

Persons, Minds, and Bodies
By Benjamin Block

University of Michigan Honors Thesis, Winter 2011

Advisor: Assistant Professor David Manley
Second Reader: Assistant Professor Sarah Moss

Each of us who is capable of thought is equally capable of deducing that he or she exists. This simple idea, that we could not be the subject of an experience without also existing, goes back to Augustine and was further popularized by Descartes (Descartes, 1996) (Cochrane, 1944). Simply by referring to myself with the referential term “I,” I seem to imply that something exists that is identical to “me.” But knowing we exist is by no means equivalent with knowing what we are. We have developed several terms with which to refer to ourselves as a kind: human beings, humans, people, persons, mankind, etc. But it is far from obvious what it takes to count as a human being or a person, for example, and even whether there are non-human persons or human non-persons. In this paper, I aim to explore a view for what we essentially are; we generally agree that we are persons, but are we persons by necessity or accidentally? For example, could I continue to exist while no longer being a person, after a terrible accident? We have common-sense intuitions about what we are, and many of these may turn out to be correct, though none goes unchallenged. Views that deny intuitions in an attempt to save others have arisen from animalists, followers of Hume, *et al.* We generally consider ourselves to be these things that think and have bodies and move around and occupy a certain amount of space. Some of these properties we have essentially, and others we could continue to exist without — for instance, I could occupy more or less space than I do now. This raises questions regarding what kinds of changes could I undergo. But the question I wish to explore is: What kind of object am I most fundamentally?

The goal of this paper will be to examine some views regarding which of the aforementioned qualities, among others, are essential to my existence, and some objections to those views. In so doing, I hope to offer a new view that avoids some of

those objections to the existing views — specifically objections concerning the relationships between persons, animals, bodies, and minds.

We can begin by asking about the precise relationship between ourselves (on the one hand), and the entities we call ‘persons’, ‘bodies’ and ‘animals’ on the other. Here it is useful to consider three claims about those things, to which I will refer often throughout this paper: Suppose I am the only person in my room. Then, in my room right now...

- (1) There exists something that is essentially a person.
- (2) There exists something that is essentially a human organism but not essentially a person.
- (3) There exists something that is essentially a body but not essentially an organism.

Which of these claims are true? Though some combinations of these three claims or their negations result in rather outlandish views, no straightforward contradiction arises from any combination of these three claims or their negations; moreover, the anti-essentialist could even deny all three outright. That said, each appears to be true when we consider the way we commonly think about each of the three objects. We can begin by considering some straightforward intuitions in favor of each claim; we will reconsider each claim at length in subsequent sections.

For example, consider (3). Most of us think that, assuming an organism dies in such a way that its body does not disintegrate, there would be something shaped like that organism that continues to exist after its death. This we call its body. But assuming an organism goes out of existence when it dies, the body cannot be essentially an organism

It also seems plausible that, were a person’s head cut off (or a person’s brain removed, save for the brain stem) and the rest of the body kept functioning by machines,

a human organism would survive the ordeal even though the person would not. Again, it seems that these objects cannot be identical, or we would have a single object that both did and did not survive the same incident.

The view that seems evident from this, and the view I am going to defend in this paper, is that these three distinct objects all exist in the same physical space. I will refer to objects that coincide spatially in this manner as “colocated.”

In defending the acceptance of (1), (2), and (3), I will first examine arguments which deny the plausibility of colocated objects altogether, regardless of whether any of those objects would be generally considered “alive.” I will then address objections specifically to collocation of living things.

Having shown that a theory of persons, organisms, and bodies which includes collocation can withstand all objections raised on the grounds of collocation, I will argue for a specific view of what persons, organisms, bodies, and minds truly are and how they are related to one another and the self, ultimately supporting the simple premise, “Persons and only persons are objects of first-person reference” (Noonan, 2010).

Part I: Colocated Material Objects

By far the most common case discussed when it comes to colocated objects is the statue and the lump. The general situation is as follows: Suppose we have a lump of some material, perhaps clay, and we fashion it into a statue depicting a specific human form. If you were to show that statue to a layperson and inquire as to how many objects that person was being shown, they would likely respond that there is only a single object — the statue. However, it seems strange to claim that the lump of clay has ceased to exist

and been replaced by the statue, so where is the lump of clay? Judith Jarvis Thomson lays out the simplest example of this quite neatly. Suppose you fashioned the lump of clay into the statue at 1 p.m. in the afternoon, and that was your first time ever sculpting. At 9 a.m. that morning, the clay existed, but the statue did not (Thomson, 1998). The clay has the property of having existed at 9 a.m., but the statue does not, so it seems intuitively they cannot be numerically the same object.

Equally puzzling is the question of what happens if the statue is squished: A lump of clay can survive squishing but a statue cannot. Particularly because of this intuition regarding persistence conditions and survival, it appears that there are at least two colocated objects: the state and the clay lump. This is the view that I support, but in order to do so I must show it can be defended against some important objections and alternative views¹ that have been raised. While my discussion of the first two objections in this section will be discussed in terms of statues and their materials, the last two will be discussed more generally in terms of material objects, though I will endeavor to show their applicability to statue-lump cases as well.

Alternative 1: Contingent Identity

There are some worthwhile objections to the view that the statue and the clay lump are distinct, colocated objects. Some of these objections arise as solutions to the problem that multiple objects occupying the exact same space seems impossible, or at least greatly counterintuitive. It would seem to be a problem for colocation if there is a better way to save the intuitions about bodies, animals, and persons that we sketched

¹ While views regarding temporal parts of objects can come into play in discussions of colocation for both animate and inanimate objects, I will be setting aside discussions of such parts from this paper.

above, without having to violate the intuition that there is only one object in the room. Based on this problem, some alternative views have been developed to escape the perceived problems with collocation. The first view I will discuss in this paper is the notion of “contingent identity” as detailed by Allan Gibbard.

Identity is thought by many essentialists to hold *necessarily* wherever there is a genuine identity relation. If A is identical to B, for all properties possessed by A, those properties must be possessed by B, and vice versa. This is simply Leibniz’s Law. As Michael Jubien put it (Jubien, 1993):

Undoubtedly there are some relations that behave in some respects like identity but which do not hold of necessity. I believe it is a fundamental error to think that one of these relations actually is the identity relation, that is, to think that identity actually doesn’t hold of necessity. I also believe it is seriously misleading to agree that none of these relations is the identity relation, but nevertheless to dub one of them ‘contingent identity.’

To claim, then, that A and B could be identical even though there are some properties — such as persistence conditions or having existed at some past time T — which they do not share appears untenable.

The proponent of contingent identity has one response compatible with Leibniz’s Law, and that response is well laid out by Gibbard. It is to say that specific material objects do not have such modal dispositional properties as having existed at some time in the past or having the potential to be destroyed in some way in the future (Gibbard, 1975). Gibbard refers specifically to the fact that the statue could be rolled up into a ball before the clay has dried, and shows that in this instance there is a lump of clay after the ordeal, but no statue. Yet he still maintains that the two are contingently identical, with contingent identity defined as such: $A = B \ \& \ \Diamond(A \text{ exists} \ \& \ B \text{ exists} \ \& \ A \neq B)$.

How does this view apply to the case of human beings? Usually, one's body seems to survive one's death, and in such a case Gibbard would not claim that one is contingently identical to one's body. But Gibbard's view raises the important question whether we can use merely modal facts — especially persistence conditions — to distinguish completely colocated objects. For example, someone might hold that she and the organism with which she is colocated both came into existence at conception and will as a matter of fact die at the same time. In that case, does it follow that she is identical to an organism? Or can she distinguish herself from her organism by considering *merely possible* cases in which, for example, most of her brain is removed and transplanted into another body? That is, can objects be distinguished by a mere difference in persistence conditions even if they *in fact* overlap completely in spacetime?

Given how intuitive and simple the assigning of persistence conditions generally seems, the burden is on Gibbard to explain why exactly we cannot ascribe specific modal or dispositional properties (such as a statue's disposition to be destroyed if rolled up into a ball) to specific concrete objects. Lynne Rudder Baker (Baker, 1997) outlines Gibbard's attack on this assignment of properties, which is also an attack on essentialism itself, as consisting of the following three premises, which she takes verbatim from different parts of his essay:

- (a) "Modal expressions do not apply to concrete things independently of the way that they are designated."
- (b) "A property, if it is to be a property, must apply or not apply to a thing independently of the way that it is designated."
- ∴ (c) "Expressions constructed with modal operators...simply do not give properties of concrete things"

The key to Gibbard's argument is that whenever we refer to an object with a modal property, we are only able to do so by using a referral term that implies the object is of a certain kind. Mark Johnston describes what is meant by this well (Johnston, 1992):

"Goliath" not only picks out the statue but it also, perhaps with the cooperation of context, indicates a particular counterpart relation or set of necessary and sufficient qualitative conditions for tracing Goliath across the worlds—if you like, the "statue" counterpart relation

It is this argument that causes trouble for theories of colocation that appeal to mere differences in persistence conditions to distinguish objects. Without being able to assign modal expressions, especially persistence conditions, to specific concrete objects, we are left only with what is evident from the material composition of an object.

Karen Bennett calls this problem the "grounding problem," based on the fact that it is unclear what could ground any of the alleged differences in the applicability of modal expressions (Bennett, 2004). Bennett also claims that there are differences in properties she claims not to be modal, which the grounding problem must also address. Her examples include an object's being innovative, valuable, or well-made.

In the case of the statue and the clay, or any similar pair of colocated objects, the objects are made of the exact same particles of matter. This collection of matter is obviously identical to itself. So, at all times during which an object is not being referred to by a certain name, the only properties the object has derive from being composed of those particles of matter. Accordingly, any objects constituted by those same particles must have all and only the exact same properties, and would thus be identical. Any differences in properties would only arise through someone's referring to the object a certain way, but the act of referring to an object does not actually alter the object.

Accordingly, while we may consider an object to have different properties when discussed as a lump than it does when discussed as a statue, the actual concrete object does not innately possess any of those inconsistent properties. These are the claims that Gibbard puts forward to argue for contingent identity, but Baker is right in pointing out that the first of his three premises is contentious and open to counterexamples. Baker presents such a counterexample as follows (Baker, 1997):

“... modal expressions include not only predicates like ‘is essentially a statue’, but also many other kinds of predicates. Suppose that a surgeon removes a bullet from a wounded soldier’s shoulder, and later presents the bullet to the injured soldier and declares: ‘This thing could have killed you.’ Then it seems true of that particular bullet, independently of the way that it is designated, that it could have killed the soldier. In general, predicates ascribing abilities and powers to concrete things, independently of the way that they are designated, entail that modal expressions apply to concrete things.”

Her argument here appears to be correct, though it would be strengthened by not referring to the object as a “bullet” each time; her claims are more obviously true *qua* bullet, as Gibbard would say, than they are in general pertaining to the indicated object. We can thus say, concerning that very object, that *it* would have killed the soldier. To illustrate why Baker is right, we must examine Gibbard’s tactic in coming to the premise Baker is denying. He uses the concept of possible worlds, stating that we cannot know if a given object exists in another possible world independent of the way to which it is referred. Suppose we have two designated terms to refer to the same concrete object — for Gibbard they are “*Goliath*” the statue and “Lump1” the lump of clay. If we ask whether *Goliath* exists in another world, the question is meaningless unless it is understood that *Goliath* refers to the object *as a statue*, or *qua statue*. “Meaningful cross-

world identities of such things as statues, it begins to seem, must be identities *qua* something: *qua* statue or *qua* lump, *qua* Goliath or *qua* Lump1. It makes sense to talk of the ‘same statue’ in different possible worlds, but no sense to talk of the ‘same thing’ ” (Gibbard, 1975).

