candidate. We think the heavy lifting has now all been done. The ancillary hypotheses or premises to get us to a theistic ethics aren’t as difficult. They are not where the atheist and theist fight their battles.³⁵ If God made us or our soul, then it seems very likely He did so for a purpose. Like artifacts, we created beings would be fulfilling our proper end if we functioned in the way intended by our maker. Of course, we have not said anything specific what that a theistic ethic would look like and different religions disagree and theorists within the same denomination disagree with each other. But these are disputes between theists, not between atheists

³⁵ We don’t deny that that other reasons to doubt God’s existence like the Problem of Evil that might put weight on the scales thus providing reason to favor a non-soul theory despite its costs.

and theists. We are content here with arguing that autonomy needs a soul and a divinely created soul is the best candidate. That takes us very far in that it limits the rivals to theistic alternatives.