He summarizes his reasoning behind this as follows: A specific designator will always refer to an object of some specific kind, and for each kind there is a specific set of persistence conditions (and presumably other modal expressions as well). “In rare cases, at least,” he writes, “one thing will be of two different kinds, with different persistence criteria, and whereas one proper name refers to it as a thing of one kind, another proper name will refer to it as a thing of another kind” (Gibbard, 1975).

The consequences of accepting this view, which must be accepted for claim (a) to also be acceptable, are counterintuitive. An object of two kinds is an object that has two sets of persistence conditions, but it only has each set of persistence conditions relative to the method to which it is referred. What are the persistence conditions, then, of an object never referred to as being of a certain kind, even if it is *in fact* of two kinds? Gibbard is not simply saying we *don't know* or *never express* the persistence conditions — he says “modal expressions do not apply” to that object. It seems to me, then, that an object never referred to or perhaps even thought of as having any kind would have no persistence conditions at all. An object with no persistence conditions, I believe, is simply inconceivable. The only salvageable, intuitive conclusion to draw from the statement “that has no persistence conditions” would be that the object will endure forever, but surely Gibbard does not think that, without referral, every concrete object would be indestructible and eternal.

The one alternative I can come up with, that an object not referred to *qua* something has the persistence conditions of *all* the kinds of which it is a member, leads to an even more egregious problem. Given this view, an object could both survive and not survive the same event.

This is why Gibbard says it makes no sense to talk of the same “thing” in different possible worlds. Without referring to the object *qua* something, it seeming has either no persistence conditions or inconsistent persistence conditions. But I believe that the better view is one that allows us to speak of the same thing across different possible worlds. Given a view that claims each object can be of only exactly one kind, we are able to speak of objects across different possible worlds without worrying about the method to which they are referred, and we avoid cases like those just described where extremely counterintuitive results are reached regarding persistence conditions.

Gibbard’s argument, as it poses a threat to colocation, is reliant on the existence of objects which are of two or more kinds. But as we have just examined, a view in which an object has multiple possible sets of persistence conditions, all of which only apply based on the method of referral to that object, leads to very counterintuitive consequences. In order to pose a threat to colocation, Gibbard’s theory would need to show that there is no available reason *not* to claim identity between the statue and the clay. But because his theory offers what seems to be a counterintuitive explanation for the persistent conditions of contingently identical objects, it does not manage to show identity holds between those objects with any certainty. Given that doubt over Gibbard’s central premise, we have no reason to choose a contingent identity view as superior to a colocation view.

Alternative 2: Dominant Sortals

There is another form of objection to colocation which is raised, among others, by Michael Burke. His view is motivated by the common-sense objection to colocation that it leads to a counterintuitive economy of objects.

Burke's arguments have the same goal as the tweaked version of Gibbard's view: To show that there is no available reason *not* to claim identity in cases where A and B occupy the same physical space. The reason we discovered not to claim identity in the contingent identity objection was the irreconcilability of multiple sets of persistence conditions, all of which apply conditionally based on methods of referral. Burke avoids this problem by claiming that, while there do exist objects of two kinds, only the persistence conditions of one of those kinds, or "sortals," applies. He calls the sortal whose persistence conditions apply the "dominant" sortal (Burke, 1994). Dominance, he argues, is held by whichever sortal encompasses the broadest scope of properties:

An object's dominant sortal is the sortal that tells the object's sort, the sortal that tells what the object is. Of course, this answer raises a new question, in response to which I offer (as a first approximation) the following criterion: Of the sortals satisfied by an object, the one that tells the object's sort is the one whose satisfaction entails possession of the widest range of properties.

The statue, he argues, encompasses a wider range of properties than a lump of clay. They do not need to be specific properties — merely a range of properties. To clarify, a statue is not necessarily made of any one particular substance, but for every statue there will be at least one substance of which it is made. Therefore, Burke argues, it has an equal range of substantial properties as a lump of clay. But in addition to that, it has properties like form which exceed the range of properties held by the lump.

But simply because there is only one dominant sortal does not mean that an object's other sortals do not still apply. Burke argues that the statue is still a member of the "clay lump" sortal — it simply has its persistence conditions defined by another sortal (Burke, 1994). The obvious question, for which I don't think Burke has an adequate defense, is "Why do the persistence conditions of the other sortals suddenly not matter?" As described before when tweaking Gibbard's view, Burke genuinely believes that the original clay lump ceases to exist when turned into a statue, as it loses its old persistence conditions and acquires a new set with its new, dominant sortal. The fact that the original lump ceases to exist and is wholly replaced by the statue-lump does seem strange, but arguing against that specific consequence of this view is both complicated and unnecessary for refuting Burke's overall argument.

Burke's argument is based on the idea that a sortal — not just a *dominant* sortal, but any sortal — is defined as a term that answers the question "What is it?" for some potentially existing object. This is an idea he borrows from Wiggins. However, Burke's view on what serves as a genuine answer to this question seems unjustifiably restrictive (Burke, 1994):

Consider the sortal 'piece of copper'. For some objects, *mere* pieces of copper, 'piece of copper' answers the what is it question. But what would stop us from saying ... that there are other pieces of copper, Statue, for example, for which 'piece of copper' fails to answer that question?

"Failing to answer" a question, as I read the phrase, could either mean giving no response at all (which is not the case here) or giving an answer that is wrong. It does *not* mean giving an answer that is true despite not being what some might call the "best" possible answer. Indicating a copper statue and calling it a piece of copper *does* answer what it is,

and does so *correctly*. Burke himself thinks the copper statue meets the criteria for being a ‘piece of copper’ and claims that ‘piece of copper’ is a sortal which the statue possesses — he simply does not think that is the *dominant* sortal. Despite the above quotation, both Burke and I believe that the copper statue must surely *also* be a piece of copper, even if it is not a piece of copper *essentially*. Burke uses the fact that “piece of copper” supposedly does not answer “What is it?” to skip over the examination of what happens to the persistence conditions of non-dominant sortals. In other words, he ignores the fact that the object still *is* a piece of copper despite not being a piece of copper *essentially*. I will now attempt to show the problem that arises from an object having non-dominant sortals.

We have a copper statue. Given this, there are three claims we could make all of which Burke acknowledges are true: “We have a thing that meets the criteria for copper statue,” “We have a thing that meets the criteria for statue,” and “We have a thing that meets the criteria for piece of copper.” It is uncertain which of the first two sortals mentioned would be chosen by Burke as dominant, but it clearly would not be the sortal “piece of copper.” Suppose, now, that we squish our copper statue. Burke claims that now we have an object for which the dominant sortal is “piece of copper.” This means it cannot be the same object we had before, in any sense. The original thing that met the criteria for “piece of copper” has been destroyed. But now consider the following argument:

1. An object survives so long as no event that would destroy it occurs.
2. The event of being squished does not destroy pieces of copper.
3. C meets all the criteria for being a piece of copper.
4. C is squished, and no other event occurs which impacts S in any way.

The third premise could be rewritten as “C is a piece of copper.” Burke may not say that

is the “best” answer for what C is, but insofar as the statement is an acceptable rephrase of the third premise (which I believe it is), the statement holds. Based on the four given premises, we reach the conclusion:

5. C survives.

Burke’s view leads to a scenario where an object can be destroyed by an event that should not destroy it. Burke’s likely counter-objection would be to say that I am wrong in equating my third premise with the phrase “C is a piece of copper.” But to say that an object can meet all the criteria of being a certain kind of object yet still *not* be that kind of object (even non-essentially) is extremely counterintuitive. On what grounds can an object be excluded from a sortal if it meets every single criterion? Given this, it seems that Burke’s view leaves much room for doubt regarding identity, again for reasons pertaining to persistence conditions. He must give some explanation of how the thing that is a piece of copper can be identical to the thing that is a statue which does not suffer from the flaw explicated above.

And his closest attempt fails to do so. He says that the piece of copper is destroyed, but the *mere copper* of which that piece was constituted persists. That mere copper is not itself an object, he says, but a collection of copper atoms with no single sum. First, I do not hold this view to be tenable for reasons that will be explained in Objection 4. But even if we grant Burke this claim, it does not save his argument. E. J. Lowe writes in response to Burke, “If it is admitted that many objects can collectively occupy the same place as one other object, in what way is this supposed to be preferable to admitting that one object can occupy the same place as one other object?” (Lowe, 1995). Burke is either forced into inconsistent claims regarding survival or fails to achieve his original goal of

avoiding colocation. Accordingly, his “dominant sortals” view does not sufficiently pose a threat to views including colocation.

There are also cases in which an object’s dominant sortal is unclear². Suppose I have a flashlight with a built-in radio. Flashlights, radios, and flashlight-radios all seem to cover basically the same range of properties. They all have properties of substance, purpose, emitting a certain kind of thing, etc. There does not seem to be any clear way to determine what the dominant sortal is. Suppose that I hack off the part of the object with the light bulb in it, but the radio still works fine. Have I destroyed the object? That would depend on its dominant sortal. If it was radio, the object survived; but if it was flashlight or flashlight-radio, it didn’t. Yet we had seemingly no way to know which it was, so we do not know whether the object survived or what its persistence conditions were.

Objection 3: Additive Properties

A third concern hinges on the problems that colocated objects present in terms of duplication of physical properties — not simply invoking our intuition of “one object in any one space,” but pointing out the consequences of colocation as relevant to properties such as mass.

It is this issue that leads Dean Zimmerman to reject colocated objects altogether (Zimmerman, 1995). He lays out the problem in terms of bodies and cellular tissue. The intuition that these objects are distinct arises from the fact that a body endures over time while specific cellular tissue structures dies and shift and are replaced. Supposing the cellular tissue at a given time has a mass of 140 pounds, the body also seemingly has a

² Thanks to David Manley for developing the idea behind this argument.

mass of 140 pounds. We can apply this to what Thomasson calls a Principle of Additivity of Weights: “where A and B are not identical, if A weighs n kilos and B weighs m kilos, A and B together weigh $(n + m)$ kilos” (Thomasson, 2006). The body and the cellular tissue, to those who accept colocated objects, are not identical. The body weighs 140 pounds and the tissue weighs 140 pounds. Accordingly, it seems that, when we put the tissue and the body on a scale together, it should read 280 pounds, but obviously it does not. This same dilemma would arise regarding a 140-pound marble statue and the 140-pound “chunk of marble” of which it is composed. Colocation theorists must offer some solution to this problem of the additivity of mass.

One option we could propose is to say that a constituted object has its properties in virtue of the thing(s) it is constituted by. But Zimmerman objects to such a view, saying it does not give enough credit to those objects which are constituted by others, seemingly setting them apart from the unconstituted physical objects (Zimmerman, 1995):

Of course we are free to say that if my body is coincident with a mass that weighs 140 pounds, then it too can be said to weigh 140 pounds in virtue of its close relationship to something that really has this weight. But why should my body — a physical object in its own right, made of physical parts each with its own proper weight — have to “borrow” its weight from some other object?

Given this problem with any view that involves borrowing, there are two alternative solutions to the additivity problem which I think show promise.

Thomasson’s solution is to claim that any two terms whose objects are “analytically interrelated” neither compete nor duplicate any properties that they share. She describes this in the following terms: “(when a and b both have property P) there is no rivalry for possession of a single property instance, nor any ‘doubling up’ of properties where a ’s being P is an analytic entailment of b ’s being P ” (Thomasson, 2006).

What is referred to here by “analytic entailment” or “analytically interrelated” is the phenomenon wherein, by the use of human reason, we can understand one term or claim to necessitate the existence or truth of another term or claim. Thomasson uses the example of a group of simple particles “arranged baseballwise” being causally responsible for breaking a window. One could also simply say, if one accepts the existence of baseballs, that a baseball was causally responsible for breaking the window. Because our reason allows us to understand that “baseball” is an entailment of particles arranged “baseballwise,” there is no rivalry for causal responsibility here. Nor for mass, velocity, or any other such shared property.

Sydney Shoemaker offers a similar form of solution wherein properties are shared but not necessarily borrowed. Shoemaker defines two kinds of properties: thin properties are those that are shared by colocated objects, while thick properties can be ascribed to some but not all of the present colocated objects (Shoemaker 2008). It is important to note that a property’s being thin or thick can vary based on context, though “microstructural properties” like shape, mass, electrical charge, etc, tend to be thin properties. Through this conception of properties, a 140 pound mass is shared wholly by both a body and a batch of cellular tissue, even though they are not identical. Persistence conditions, which are a thick property, are not shared.

I am inclined to think that either of these solutions are acceptable. While there is no logical proof of their truth, they seem to comply with our intuitions and evidence from the real world. Providing that the other sections of this paper regarding collocation give us reason to believe colocated objects exist, there must be some explanation for why a scale on which I stand displays the weight of my body — not double, triple, or some other

multiple of that weight. Thomasson in particular offers an attractive solution in that, by her conception, each object independently has a certain weight, without any form of sharing necessitated. Such properties are simply not doubled up when one object's possessing that property is entailed by the other's.

Objection 4: Arbitrariness of Composition

Peter van Inwagen attacks colocation on its most fundamental level: composition, on the grounds that determining its occurrences is too arbitrary (van Inwagen, 1987).

One who accepts colocation likely also accepts that all material objects greater in size than a single, simple particle are colocated with some aggregate of simple particles (which I will henceforth simply call "simples," as does van Inwagen. With this term, I refer to the smallest, indivisible particles, but I do not wish to name any specifically, should science later prove them divisible.) In general, we think that these material objects — a statue, for existence — exist in virtue of being *composed* of some aggregate of simples. In the case of the clay lump and the statue — as well as the case of the body, organism, and person — we are concerned with whether some single aggregate of simples can come to compose two distinct objects, one of which constitutes the other. Without composition there would be no constitution, it seems. There would be no objects except simples, and because constitution is a one-to-one relation, no simple can constitute anything except itself.

Peter van Inwagen's view is similar to nihilism (the view that there is no composition whatsoever), but he makes an exception for living things. He bases this exception on the fact that we know ourselves to exist on Augustinian/Cartesian grounds,

and surely we must be made of *something*. This thing that we are is essentially caught up in is a “life.” It is also, in all real-world cases, an organism. But van Inwagen would not agree that there is any colocation of objects save for the “life” and its aggregate of simples. There is no body distinct from the organism distinct from the person (van Inwagen, 1990).

But why is this the case? Why are we unable to say that some aggregate of simples composes a body which then constitutes an organism which then constitutes a person? There are two questions van Inwagen puts forth regarding composition, intending to show that each has no good answer and thus show how rarely composition occurs. We will concern ourselves primarily with the first, which van Inwagen refers to as the Special Composition Question.

This question is posed as follows: “What would one have to do—what *could* one do—to get the *xs* to compose something?” I will here examine van Inwagen’s discussion of this question and his claim that there is no suitable answer available for it. In attempting to find a solution to van Inwagen’s problems with this question, I will also be concerning myself about when some *xs* come to compose two distinct *ys*, as that is what occurs in the cases of person-organism colocation.

Peter van Inwagen proposes several of the popular possible explanations for what could be sufficient condition for composition to occur. They are contact, fastening, cohesion, and fusion. In discussing contact, van Inwagen addresses the idea of arbitrariness in the following way (van Inwagen, 1987):

One might suspect that there is no answer to these questions laid up in heaven, and that how we answer them—assuming they're worth answering—is going to be simply a matter of convention. But I think that we can see that there are at least

some cases in which mere contact is not sufficient for the production of a new object.

The way van Inwagen writes here, it seems clear that he is looking for some principle that holds universally. Were it acceptable that contact, or any of the four criteria discussed, be sufficient for composition *sometimes* but not always, there would be no reason to dwell on the cases in which they are insufficient. But van Inwagen believes that showing any single counterexample negates the idea of contact entirely, and the same for fastening, cohesion, and fusion. Basically, van Inwagen is looking for an answer that removes the arbitrariness that he knows some people will claim is inevitable in the above quote.

And for each of the four criteria, van Inwagen does provide a counterexample such that creating the proper relationship between two objects does not make them into a distinct, new object. One such example for fastening is the relationship between a nut and a bolt (van Inwagen, 1987). When they are fastened together, we believe we have a nut fastened to a bolt, not some new object called nutbolt. In *Material Beings*, van Inwagen writes, “it does not follow from the mere existence of certain objects that there is an object that has them all as parts; not every set of objects has a sum” (van Inwagen, 1990).

There are two principle methods of objecting to van Inwagen’s claim. One is to deny the above quote and claim universal mereology, that for every possible combination of objects, there genuinely exists an object that is their sum. This leads to the existence of some rather odd objects. On the more believable end of the spectrum we have things like nutbolt, but then there is also the object that is both a frog and a rocket ship, or both a horse and a stool. I suppose these objects would be called frog-rocket and horse-stool.

But this view alone would be insufficient to show collocation of persons and organisms. While universal mereology as laid out so far enables us to say that the simples

arranged organismwise compose an organism, we have not yet established any reason to claim they can compose an organism *and* something else. We must find a method for which the xs can compose more than one y.

Here, the universal mereologist in favor of colocation could argue that not only is every possible combination of objects the necessary-and-sufficient set of objects for some sortal, but that every possible combination of *properties* is necessary-and-sufficient for some sortal as well. So when we are faced with a bunch of simples arranged organismwise, we can consider an object composed of those simples and with the necessary properties of an organism, or an object composed of those simples with the necessary properties of a person.

On this view, there are now certainly bodies, organisms, and persons all collocated with one another and composed by the same simples, but each having different properties considered in that combination. But there is also an object that is all three of them, and an object that is all three of them plus the planet Jupiter.

Ned Markosian describes this universal mereology view as the main opposition to nihilism and to van Inwagen, but it is only the case if we are truly unwilling to accept any level of arbitrariness in our conception of composition (Markosian, 1998). I believe that not only ought we accept arbitrariness, but that even van Inwagen has not escaped arbitrariness in his view, despite that seemingly being his goal.

When van Inwagen says the only composed objects are “lives,” it is not sufficiently clear what actually counts as a life. Viruses, for example, have been a subject of much controversy regarding whether or not they are genuinely living things (Villarreal, 2004). But what about a case of *dicephalus*, a pair of conjoined twins with

two heads and a principally shared remainder of the body? (I will discuss such cases further in the section on bodies later in this paper.) Here we have what seemingly must be described as two persons. They have different thoughts, different desires, different minds, etc. But there is only one organism to be composed. With this example, I do not intend necessarily to disprove van Inwagen, but to show that even with his view, we may have to allow colocation. Either van Inwagen will insist that there is only one life, in which case there is only one individual, or he will have to allow that there are two lives composed of the same simples. Because of this, we have two colocated individuals.

But when we look at death, we seem to be forced into accepting that even van Inwagen's view is arbitrary. At what point is a human life no longer present. If an organism suffers complete brain death, the remainder could still survive and the rest of the body could continue to function. This seems to still be a life. But what if, one by one, we remove vital organs and limbs and such, replacing them with machines, until all that's left is a functioning, beating heart kept going by some machine. Is *this* a life? If not, at what point did the life cease to exist? Even if van Inwagen wishes to argue that only a conscious mental life counts as a life (for only things with that kind of life could know they exist), there is a lack of clarity in coma patients and similar cases of greatly reduced brain function.

If we accept this arbitrariness, then we are able to formulate a more common-sense view of which objects exist in the world. We can claim that nutbolt is not a real, distinct object, but that a block tower is distinct from the blocks that compose it. Essentially, this view boils down to the realization that the objects we treat as existing tend to be the ones for which we have developed words. If NUTBOLT had been a

concept since the dawn of language, I likely would have far different things to say regarding it. Though in this alternate version of our world, perhaps it would not be the word “nutbolt” that were greatly altered, but the word “exists.” It is wholly conceivable that the language could have evolved such that “exists” or some other word is agreed to apply to all the things for which universal mereologists claim it applies. But for many, this might necessitate the idea of existence being altered from its current general meaning. Composition is arbitrary because it is a part of how we generally perceive the world around us — but our perceptions of the world are undeniably arbitrary, and there is no problem with that.

In short, there are at least two ways to avoid van Inwagen’s objections: one is to adopt universal mereology, and the other is to adopt a position that explains and justifies the arbitrariness about what exists that makes van Inwagen uncomfortable. I do not intend in this paper to endorse either of these positions. I merely hope to show that either is an acceptable alternative to van Inwagen’s view that would allow us to accept a colocation view regarding persons.

Part II: Colocated Thinking Objects

In examining objections to colocation in terms of general physical objects, we have found no good reason to reject the intuitive ideas that a clay lump is distinct from and survives the squishing of a clay statue. It appears a sensible belief that two objects can be coincident but not identical, in large part due to their persistence conditions. But there is another form of objection to the idea of colocation which applies uniquely to people and other thinking, living things. This objection is based on the “too many

subjects” problem, sometimes also called the “too many thinkers” problem. Rather than attacking collocation for being counterintuitive in terms of the quantity of physical objects present, it attacks collocation for presenting a world in which actions that ought to have only one subject would in fact have two or more.

The view whose proponents most commonly advance this objection would also endorse that, of the original three claims, only (2) is true. This view is often called animalism, and tends to use the term “human animal” as opposed to my original choice, “human organism.” I think, however, that the meaning animalists intend to convey with the term “animal” is better related through the term “organism,” as many may consider animals (say, Fido) to have certain mental traits by necessity which animalists do not intend in using the term. (That is, they might be as tempted to say that Fido has passed away if his fore-brain is dead, as they would in the case of a human.) Also, Eric Olson, one of the principle modern proponents of animalism, often uses the word “organism” in his writings on animalism as an alternative for “animal.”

What is the “Too Many Subjects” Problem?

The too many subjects problem is as follows: Suppose that original claims (1) and (2) were both true. Most who accept that (1) is true, including myself, believe that (1) is the object with which we are identical. We also believe that we are thinking creatures, and accordingly believe that the objects which are essentially persons are also thinking. But we also believe that we are collocated with and constituted by our organism. Though we, the undeniably thinking thing, are not identical with our organisms, it still seems from a common-sense, scientific perspective that the organism itself is thinking. After all,

it has a brain that is functioning properly functioning, and as Olson writes, “What stops a typical human animal from using its brain to think? Isn’t that what that organ is for?” (Olson, 2003). David B Hershenov seconds this, pointing out the complete sharing of atoms between the person and organism: Since the person can obviously think, the organism should also have such a capacity as a result of possessing the same brain as well as every other atom of the person” (Hershenov, 2005).

The consequence here is that each thought a person thinks has two subjects — the object-essentially-a-person and the object-essentially-an-organism. This is not a contradiction in of itself, but Olson claims — and I am inclined to agree — that “this is not an attractive picture.” I will here take a moment to explain why the too many subjects problem is a problem that must be avoided, not simply another perhaps-counterintuitive philosophical consequence.

Olson first states that have having both the person and organism as thinkers simply appears to be a problem of “overcrowding” (Olson, 2003). There is something about an action as unique to intelligent creatures (and uniquely first-person) as thinking is that makes it difficult to accept the notion of a thought action having multiple subjects, even if those subjects are colocated with one another.

The situation seems especially strange when examining the genesis of an idea. Upon hearing the idea, someone may ask “Who thought of this?” The answer would be that both the thinking animal and the thinking person thought of it. But we do not want to say that each thought of it “independently” of one another. They are constituted of the same matter, including the same neurons. It would have been a shared idea, then, with no lone conceiver. I do not think it will be possible for me to find a case wherein

overcrowding alone results in a scenario that is logically unacceptable, but I do believe that this idea case is more difficult to accept than the notion of both an animal and a person being the subject of a pure physical action, such as standing on a stage. To put it another way, returning to Shoemaker's concepts of thin and thick properties, I do not think that the property of having generated a thought or idea can be a thin property.

Olson finds one of the greatest problems of having these multiple thinkers to be that each has a seemingly equal claim to being the proper object of the term "self" (Olson, 2003). In a case where each simultaneously thinks, "I am the object to which Benjamin is identical," or, "I would not survive an accident in which my frontal cortex is destroyed" there is no method available to explain how one could be correct but not the other. (But clearly, there is only one thing which I am essentially and only the animal could survive frontal-cortex destruction.) If both are thinking in the same space at the same time with the same thought, it seems incredibly strange for one to be right and the other wrong — but perhaps it is not impossible³. What *is* impossible is that there be two things to which I am identical that are not identical to each other. If the organism and person are not identical, then one of the thinkers must be wrong, but, as just stated, there

³ There is another version of the two many subjects problem for which this comes into play. Suppose we hold a universal mereology view that includes an object's ability to hold any set of properties by necessity and thus allows collocation. There could be multiple objects, all of which are persons, collocated with one another. In this case, I would argue that the only one identical to "I" or any other proper noun is that with the sortal "person," and that no contradiction arises from differing thought values with the other person-objects because they are not identical to the primary thinker, and are instead constituted by her. But this version of the problem is uncommon. Most people believe that such properties generally considered phase sortals (Oderberg, 1996) can never be essential to an object at all. In this case, there would be no object whose sortal is "redhead person" to be wrong when thinking "I am not essentially a redhead."

is no good method available to determine which one, as both the thinking organism and the person otherwise seem equally good candidates for the self.

It is for this reason that Olson does not accept both (1) and (2). In the section that follows, I will examine the argument he uses to show that only (2) is true, and attempt to reject it in such a way that also avoids the too many subjects problem.

Rejecting the “Thinking Animal” Argument

In order to introduce the too many subjects problem in his essay, Eric Olson uses the “thinking animal” argument, which he presents exactly as follows (Olson, 2003):

O1. $(\exists x)(x \text{ is a human animal} \ \& \ x \text{ is sitting in your chair})$

O2. $(x)((x \text{ is a human animal} \ \& \ x \text{ is sitting in your chair}) \rightarrow x \text{ is thinking})$

O3. $(x)((x \text{ is thinking} \ \& \ x \text{ is sitting in your chair}) \rightarrow x = \text{you})$

O4. $(\exists x)(x \text{ is a human animal} \ \& \ x = \text{you})$

It seems safe to assume that, for all mentions of human animals, Olson refers to an object that is *essentially* a human animal, not accidentally. If this were not so, then all four premises would hold even if I am essentially a person but only accidentally a human animal – but that is not the conclusion Olson wants to reach.

Let us see what claims the thinking animal argument makes, premise by premise, relative to the view which accepts original claims (1), (2), and (3). The colocation theory I support suffers no blow from the acceptance of O1. My acceptance of (2) clearly necessitates my acceptance of that premise. But Olson’s conclusion, O4, is inconsistent with my view — at least once we establish that the object-essentially-a-person is that with which I am identical. The premises that ultimately lead to the negation of claim (1) are O2 and O3. Were either of these premises shown to be false, or at least sufficiently

dubious, we would not be required to accept O4. Between O2 and O3, Olson seems to consider the former to be necessary given the presence of a brain and the latter to follow necessarily from that given (Olson, 2003). Because O3 is reliant on O2, our target is premise O2. If the organism is not thinking at all, then there is no object competing with the person for thought-subject, only one thinking object in my chair, and no reason to equate my self with my organism.

In the rest of this section, I will endeavor to show that the rejection of O2 is not contradictory to the way we think about the world, and explain how its rejection leads to a consistent view that accepts (1), (2), and (3).

Olson claims that the only scenario in which I am thinking but the human animal in my chair is not thinking is one wherein it is impossible for animals to think: “You think, but the animal doesn’t. The reason for this can only be that the animal can’t think. If it were able to think, it would be thinking now” (Olson, 2003). While this claim does not seem to follow from any logical necessity, I believe the idea that organisms cannot think to be true. To some, this statement may seem a ludicrous denial of the way nature works. But much of this disagreement centers around what I believe to be importantly different possible meanings of what it is for an object to be thinking, only one of which ought to properly be called “thinking.”

Imagine that you are using a computer to conduct research. The action we will consider here as an analogy to thinking is “researching.” There are two possible candidates for “researcher” here: you and the computer. Both you and the computer could be said to be searching for relevant information, but the sense in which each object is performing this action is greatly different from the other. In the act of researching, the

computer is a tool. It is being used by the person to complete the action in question. In common conversation, it seems unintuitive to say “the computer researched the average yearly rainfall in Brazil,” and much more accurate to say that *you* researched it. I believe this intuition is correct. The computer, while a tool used in the act of researching, is not itself the researcher; *you* are.

The organism, then, is not thinking. You, the person, are thinking, and the organism, especially its brain, is the tool which you are using to perform that action. Thought production is still a biological process consisting in the firing of neurons every which way in the brain, but the action of truly “thinking” is performed by the person, not the organism.

But even given the animalist view regarding thinking, there are problems animalism faces in terms of the too many subjects problem⁴. If the organism were identical to the mere body, then an organism would not die until the body were destroyed, but this does not appear to be the case. Bodies, many generally agree, continue to exist after an organism’s death. Because a body would have the same brain as the organism, it would too be thinking, so animalism must deny that there is such a body, and thus it seems that nothing survives our deaths. Also, animalism must deny that organisms have brains as proper parts, because the brain and the organism would be both be thinking, and there would again be too many subjects.

But given my view, wherein the brain itself is not thinking but merely serving as a tool for thought, there are several alternatives available for how thinking could manifest itself in the person. I tend to favor the belief that thought is epiphenomenal qualia, as

⁴ Thanks to David Manley for suggesting this *tu quoque*.

detailed by Jackson. This view states that thought is simply a side-effect of neuronal activity with no causal powers (Jackson, 1982). Thought could also be a mental property of persons. It is not essential that we detail here the precise nature of what thought is — we need simply establish the possibility that the subject of thinking be distinct from the physical, biological organism. And if we accept property dualism and/or the legitimacy of epiphenomenal qualia, we can reasonably attribute the properties of consciousness and thought with the person rather than the organism.

In this section, I hope to have given solutions to the varieties of too many subjects problem that will allow for the acceptance of (1), (2), and (3) while maintaining an acceptable view of what thought *is* and what it means to be thinking. Having addressed this concern, I believe I have shown that a view which claims persons are colocated with organisms and bodies can withstand the popular objections and is worth exploring. In the following sections, I will detail the view that I believe to be strongest regarding persons, minds, and bodies, as well as explore some of the consequences of this view.

PART III: What is a Mind?

What I refer to with the word “myself” is a “person,” but it is also a thing that can think. From the discussion in the previous section, it seems quite likely that my self is *not* identical with my brain or the organism in which it is included, because these things cannot think — though they are tools generally necessary for the production of thought. If there were some thinking thing that were *constituting* me as opposed to *being* me, that would result in the too many subjects problem; both the thinking person I am and the thinking object constituting me would be equal candidates for “myself,” and there would

also be an overcrowding problem (Olson, 2003). If there is a thing to be called a “mind,” then, it must either be an unthinking mind or be identical to the person I am rather than constituting it. I am reluctant to choose the latter, because I believe that persons can be constituted by bodies while minds cannot.

Because of this problem regarding the potential identity between myself and a mind, I will put forward the view that a mind is actually an unthinking entity. Instead, I would call the mind a “consciousness engine” — an entity that is fully responsible for the production of thought, but is not itself thinking.

The simple parallel to draw here would be “a mind is a brain,” but it is not quite that simple. Were a mind to be simply identical to the material brain, it would persist after death so long as the brain itself were not destroyed. For there to be a mind, there must be some activity that could be called mental. Campbell and McMahan put forth a view that solves this, wherein they argue that a functioning brain is not constituted by a “mere brain,” but then comes to be replaced entirely by a mere brain when function ceases (Campbell & McMahan, 2010). They do this, however, to avoid the too many subjects problem by creating only one thinking candidate. This means they also believe there is no thinking thing constituted by the functioning brain, and that the self is identical to the functioning brain. But I disagree with this view on two points: First, it considers the self to be identical with something that cannot be constituted by a body; second, it claims that a functioning brain pops out of existence and is replaced by a mere brain at death, rather than saying the two are collocated.

A mind, by my conception, is materially constituted by a brain but has the essential property of functioning in a specific way — yet it is *not* itself thinking. The two

questions I will explore in the remainder of this section are “In what specific way must a brain function in order to constitute a mind?” and “What exactly *is* a brain?”

Functions Necessary to the Mind

In the discussion of the thinking animal argument, I sketched a very rough criterion of the functioning brain, wherein functioning was defined as “operational in a way that allows for thinking,” and made distinct from the mere regulation of bodily processes. While this criterion is on the right track, it is insufficient in its level of detail.

The functioning which I argue is required of the brain in order that it constitute a mind is of a level and kind that allows for psychological continuity. But we cannot simply leave it at that. Among philosophers who advocate psychological continuity theories, there is much debate over the specific criteria required for that continuity to exist in a manner sufficient to continued personal identity. Some, including Shoemaker, feel that memory is an important criterion in personal identity (Shoemaker, 1959). Parfit’s view is rooted more in psychological connectedness and choosing the best psychological candidate for being a future version of oneself (Parfit, *Personal Identity*, 1971). Scott Campbell argues for a Humean view that continuity, and even persons themselves, exist as a series of bundled mental states (Campbell S. , 2006).

The psychological continuity view which I will advocate is similar to that put forward by Unger (Unger, 1990). Unger considers his view to be rooted in physical continuity, but Shoemaker is correct in pointing out that Unger’s view merely advocates a unique and less discussed form of psychological continuity (Shoemaker, 1992);

Unger advocates for what Shoemaker refers to as “core psychology” as that which must be continuous to establish personal identity over time. Shoemaker defines this core psychology as simply “basic psychological capacities,” and places it in opposition to “individual psychology,” which is composed of memory, dispositions, and general conscious thought which we are able to perceive in ourselves (Shoemaker, 1992). Core psychology seems to be deeply rooted in the functioning of a brain, and its continuity, therefore, is that brain’s persisting ability to function. So Unger’s view of essential continuity being physical requires that the psychological which must be essentially continuous for a person to exist over time cannot be transported between brains.

Shoemaker attempts to show the Unger’s view is inferior to those based on individual psychology by envisioning a world wherein brains transmit their states to one another across a grid of electrical wiring, each one shutting down until it receives the psychological state of a neighboring brain. The core psychology of each brain stays put in that brain, but the individual psychologies jump around from brain to brain. Shoemaker claims that, intuitively, the persons are those who jump around (Shoemaker, 1992):

If we allow that there are persons in Brainland at all, as we clearly should, we have here a case in which continuity with respect to individual psychology clearly trumps continuity with respect to core psychology. While the brains in Brainland stay put, the persons move around.

The intuitive appeal of Shoemaker’s claim is obvious. For much of their individual existences, the physical brains of Brainland have no individual psychology at all. And at each instance when the brain *does* have individual psychology, the brain has memories of experiences that do not involve itself. Neither of these, however, is

conclusive proof that individual psychology trumps core psychology in terms of being essential for a person's continuity through time.

Suppose an individual brain is hooked up to a machine with astounding capabilities like those machines in Brainland. Every minute, the oldest remaining thought or memory in that brain is sent out of the brain via the machine, destroyed, and replaced with an artificially created thought or memory. This new memory does not necessarily involve the brain in which it is now contained. Interacting with the person possessing this hooked-up brain may seem strange: the person would certainly appear forgetful, and might have sudden changes in taste or disposition. But for the most part, it would likely feel as though we were interacting with a single person. Because the brain presumably has quite a few thoughts and memories, each one is present in the brain for some time. But if the brain were hooked up to the machine for long enough, it would have no continuity of individual psychology relative to its pre-machine self. In this case, I believe that core psychology trumps individual psychology.

Shoemaker even concedes that there are cases in which other psychological continuity theorists may find themselves relying on the idea of core psychology rather than individual psychology (Shoemaker, 1992):

“Nor is there any reason why an advocate of a standard psychological continuity view could not hold that this sort of continuity is sometimes sufficient for personal identity—i.e., that there can be cases (tragic ones) in which there is continuity with respect to core psychology and none with respect to individual psychology (because individual psychology has been obliterated), and that in such a case the person could still survive.”

Given that Shoemaker makes this concession, I am inclined to find flaw with Brainland, and I believe there is such an interpretation that would allow for core

psychology to remain essential even in that case. Given the way that the brains in Brainland interact, it does not seem unreasonable to claim that all of Brainland is a single mind. Interpreted as one object, Brainland has continuous core psychology with neuronal signals traveling about the mind in a fashion that, while obvious not identical with how neurons generally function in the human brain, is sufficiently similar that the comparison holds. My reasons for making this one-mind claim regarding Brainland will be made more clear in the following section on brains, so I will not linger too long on Brainland at this moment.

For all the reasons described above, I believe that core psychological continuity is the best candidate for a continuity theory that prevents intuition-defying gaps in a person's existence and holds across as many strange or tragic cases as possible, but there are two specific scenarios in which I want to ensure that this view holds — sleep states, and coma states — as these are very real scenarios which can expose holes in consciousness-based theories of continuity. While the core psychological continuity theory is not strictly “consciousness-based,” it is rooted in the *potential* for consciousness, and thus will face similar, though importantly different obstacles in the form of these two scenarios.

In their essay “Brain Death and Personal Identity” Michael Green and Daniel Wikler put forward an argument for brain death equating person death (Green & Wikler, 1980). In so doing, they define brain death in a way quite similar to what I would call “cessation of core psychological continuity.” Approaching the subject from a more medical perspective, they offer a clean and simple explanation of how a view based in core mental capacities deals with cases of sleep and coma:

“Of course, our view does not imply that a person dies with his last moment of consciousness. What matters is the preservation of the substrate, not the psychological states which it produces. Hence a person who suffers brain death during sleep dies at the time of brain death, not the time of onset of sleep. Similarly, a person in a persistent coma might be alive if enough of the brain remained structurally and functionally intact.”

The argument seems rather clear with regards to sleep. As I stated with core psychological continuity, it is the *capacity* that must endure (and that capacity is preserved in the substrate), not necessarily any actual execution of that capacity. Even those who enter a deep, unconscious sleep continue to exist during that sleep because the substrate and its capacities remain intact and functioning throughout, where functioning is defined more based on potential.

Imagine the sleeping brain to be like a television that is currently turned off. It is not currently displaying a program, but it has the capacity to do so as long as its internal workings remain as they should be, and it is said to be a “functioning” television in the sense that were you to hit the power button, it would begin to display programming. In this analogy, I am treating waking up or being awoken with having the power button pressed.

The case with comas is more difficult to decide simply, and I think the vagueness in Green and Winkler’s statements regarding it is unavoidable. This is not the fault of the philosophical view, however, but simply a consequence of modern science being still incomplete. In many cases, it is impossible to know whether a coma is genuinely irreversible, or even whether the individual in the coma is fully unconscious. In cases of a genuinely irreversible, unconscious coma, even if some substrate of the brain is intact, this is insufficient to core psychological continuity. Implicit in the term *capacity* as I use

it with core psychological continuity is the fact that it is *metaphysically* possible at some point in the future for that capacity to be acted upon. If the substrate is physically intact but no set of circumstances could come about such that it would once again generate individual consciousness, the person is no longer alive. Mere incapability of present science to have the substrate continue in those capabilities does *not* itself mean the person is no longer alive. It is simply the most accurate indicator available that a person might be no longer alive. Unger offers a physical example of the sense in which I mean the term capacity in defending his view that a brain can undergo disabling events while still retaining its capacities through time (Unger, 1992):

When disassembled, my special car does have the capacity to ("start up" and) go 70 mph. When disassembled, it's just that she's "in no shape for" that capacity to be exercised. Put her back together, and she'll exercise the capacity. ... When my brain is quartered, then I'm "in no shape for" my capacity to think to be exercised. But, realized in those quarters, there still is that capacity of mine. Put me back together, and I'll exercise it.

Returning to the television analogy, even if the inner workings of the turned-off television are intact, we would not say the television "works" if there is no way to ever turn it on and never could be. But for patients in comas who are either potentially conscious or may regain consciousness, we cannot say with certainty whether there is some physically possible solution that would allow their brains to generate individual consciousness again. This leads to the unavoidable vagueness of determining whether a person still exists in modern coma cases.

I hope now to have offered a consistent theory of what functions a brain must have in order to constitute a mind. It must have core psychological continuity, defined as the continued capacity for individual psychology, which is essentially interchangeable

with “consciousness.” But in an endeavor to be able to identify which things constitute minds and which things do not, it is insufficient simply to know how a brain must function in order to constitute one. We must also be sure that the object which we are examining is, in actuality, a brain. Because of this, I will now endeavor to explain what exactly I intend to refer to in this paper with the term “a brain.”

What Brains Are

The reason that I established the necessary functionality of a mind before exploring the criteria for an object’s being a brain is that the physical entity that is a brain is itself partially defined by this functionality. One criterion of a brain, I stipulate, is that it contains parts sufficient to produce core psychology. The distinction between these criteria for a brain and the earlier criteria for a mind is that the brain need not be actively functioning in any sense whatsoever. A completely dead brain is still a brain so long as it has parts that, under different circumstances, would produce a core psychology.

Beyond this requirement, however, I stipulate no other restrictions regarding capabilities. A brain lacking any part that would give it the ability to comprehend language is still a brain. The same goes for all other individual functions except the most general possible form of the potential for consciousness production. Because of this, it follows that there is no required size or shape that a brain must possess.

Another vital criterion for an object’s being a brain is that it must have interacting parts. Again, with the use of “interacting parts” I intend the meaning, “parts that could interact if the brain were alive, regardless of whether the brain is actually alive at the present moment.” I do not require that each part of the brain be able to interact with each

other part directly, but that each part, through some chain of neural impulses, can interact with each other part or neural impulses sent from each other part. (Whenever I use “neural” or “neuronal,” this is intended to refer to an object that is a functional equivalent of a neuron.) These interactions must be of a nature relevant to the unique functioning of that kind of object. Were two such objects to collide physically yet yield no mental consequence by doing so, this interaction would be insufficient.

This kind of view is in some contrast to the view held by Wigan, who argues that this level of interaction is insufficient to unify the brain organ, much like how we do not consider the eyes to be “one eyes” simply because the information from both indirectly meets up at some point (Wigan, 1844, 1985). I believe, however, that the nature of the brain’s communication with itself, including between hemispheres, is sufficiently unified to render the conventional brain a single organ.

For example, imagine there are exactly five physically attached brain parts labeled A through E. If A and B can interact, C and D can interact, and D and E can interact, with all interactions of the kind sufficient for brain-unity, then there are two brains present. A and B are closed off from the rest, so they are their own brain. C, D, and E also compose a brain, because, even though C does not interact directly with E, the signals from both are capable of interacting in part D. This statement that they are brains is assuming, of course, that AB can generate core psychological continuity and CDE can also independently generate core psychological continuity. The fact that they are all physically attached is inconsequential.

Given this version of what I will call the “interaction criterion,” there appears to be a good deal of vagueness regarding where the brain spatially ends, especially when we

consider a brain in an average human organism. The brain interacts with parts of the body that we traditionally do not consider to be part of the brain, and which I do not want included in the brain by my conception either. The entire nervous system, for example, is not all just part of a brain. The same is true of the heart, the lungs, and every other organ which interacts with what we traditionally call the brain in a human organism. When pricked, an arm sends pain signals to the brain, and when a brain sends a different kind of signal to the arm, it may lift, bend, or perform some other action. Clearly there is interaction between brain parts and arm parts, so a criterion must be developed to prevent body parts like arms from being considered part of the brain.

What makes the brain unique from the rest of the nervous system or from other vital organs seems to be its ability to *process* information, rather than “mindlessly” sending outputs or receiving neural inputs. So it would be insufficient to say that each part of the brain is capable of both sending and receiving neural signals. As I just stated, this would make an arm a suitable candidate for being part of the brain. What is unique about the idea of a brain “processing” information, is that the neural output given off by the brain is based on a neural input rather than another form. When an arm sends a pain signal to the brain, the input that generated this output was not neural. More likely, it was pointy. And when the arm receives a neural input, such as a command to bend, this is not translated into a neural output but into a physical, bodily action. (This action may result in a separate neural output being sent back to the brain giving the sensation of the arm bending, but the input that directly led to this output was a biological, physical input — not a neural one.) The third criterion for the brain, then, is as follows: “Each part of the brain must be capable of generating a neural output *based on* a neural input.” Not being

myself a neuroscientist, I do not wish to go into exact specifics on what is meant by “based on” in this criterion, but my simple attempt will be to say, “the neural output was the result of a purely neural process in which the neural input was the initiator.” (This purely neural process need not take place entirely within any single part of the brain.)

I speculate that there may be an objection to this view on the grounds that I have not yet specified what is meant by “part.” A neural fiber, for example, seems to be part of a brain yet is merely a pathway. It would be a stretch of language to say that it generates output, and perhaps even to say that it receives input. In a similar vein, it does not seem that any single proton in an atom in the brain is capable of receiving input or generating output. Therefore, I must specify what I mean by “part.” And to do so, I will actually rely on the “processing criterion” I have just discussed. A candidate for a “part of the brain” is any chunk of matter that is capable of receiving neural inputs and emitting neural outputs, along with any *purely neuronal* pathways through which this process occurs. By this criterion, again based on function over form, neural fibers are included with the part of the brain that utilizes them, and something so small as an atom does not alone have sufficient capabilities to be considered a “brain part.”

But even this improved version of the “interaction criterion” leads to some difficulties, however, when looking at a brain whose parts interact in a manner that could be called “incomplete,” as happens in many patients who have the corpus callosum, the greatest concentration of nerve fibers connecting the traditional brain’s two hemispheres, severed. (This procedure is undertaken most commonly as a treatment for epilepsy.) I believe that, if this conception of a brain is to be viable, it must be able to accommodate situations such as these and yield an acceptable, if not immediately obvious result.

In “Brain Bisection and the Unity of Consciousness,” Thomas Nagel details both the findings of research conducted on patients with this kind of poor inter-hemispheric communication and the consequences it has regarding the number of present minds in these and all individuals (Nagel, 1971). I do not wish to take the time here to fully recount all of the findings from the various studies, though they are fascinating in themselves, but I will attempt to summarize those elements which are relevant to the current endeavor.

In the vast majority of normal situations, these patients functioned in a matter that indistinguishable behaviorally from those who have not undergone such an operation⁵. If shown an image for any significant length of time, they could afterwards state what object had been displayed to them. Their abilities to speak, read, and write were, in most situations, unimpaired.

But in situations specifically engineered to send a visual, aural, or olfactory stimulus to only one hemisphere of the brain, either through flashing images in a certain way of blocking off an ear or nostril, the patients were unable to report on what sensation had been presented in the same way as an individual with an intact corpus callosum. Here is a telling example, from Nagel’s writing (Nagel, 1971):

“A smell fed to the right nostril (which stimulates the right hemisphere) will elicit a verbal denial that the subject smells anything, but if asked to point with the left hand at a corresponding object he will succeed in picking out e.g. a clove of garlic, protesting all the while that he smells absolutely nothing, so how can he possibly point to what he smells. If the smell is an unpleasant one like that of rotten eggs, these denials will be accompanied by wrinklins of the nose and mouth, and guttural exclamations of disgust.”

⁵ This is, in part, because the corpus callosum is not the only bundle of nerve fibers connecting the two hemispheres, and some direct interaction is still possible after its severance.

Cases like these bring up troubling questions about how many brains, minds, and even persons are present in individuals lacking the corpus callosum or suffering from poor brain communication for some other reason. Nagel writes that we seemingly have five alternatives from which to choose regarding the number of minds in such a patient, and the number of minds is precisely what I am not attempting to determine. Ultimately, however, he comes to the conclusion that none of them are acceptable. I will first determine which seems the most plausible explanation given my conception of the brain, then see if there is some way to overcome the obstacles with Nagel believes are sufficient to rule that option out. The five choices are as follows, taken verbatim from Nagel:

- (I) The patients have one fairly normal mind associated with the left hemisphere, and the responses emanating from the nonverbal right hemisphere are the responses of an automaton, and are not produced by conscious mental processes.
- (II) The patients have only one mind, associated with the left hemisphere, but there also occur (associated with the right hemisphere_ isolated conscious mental phenomena, not integrated into a mind at all, though they can perhaps be ascribed to the organism
- (III) The patients have two minds, one which can talk and one which can't.
- (IV) They have one mind, whose contents derive from both hemispheres and are rather peculiar and dissociated.
- (V) They have one normal mind most of the time, while the hemispheres are functioning in parallel, but two minds are elicited by the experimental situations which yield the interesting results. (Perhaps the single mind splits in two and then reconvenes after the experiment is over.)

Nagel does not waste much time in refuting (I) and (II), and I will follow suit.

There is clearly consciousness in both hemispheres, and both function, making both hemispheres candidates for brains and minds by my criteria. There is no reason to make exception for the right hemisphere in these cases (Nagel, 1971).

Option (III) is unappealing because there is clearly interaction between parts across hemispheres in the vast majority of the patients' real-world experiences. As has been the case with many of my conceptions given thus far, a brain and a mind are defined by their *capacities*. These hemispheres do have the *capacity* to interact — even though circumstances can be manipulated such that they will not do so in all cases. This point also argues against option (V), because even when the hemispheres do not in actuality interact, it is clear from other experiences the patients have that the capacity is intact.

This leaves option (IV), and it is the option to which I am most attracted⁶. Nagel, however, is inclined to reject it. He gives what I interpret as two main objections to the acceptance of (IV). The first is that the mind appears to be capable of multitasking in a way that a single mind ought not be able to. He writes, “in these patients there appear to be things happening simultaneously which cannot fit into a single mind: simultaneous attention to two incompatible tasks, for example, without interaction between the purposes of the left and right hands” (Nagel, 1971). His second issue taken is that a single mind implies a single consciousness, but if the hemispheres are truly not interacting yet still each functioning like a conscious entity, then there must be two loci of consciousness-generation, which he claims to be indicative of two minds.

The first of these two objections does not actually pose a threat to the acceptance of (IV), I believe. His use of the word “incompatible” seems to presuppose assumptions about compatibility all of which are based on the average human mind.

⁶ I must be clear that this is only with regard to specific versions of this case. In most of the cases described, a patient is able to achieve some minimal direct communication between hemispheres through minor nerve fibers. This allows for indistinguishable functioning most tasks from those whose brain hemispheres are wholly connected. Were the interaction to be only indirect — that is, the interaction was by each hemisphere's realizing what the other is doing and following along rather than a physical interaction of parts — then I would claim there are two minds, and that this would be a fission case. I discuss such cases in depth in a later section of this paper.

Simply because a mind is capable of things which the average human mind is incapable of does not mean it cannot be a single mind. Suppose that humans were to evolve in such a way that brains and minds capable of handling these kinds of now-“incompatible” tasks became the norm. We would not say that these individuals necessarily have two minds if they were also capable of having full interaction and unification of consciousness.

Susan Leigh Anderson offers a more specific explanation on what she speculates is meant by incompatible, which she has since replaced with the word “contradictory”:
“Why are these pairs of actions considered contradictory? I think it is because they are different actions each of which are truly ascribed to the individual(s) in question and are not performed by different spatial parts of the ‘individual(s)’ ” (Anderson, 1976). This does not seem to make much sense, either. Two actions cannot be contradictory solely because they are undertaken by seemingly separate individuals. It must also be true that they *cannot* be performed by different spatial parts of an individual. In the cases discussed here, it could at most be said that the actions *are not* being performed by different spatial parts, but that is not enough to claim contradictoriness.

Before moving on from this entirely, there is a stronger version of this same vein of objection which must also be overcome. Roland Puccetti, who advocates that all humans with the average human brain have two minds all the time, raises the same objection as Nagel, but frames it as a contradiction (Puccetti, 1973). He claims that in these experimental situation, the patient must be said to both see and not see the visual stimulus. Therefore, he claims, to say there is one mind is to break Leibniz’s Law. But this version of the objection also seems to me to be flawed. He claims that those arguing

for Nagel's option (IV) would, and perhaps must, accept *both* that the patient has one mind *and* that the single mind both knows and doesn't know what object it saw.

This does not seem a necessity at all. I argue that if any part of the single mind knows, then the mind simply knows, and the other part of the mind's denial does not overrule that fact. Suppose that you are asked to recall an obscure fact you learned some time ago. You may attempt but fail to do so, and thus respond "I don't know that fact." But then suppose the next morning you wake up and suddenly remember, because your brain has been attempting to do so, unbeknownst to your conscious self. It seems that your mind knew the fact the whole time — just not the part of the mind relevant at the time you were asked. It seems from this example that it is not necessary for an *entire* mind to be aware of Fact X in order for the claim "that mind knows Fact X" to be true.

Joseph Margolis points out that Puccetti himself acknowledges the possibility of cases that seem exceedingly similar to those described in these experiments yet still only include one person (Margolis, 1975):

Puccetti himself acknowledges the possibility of "a schizoid personality with divergent impulses; or of dissociative reactions in the same person." We may admit, with Puccetti, that the commissurotomy patient does not exhibit these syndromes. Still, either they show how it is possible for there to be two minds in one person or else they suggest that Puccetti has not helped us to see how to determine whether there is one mind or whether there are two minds, or more than two minds, in one person or experimental subject.

While Puccetti can and does argue, as Margolis says he can, that there are two minds to a person, Margolis is correct that Puccetti's arguments could just as easily lead to the conclusion I am driving at: option (IV) in Nagel's scheme.

Now I am ready to move on to Nagel's other objection to (IV) — the presence of two loci of consciousness. Specifically, he states that the acceptance of a single mind in these patients must be willing to claim a single consciousness is being produced by two independent, yet sufficiently complete control systems (Nagel, 1971). Nagel does not argue this point much further, except to make the objection already addressed above, and I see no reason not to accept precisely what Nagel says I must. The experiments Nagel discusses seem to give evidence that this is in fact how the average human mind works. Clearly there is more than one part of the brain sufficient for consciousness, or else one of the hemispheres would be left unconscious in the trials. It seems strange to think that, at the moment the corpus callosum is severed, one hemisphere suddenly gains a capacity for consciousness that it lacked before. So if this is how the mind works all the time, with two or more different potential control centers creating a unified consciousness, there is little reason to find it objectionable when the same thing happens in these patients. Simply because consciousness manifests itself in distinct ways does not mean we have numerically distinct consciousnesses from those we had before.

The one difference between the two cases, on which the entire objection now stands, is the word "independent." In the average human mind, the control centers are capable of directly interacting, thus making a single consciousness more plausible. I offer several responses to those who believe that the "independence" makes a crucial difference. First, my conception of interacting parts and its reliance on the *capacity* to interact calls into question whether it is even correct to call them independent in this case. But the real problem brought up by independent is the fragmentation of the consciousness. And to address this, I merely return to my argument against Puccetti: It is

not unreasonable that one part of a single mind be unaware of what is happening in another part of that same mind. Ignorance of conscious experiences is included in this argument. Acceptance of the view I put forth here regarding consciousness and the non-essentiality of strict unity allows for multiple bundles of individual psychology to float around in the Brainland mind, as well.

We now, finally, have what I consider to be a viable set of criteria for what can be properly called “a brain,” which, when considered together with our discussion of the mind, allows us to identify where minds are present. And accordingly, because a mind is necessary and sufficient for a person, we can identify where persons are present as well.

Part IV: What is a Body?

But a mind is not the only object which can compose a person. The majority of persons also have bodies (though brains in vats are people, too). These bodies are non-essential to a person’s existence, but when they exist in proper relation to a mind, they come to constitute a person. In discussing the body, it can be appropriate to consider the brain, or the mind it constitutes, to be a mereological part of the body. But we must note that, while the mind is the most essential part of a person, it is *not* the most essential part of the body. As we see in brain transplant cases, the body remains when the mind and person move to another body. Still, we must remember that this does not mean that both minds and brains can and should be considered a part of the body as well.

No body can constitute a person unless it includes a mind. The body is to be conceived, then, as constituted by all the matter that has a sufficient relation to the mind, either directly or indirectly. (It follows from this that one mind can never have two

bodies.) The nature of this “sufficient relation” can vary, and there are different possible relationships they could possess each of which is sufficient to some matter’s being included in the body. These viable relationships include control, support, regulation, or sensation. Also included in the body are any parts having a sufficient relationship with another part, which in turn has an aforementioned relationship to the brain.

With control, I refer to the ability we have to move parts of the body at will. Thus our limbs, our lungs, and several other parts of the traditional human body come to be included. By support, I refer to all those organs by whose functioning the mind is kept functioning. By regulation, I refer to the mind’s role in keeping several organs working in tandem and maintaining a bodily balance. This is essentially the same as support, but with the roles of the brain and the body reversed. And lastly, by sensation I refer to those parts of the body that give us sensations as conscious states by sending signals to the mind. This set of criteria is an attempt to lay out the functions of the nervous system in such a way that, were a species to evolve in such a way that these functions were accomplished by a mechanism biologically different from nervous systems in currently existing organisms, this conception would still be applicable.

There is some debate over whether a person’s mind must *accept* a body part — in addition to the neural criteria I have given — for that body part to be reasonably considered part of that person. Feinberg and Keenan describe several cases wherein patients suffered disorders that caused them to treat certain parts of their body as not being their own (Feinberg & Keenan, 2005). Of the many disorders they discuss, the most relevant to our discussion of the body here is asomatognosia. Asomatognosia is “a condition in which the patient demonstrates delusional misidentification of a part of the

body. Asomatognosia typically occurs in a patient with a right hemisphere lesion, left hemiplegia, severe sensory loss on the left side, and left hemispatial neglect” (Feinberg & Keenan, 2005). In clinical conversations between Feinberg and patients, the patients would refer to a part of their body as being its own individual, or having gone on vacation without asking.

Because of the biological states in these patients, it seems the body part in question is *almost* distinct from the body. There is no control relationship between the brain and that arm. But Feinberg and Keenan describe the disorder as occurring in patients with “severe” sensory loss, not necessarily complete sensory loss. In a case where there is still a sensation relationship between the arm and the mind, I would argue that the arm is in fact part of the body. The control relationship to the body part is not a necessary one, and the sensation relationship, like each of the four kinds of relationships detailed earlier, is sufficient on its own. I do not wish to define any part of a person in a way that a person’s perception of biological states is given credence over genuine biological states. It does not matter whether a person perceives his or her body as being part of him or herself — it matters whether it has the correct biological relationship. Were a limb to be truly “dead” — that is, with no functioning biological relationship to any other part of the body — then this would make it no longer a part of the person, despite any physical attachment.

Because my view says that all matter that bears a sufficient relation to a single mind counts as a single body, it does not allow for a mind with two bodies. It does, however, allow for a body “shared” by two distinct minds, thus constituting two distinct persons. Abigail and Brittany Hensel are *dicephali*, conjoined twins who each possess

what we could colloquially call a distinct head but share many parts of their bodies. From a TIME Magazine report on the twins (Wallis & Doman, 1996):

“Though they share many organs, including a single large liver, a bladder, intestines and a reproductive tract, their nervous systems are distinct. Tickle Abby on her side anywhere from head to toe, and [Brittany] can't feel it--except along a narrow region on their back where they seem to share sensation. The girls experience separate hungers and separate urges to urinate and sleep.”

According to the same report, the Hensel twins are able to conduct coordinated physical tasks, including walking, running, and swimming. It remains a scientific mystery how exactly they are able to accomplish this without having a single nervous system.

Were my conception of the body based solely on the nervous system and its functions, there would seem to be two minds, two persons, *and* two bodies present in this case. That is why a proper conception of the body must include the entire organism, accepting that some elements of the organism bear sufficient relationship to one another to make them part of the same organism, while not *all* bearing a sufficient relationship to a mind (though at least one must). With the Hensel twins, the entire organism, excepting each of the two minds contained within it, composes a single body that constitutes Abigail and also constitutes Brittany. Even though the mind of Abigail may have no sufficient relationship to the arm on Brittany's side of the body, that arm has a sufficient biological relationship with parts of the organism that *are* sufficiently related to Abigail's mind.

Part V: What Sorts of Persons Are There?

At this point, I have finished my descriptions of what minds, bodies, and persons are. All of these conceptions have focused on function over form, and a consequence of this is that there exist all sorts of creatures which meet my criteria for being persons.

Let us begin with animals. (It is for this reason that Olson and other animalists using the term “animal” to refer to the organism is particularly misleading.) Fido seems to have a mind, and a body, and thus would be a person. The same goes for every animal with conscious experiences. And our intuitions regarding brain transplants, with which my view is in line, support the idea that were Fido’s brain moved into another body, that body would now constitute Fido — even if that body were not a dog body. Animals are as much persons as persons are animals. This shows the animalist thesis “all persons are animals” to be true in a roundabout way that doesn’t quite match their intentions.

And I do not want to simply throw in a DNA criterion to persons to make them distinct from other animals. First of all, this seems not to comply with cases of inter-species brain transplant, should such transplants prove possible. If a human brain were put into another animal’s body, the majority of DNA in that organism would be of the other species. Or suppose, even more outlandishly, that we fused one hemisphere of a human brain with a hemisphere of a dog brain (the identity consequences of fusion will be discussed in the next section). Would the ensuing person be human? This is unclear as well. Finally, supposing a computer or other synthetic object could be developed that can generate consciousness, we would have a person that is not even biological, so we ought not include a biological criterion like DNA.

Perhaps what is necessary to make our discussion here comply with the common intuition that the word person refers to a human is to not use the word person at all. As

we saw with van Inwagen, though, “life” or “mental life” would be too vague. As of this moment, I do not believe that such a term exists to distinguish human persons from other kinds of persons. Tragically, cat-person and dog-person already have non-philosophical meanings, preventing us from using the obvious choice.

Though in some sense, even such a line as this may be inappropriate to draw. At the core of things, we are the same kind of thing as Fido — we are organisms with minds and consciousness. While some may wish to draw a line based on specific cognitive powers (self-conception, moral understanding, rationality), such lines seem to always lead to cases in which they blur. Suppose a human suffers some trauma that damages their mind in such a way they lose one of these capabilities. Are they no longer a person? This, to me, seems like too cruel of an outcome, though perhaps those who argue for such criteria would genuinely accept that there is no person present anymore.

Part VI: Puzzles of Personal Identity

While the view I have laid out thus far offers an option for what we as persons *are* and what we are *constituted of*, I will now explore how the view handles the popular puzzles of personal identity through time. For many, it may be less important to know what we are as persons, and more important to know whether some person in the future or past is the same person as his or her self in the present.

Before addressing the specific puzzles individually, I will describe what I consider to be the conditions for which personal identity persists through time given my view. Because a mind is both necessary and sufficient condition for a person, I see the two as being related such that each distinct mind’s persisting through time guarantees its

corresponding person's persisting through time. If a mind should cease to exist, even if it only does so by replacement with a distinct mind, the person also ceases to exist and a new person comes into being whenever a new mind comes into being. The question we must ask in the face of each of the following puzzles, then, is which minds presented are identical to one another and which are distinct from all the others.

Puzzle 1: Fission

Perhaps the most discussed puzzle is that of fission, or the split-brain. In this problem case, a man's brain is split in half by hemisphere, and each hemisphere is placed into a new body, distinct from the original. This simple yet perplexing example has been brought up by many, including Wiggins, whose use is summarized here by Parfit along with what he considers the potential outcomes in terms of personal identity (Wiggins, 1967) (Parfit, 1971):

Wiggins then imagined his own operation. My brain is divided, and each half is housed in new body. Both resulting people have my character and apparent memories of my life. What happens to me? There seem only three possibilities: (1) I do not survive; (2) I survive as one of the two people; (3) I survive as both. The first and second options are somewhat self-explanatory, but I will clarify the third to mean that the subject survives with two bodies and a divided mind, not with two minds both identical to the original.

I will begin by addressing the second possible outcome, as I feel it is easiest to dismiss. In this case, we suppose equipotent hemispheres of the brain, and given this, there would be no method to distinguish one hemisphere as being a continuation of my mind and the other as not being as such. Even were the hemispheres not equipotent, as is

the case in reality, either both have core psychological continuity with the original or neither does. If neither does, you obviously did not survive at all, and if both do, then it seems silly to say the one of them is not you while the other is, solely because of the degree of continuity present. Were either half of the original mind destroyed, it would be consistent with my view that the person should survive as the remaining half. One could rephrase the fission operation as “destroying each half with regards to the other half.” That is to say, as far as each hemisphere is concerned, the other hemisphere may as well have been destroyed — the distinction between removal and destruction bears no relevance for the hemisphere from whose perspective we consider the other. Given the fact that each hemisphere would survive this on its own, it makes no sense to claim that only *one* of the two survives it when it occurs simultaneously for both.

We ought now to turn to the third possible conclusion, as any argument showing it to fail guarantees the correctness of the first (assuming, as I am, that Parfit is correct in regarding these three possibilities as the only available options). The first conclusion is the choice of many who feel that intuitively there must be personal identity preserved in such a case, but great lengths must be gone to in order to accommodate this view, and those lengths are not necessarily consistent with my view laid out thus far.

But first, let us establish quickly why this is difficult to accept on face value. My conception of the mind does not allow for the “divided mind” as described here. While Unger may believe that the divided brain still has the “capacity” to interact given his aforementioned car example, I am hesitant to say that these two mind meet my criteria regarding interacting parts. That means that there are two minds here, and given my view’s one-to-one correlation between persons and minds, there must be two persons as

well. And these persons cannot both be identical to me because they are not identical to each other. They are in different places at different times, have different experiences at the same time, have different physical shapes, etc.

In order to find a solution that makes this conclusion acceptable, philosophers have had to propose odd conditions to have been true *prior* to the fission operation. David Lewis, among others, proposes that prior to the operation, there were two minds with shared consciousness (Lewis, 2003) (Mills, 1993). Parfit deems this unreasonable on the grounds that we must maintain the ratio of one consciousness to one person, and holders of my view must agree (Parfit, 1976). Mills, in response to this whole debate, argued that the fission patient had two distinct consciousnesses before the operation that were identical in terms of content, but not numerically (Mills, 1993). This is inconsistent with my view on the grounds that there is a unified, self-interacting mind present, yielding one consciousness and thus only one person prior to the operation. This accommodation for accepting the third conclusion of the fission case also appears to fail. At present, I can think of no other twist to put on the case such that my view can consistently agree with the third conclusion.

We appear to have fallen into the first conclusion, that the original person does not survive. Yet even this appears to have logical problems. As I said earlier, either both minds have continuity with the original or neither does. It seems clear that both *do* have this continuity. How, then, do I not survive if my entire substrate is intact, though separated? It is counterintuitive, but the best possible solution my view can offer in fission cases is as follows: “So long as it can still function as a mind, a mind can survive the irreversible loss of interactivity with some chunk of itself, unless that chunk is also

itself functioning as a mind.” If the fission operation fails and only one hemisphere lives through the operation, you would then persist as that person. But when both survive, they are new persons distinct from you though undeniably similar in many ways. As I said, this is counterintuitive at face value. One alternative that may be easier for some to accept is as follows: “A mind can survive the irreversible loss of interactivity with some chunk of itself so long as it can still function as a mind, unless the chunk *could have* also itself functioned as a mind.” Under this view, a person does *not* survive the destruction of one hemisphere, or any amount of mind that, if separated and hooked up to life-support machinery, would function as its own separate mind. In an odd way, this may actually appeal to common sense. It could be argued that if a person has lost a mind-sized chunk of their own mind, they are changed enough to make them a different person. I tend to find this last view appealing, though some may say it makes too much of certain brain injuries. I believe that either of these conclusions are more acceptable than the idea that personal identity is a concept inconsistent with fission discussions, as Parfit claims (Parfit, 1971).

Puzzle 2: Fusion

Let us suppose now that instead of breaking a mind apart, we are bringing two minds together. There are two different cases regarding fusion which I think it will be necessary to discuss: First, a case in which two normal human brains are fused. Second, a case in which two once-unified hemispheres have been split and are then reunited after some time.

The case for the former I believe to be rather simple. Before the fusion operation, we have two individuals for whom we have no reason whatsoever to claim identity. Suppose we are able to fuse their entire brains together into a four-hemisphered mega-brain. We now have one mind, which means we have only one person. That person cannot be identical to both of the originals, because then the originals would have been identical to each other. And it cannot be identical with only one of the originals for reasons similar to those described in fission. Therefore, the mega-mind and its person are distinct, and neither of the original persons continues to exist, though they have psychological similarities with the new person.

The case with a reuniting of hemispheres that have been split intuitively seems different. If we return once more to the example Unger gives regarding the car, it seems that perhaps there was one person there all along with a divided mind whose capacity to interact was simply temporarily disabled. But I am reluctant to accept this. The moment that the two hemispheres began to have distinct experiences for which they had no discernible means of communicating to each other, they established themselves as new persons. Because of this, I believe that the “reunification” case yields the same outcome as the above case. There is a new post-fusion person, and that person is not identical with either of the pre-fusion hemisphere-persons, nor with the pre-fission “whole”-brained person. In continuing to lay out conditions wherein minds survive, we can create a parallel to what was laid out at the close of the fission section: “A mind is replaced by a new, distinct mind whenever it interacts in a sufficiently unifying way with another self-sufficient mind.” Here, the mere *capacity* to interact with a new mind does not seem enough to change the person. If a man walked around with his mind hooked up to another

a la Brainland but no interaction ever genuinely occurred, it would not seem as though this man's mind had been altered in any way sufficient to render him a different person.

Puzzle 3: Teletransportation

Let us turn now to the puzzle of teletransportation. In such a case, the entire make-up of a person, down to the smallest existing particles, is scanned by a machine. The precise locations, velocities, etc. of those particles are recorded by the machine. That person is then destroyed, and another person is built in another location using the recordings taken by the machine's scan. In the scenario, the person is said to walk, talk, and act like the person who was scanned (Ord, 2002).

The metaphysical possibilities of events unfolding one of several possible other ways than the procedure just described makes me hesitant to accept identity of persons before and after such a teletransportation. Suppose that the original person were never destroyed at all, but that the machine still created some person in the second location. Clearly these people are not identical with one another. They are having different experiences and their minds do not interact. It seems odd to suppose that, had the first person been destroyed, that would have been sufficient to make the two persons involved in the whole operation identical.

By my view, for a person to survive, that person's mind must survive, and the mind is a physical object. In traditional teletransportation, that object is destroyed. Toby Ord describes a case of a more efficient (yet less plausible) teletransporter which physically moves a person instantaneously from one place to another (Ord, 2002). This does not seem metaphysically possible, but were such a thing possible, Ord claims, we

are much more likely to consider the person pre-teletransportation to be identical with the person post-teletransportation. Regarding myself, he is correct. I am willing to accept identity here because at no point was the mind destroyed.

Hershenov conceives of yet another form of teletransportation device which disassembles the person into simples then reassembles the simples perfectly in the second location (Hershenov, 2003). This is the most complex case as far as my view is concerned. Technically, the mind has been destroyed, yet the mind is reassembled perfectly after was could conceivably be a matter of milliseconds. Regarding this case, there still does not seem to be strict *continuity* and thus I am reluctant to claim identity between the two persons. Suppose, for instance, that the machine malfunctioned during teletransportation, as seems to happen all the time in films. The person would be eternally disassembled, and we would be likely to say this person is dead. Given that a person must die to go through teletransportation, it seems unlikely to me that the person created by the machine is identical with the original. However, perhaps it is still possible.

The question is whether it is possible for a brain to constitute a mind, cease to exist entirely, then return to existence as the same brain constituting the same mind. Of all the questions discussed in this paper, this is the one of which I am the least sure. In some cases, it even appears arbitrary. For instance, suppose a child built a house of blocks, then tore it down. A week later, that child builds a house in the exact same spot with the exact same arrangement of the exact same blocks. My intuition tells me that this second house, despite being similar in so many respects to the first, is not identical with the first. Perhaps that is because, no matter how many other properties they share, they seemingly cannot share the property “time at which this object *was created*.” But this

property is not shared only if we are already assuming that the second house is not identical to the first. We could just as easily say that the houses are identical and thus both share the property of being created at some time $t1$, being destroyed at another time $t2$, and reappearing at a third time $t3$.

But then there are also cases of time travel. In such a case there is still a time gap in a person's existence. The difference is that at no point does the person appear to have actually been destroyed or recreated. Perhaps here we must make a distinction between an object's being destroyed at time t and an object's not existing after time t . In the time travel case, it seems that the person might not exist after time t , but it does not seem that the person was destroyed. In the case of the block house, a destruction event did occur. The same goes for Hershenov's teletransportation case.

My intuition tends to favor the idea that the houses are distinct. I think that any time an object comes into existence, it is allowable to say it was "created" at that time. And I do not believe that any single object can have two times at which it was truly "created." Given this intuition, I would say in the teletransportation case that the persons are not identical. Otherwise, the minds would be identical, yet seemingly have two creation times. For the case of revival, it is the same end result. Even though the brain could have only one creation time, it seems the mind it constitutes would have two distinct creation times were the pre-death and post-revival minds identical. Having said this, I want to emphasize that I am open to a view that denies this principle, because other of my intuitions lead me to believe that the persons ought to be identical in one or both of those two cases.

Conclusion

In this paper, I have endeavored to give a consistent view of what we refer to when we use the term “I.” Like many before me, I have called this object a person. I believe that a person is always composed of at least one mind (which itself is a brain essentially functioning to support thought), and along with that brain a person may have other body parts. The body as a whole constitutes both an organism and a person. The acceptance of colocation required to accept this view has been shown as tenable, despite several possible objections from those focused on maintaining a common-sense economy of both objects and thinkers.

I have argued that the capacity for consciousness is essential to persons, and that the continued capacity for consciousness is what makes a person endure through time. I have argued that Unger makes the best argument among prevalent philosophers for what is essential in psychological continuity, and share his belief that it will always go hand in hand with physical continuity in the real world.

Having detailed my view of what criteria are required of brains, minds, bodies, and persons, I have endeavored to show that my view is able to reach a consistent if counterintuitive conclusion when faced with some of the popular puzzles regarding personal identity.

I believe the view I presented regarding persons to be similar to common intuitions. We do not think we are identical to our bodies or brains, and we believe those things will persist after we die. Our intuition tells us that our consciousness makes us who we are and that without consciousness, we would not exist. Wherever my view leads to

consequences that deviate from common sense, I hope to have shown that those consequences are acceptable in a consistent view, even if they are counterintuitive.

Bibliography

- Anderson, S. L. (1976). Consciousness and Numerical Identity of the Person. *Philosophical Studies* , 1-10.
- Baker, L. R. (1997). Why Constitution Is Not Identity. *The Journal of Philosophy* , 94 (12), 599-621.
- Bennett, K. (2004). Spatio-Temporal Coincidence and the Grounding Problem. *Philosophical Studies* , 118 (3), 339-371.
- Burke, M. B. (1994). Preserving the Principle of One Object to a Place: A Novel Account of the Relations Among Objects, Sorts, Sortals, and Persistence Conditions. *Philosophy and Phenomenological Research* , 54 (3), 591-624.
- Campbell, S. (2006). The Conception of a Person as a Series of Mental Events. *Philosophy and Phenomenological Research* , 73 (2), 339-58.
- Campbell, T., & McMahan, J. (2010). Animalism and the varieties of conjoined twinning. *Theoretical Medicine and Bioethics* , 31 (4), 285-301.
- Cochrane, C. N. (1944). *Christianity and Classical Culture: A Study of Thought and Action from Augustus to Augustine*. Oxford: Oxford University Press.
- Descartes, R. (1996). *Meditations on first philosophy: with selections from the Objections and replies*. (J. Cottingham, Ed.) Cambridge: Cambridge University Press.
- Feinberg, T. E., & Keenan, J. P. (2005). Where in the brain is the self? *Consciousness and Cognition* .
- Gibbard, A. (1975). Contingent Identity. *Journal of Philosophical Logic* , 187-221.
- Green, M., & Wikler, D. (1980). Brain Death and Personal Identity. *Philosophy & Public Affairs* , 9 (2), 105-33.
- Hershenov, D. B. (2003). Can There be Spatially Coincident Entities of the Same Kind? *Canadian Journal of Philosophy* , 33 (1), 1-22.
- Hershenov, D. B. (2005). Persons as Proper Parts of Organisms. *Theoria* , 71 (1), 29-37.
- Jackson, F. (1982). Epiphenomenal Qualia. *The Philosophical Quarterly* , 32 (127), 127-36.
- Johnston, M. (1992). Constitution Is Not Identity. *Mind* , 101 (401), 89-105.

- Jubien, M. (1993). *Ontology, Modality and the Fallacy of Referenc*. New York: Cambridge University Press.
- Lewis, D. (2003). Survival and Identity. In R. Martin, & J. Barresi (Eds.), *Personal Identity* (pp. 144-67). Oxford: Blackwell Publishing.
- Lowe, E. J. (1995). Coinciding Objects: In Defence of the 'Standard Account'. *Analysis* , 55 (3), 171-78.
- Margolis, J. (1975). Puccetti on Brains, Minds, and Persons. *Philosophy of Science* , 43 (5), 275-80.
- Markosian, N. (1998). Brutal Composition. *Philosophical Studies* , 211-249.
- Mills, E. (1993). Dividing Without Reducing: Bodily Fission and Personal Identit. *Mind* , 102 (405), 37-51.
- Nagel, T. (1971). Brain Bisection and the Unity of Consciousness. *Synthese* , 396-413.
- Noonan, H. W. (2010). The thinking animal problem and personal pronoun revisionism. *Analysis* , 70 (1), 93-98.
- Oderberg, D. S. (1996). Coincidence under a Sortal. *The Philosophical Review* , 105 (2), 145-171.
- Olson, E. (2003). An Argument for Animalism. In R. Martin, & J. Barresi (Eds.), *Personal Identity* (pp. 318-34). Oxford: Blackwell.
- Ord, T. (2002). *Implications of fission, fusion and teletransportation to a view of personal identity through psychological continuity*. Retrieved 2011 from amirrorclear.net: <http://www.amirrorclear.net/academic/ideas...hrough-time.pdf>
- Parfit, D. (1976). Lewis, Perry and What Matters. In A. Rorty, *The Identities of Persons* (pp. 91-107). Berkeley: University of California Press.
- Parfit, D. (1971). Personal Identity. *The Philosophical Review* , 80 (1), 3-27.
- Parfit, D. (1993). *Reasons and Persons*. Oxford: Oxford University Press.
- Puccetti, R. (1973). Brain Bisection and Personal Identity. *The British Journal for the Philosophy of Science* , 24 (4), 339-55.
- Shoemaker, S. (1959). Personal Identity and Memory. *The Journal of Philosophy* , 56 (22), 868-82.
- Shoemaker, S. (2008). Persons, animals, and identity. *Synthese* , 313-32.

- Shoemaker, S. (1992). Unger's Psychological Continuity Theory. *Philosophy and Phenomenological Research*, 52 (1), 139-43.
- Thomasson, A. L. (2006). Metaphysical Arguments against Ordinary Objects. *Philosophical Quarterly*, 56 (224), 340-359.
- Thomson, J. J. (1998). The Statue and the Clay. *Noûs*, 32 (2), 149-173.
- Unger, P. (1990). *Identity, Consciousness, and Value*. Oxford: Oxford University Press.
- Unger, P. (1992). Reply to Reviewers. *Philosophy and Phenomenological Research*, 52 (1), 159-76.
- van Inwagen, P. (1990). *Material Beings*. Ithaca: Cornell University Press.
- van Inwagen, P. (1987). When Are Objects Parts? *Philosophical Perspectives*, 1, 22-47.
- Villarreal, L. P. (2004). *Are Viruses Alive?* Retrieved April 26, 2011, from <http://www.uvm.edu/~biology/Classes/011/alive.pdf>
- Wallis, C., & Doman, J. M. (1996 йил 25-March). The Most Intimate Bond. *TIME Magazine*.
- Wigan, A. (1844, 1985). *The Duality of the Mind*. Malibu: Joseph Simon.
- Wiggins, D. (1967). *Identity and Spatio-Temporal Continuity*. Oxford: Basil Blackwell.
- Wiggins, D. (1968). On Being in the Same Place at the Same Time. *The Philosophical Review*, 77 (1), 1968.
- Zimmerman, D. W. (1995). Theories of Masses and Problems of Constitutio. *The Philosophical Review*, 104 (1), 53-110.

Acknowledgements

I would like to thank my parents, first and foremost. Without them I would certainly never have gotten to this point. Thank you, mom, for raving about the legendary Carl Cohen, which led me to his freshman seminar which in a roundabout way led me here. Thank you, dad, for having done an honors thesis yourself and having worked down to the wire, thus making me feel better for having to do the same. And thank you both for being so supportive, encouraging me to write this thesis and reminding me when times got hard that the only person I really needed to satisfy with it was myself.

I definitely want to thank David Manley for being a cool, down-to-Earth guy willing to give tough criticism when necessary, while still making me feel like a philosopher and not some amateur. Thank you for reading my rambling emails and my sometimes ill-organized drafts and helping me turn them (the drafts, not the emails) into a scholarly work of which I can be proud. Also thank you for thinking so highly of my project from Knowledge and Reality. That was one of several indicators that should have led me to comedy writing a bit earlier, though I'm glad I didn't give up on philosophy.

Also, to the library staff at all the libraries, but especially to Molly Mahoney at Tanner, thank you for teaching me how to navigate the treacherous world of published philosophy and find the articles that were actually relevant to me. If it weren't for you all, I'd likely have gotten lost in a sea of non-Scholar Google results, never to be seen again.

Finally, I would like to thank the entire Philosophy department at the University of Michigan. The classes and professors here have helped me to think in ways I didn't even know I could, challenge my assumptions about the world, and win pretty much every debate or argument with my friends. I have been fortunate enough to spend four years on one of the overall smartest campuses in the world, and this department is, without a doubt, a hub of that intelligence.