

Chapter 9

Embryos, Four-Dimensionalism, and Moral Status

David Hershenov

Professor of Philosophy and Chair, University of Buffalo

I. Introduction

Philosophy journals have been full of discussions of Four-Dimensionalism in recent years. The rich resources of the Four-Dimensional metaphysics have been brought to bear upon many traditional philosophical problems. Alas, the implications of Four-Dimensionalism for bioethics have gone largely unexplored. Hudson (1999, 2001) is the rare exception. Relying upon a Four-Dimensional metaphysics, he argues that there is little reason to identify the human embryonic animal and human person. He makes the intriguing claim that if abortion is wrong, then it isn't because the human animal within its mother's womb is a person. This he rightly claims "is a very significant result" for "an overwhelming amount of the literature on abortion and infanticide (as well as much of the public debate on these topics) seems to turn on the question of whether or not the human fetus is a person" (2001, p. 153).

Hudson admits that if he were convinced of the truth of Three-Dimensionalism, then he would find it more compelling to identify human persons and human animals than accept that they are distinct entities though composed of the same atoms (2001, p. 130). Such an identification would lead him to claim that the human person is an animal with biological persistence conditions, coming into existence with the onset of life and going out of existence when life is extinguished. So if Hudson were an advocate of Three-Dimensionalism, then he might very well agree with such staunch defenders of human embryonic life as Tollefsen and George who argue that mindless embryos shouldn't be aborted because they are persons (George and Tollefsen, 2008). But if Hudson's favored Four-Dimensionalist metaphysics is true, as well as his claim that the approach removes the reasons to identify the person and animal, then perhaps it is George and Tollefsen who should drop their opposition to abortion for they acknowledge that the exalted moral status of embryos depends upon their identity with persons.

...embryos clearly cannot yet think, choose and speak. Nor are they (yet) self-conscious or even sentient. Were this to mean that embryos were not the same *kind* of beings as the readers and authors of this book, that they were not *persons*, then it would be difficult to see why they should be accorded the same moral respect that we authors and readers believe we are entitled to. There would be no obvious reason why they should not be destroyed for the sake of beings who really are persons (George and Tollefsen, 2008, p. 61).

However, I'm going to argue that opponents of abortion like George and Tollefsen can accept Hudson's metaphysics without having to abandon their belief that human animals are persons. One doesn't have to deny the truth of the Four-Dimensionalist metaphysics to consider mindless embryonic human animals to be persons. If this claim is correct, it would be good news for the pro-life movement since Four-Dimensionalism is likely to attract many followers given its very able defenders.¹

The first half of the paper will be a response to Hudson's claim that human persons can't be identified with the animals who are mindless for part of their existence. Hudson argues that if having later thinking stages were sufficient for being a person, then there would be countless entities that are persons. This assumes unrestricted composition, a principle that I will accept for the sake of argument. Hudson understands unrestricted composition to mean that necessarily for any collection of objects, the *x*s, there exists one and only object, *y*, such that the *x*s compose *y*. So there's even an object that is composed of Stonehenge and the reader.

I'll argue for person/animal identity by distinguishing the kinds of entities that have mindless and thinking stages. The causal relationship between the stages of entities that belong to a natural kind will serve to distinguish the embryonic human animal and person from other Four-Dimensional objects that likewise are mindless at one time but later think. This will leave us with two good candidates for the title "human person" – the human animal that's initially mindless and then later self-conscious, and the entity favored by Hudson that's capable of self-conscious reflection at every stage of its existence.

With which of the two candidates are we to be identified? I'll argue that our intuitions about the persistence of persons are best explained by appeal to a biological (or animalist) account of personal identity. Our intuitions that we would survive certain hypothetical changes as indicated by what appears to be prudential concern for the resulting individual can't be accounted for in terms of the persistence of a capacity for self-conscious reflection or ties of psychological connections and continuity. My contention is that only an appeal to a criterion that identifies us with a future thinker in virtue of sharing the same biological life can make sense of such responses. So it will be argued that of the countless Four-Dimensional entities that have thinking temporal parts, we're to be identified with the living human animal. Moreover, if any beings warrant the label of 'person', we do. Then by helping ourselves to Hudson's *maximality principle* which rules out the existence of a person embedded within another person, we can thus judge any entity within the typical human animal that consists solely of self-conscious temporal parts to be not a person but a proper part of the human person.

II. The Nature and Appeal of Four-Dimensionalism

Before looking at Hudson's reasons for distinguishing embryonic human animals from human persons, there is a need to sketch the core features of the Four-Dimensionalist metaphysics. The Four-Dimensionalist claims that for any period of your life, like the first half of it, there's a part of you that existed only then. Thus the reader has not only such familiar spatial parts as her hands and feet, but such temporal parts as the first half of her existence, the first day of her life, and even momentary temporal parts known as stages. Temporal parts are the distinctive components of Four-Dimensionalism. Informally, a temporal part of an entity will exist only at a time and will then overlap all of the entity's other parts that exist at that time. More formally: something is a temporal part of x during interval T if and only if (i) the object exists at but only at times in T , (ii) it's part of x at every time during T , and (iii) at every moment during T it overlaps everything that's part of x at that moment (Hudson, 2001, p. 59).² So your arm at this moment is not a temporal part of yours because it doesn't overlap all of your other present parts – legs, trunk, head etc. Once temporal parts are understood, Four-Dimensionalism can be defined as the view that “Necessarily, each spatial-temporal object has a temporal part at every moment of its existence” (Hudson, 2001, p. 59). An entity is said to *perdure* if it persists in virtue of having temporal parts.

Three-Dimensionalism is a view that denies that we have temporal parts, are extended in time, and persist by perduring. Rather, we are said to persist by *enduring*, being *wholly present* at each time that we exist. Metaphors abound in accounting for the view. It is said that “we sweep through time” rather than have different parts “spread across time” occupying different periods. However, it won't matter for our purposes to present a precise account of Three-Dimensionalism since our aim is to show that Four-Dimensional human persons are best construed as human animals.

Many reasons have been given to favor Four-Dimensionalism over Three-Dimensionalism. Hudson thinks the best defense of Four-Dimensionalism involves showing how it deals with problems of material constitution like the lump and the statue, fission, and embedded objects (Hudson, 2001, p. 58). I'll just discuss the latter two problems to give the reader unfamiliar with the subject some sense of the appeal of Four-Dimensionalism.

A widely shared intuition is that if your cerebrum is removed and transplanted into the empty skull of your clone where it resumes functioning, preserving your pre-transplant psychology, then you would have switched bodies. Such judgments suggest an implicit acceptance of a psychological criterion for identity: *Person x is identical to y if and only if x and y are psychologically continuous.*

An even more widely shared belief is that people survive with just one cerebral hemisphere when the other is destroyed by cancer or stroke. Now consider the possibility of your cerebrum being split. Then one cerebral hemisphere is transplanted into the brainless body of a clone while the other hemisphere, along with the rest of the original person's body, is destroyed. A common response is, again, that one has switched bodies. The person with the cloned body has your psychology and thus seems to be you.

Now consider that neither of your cerebral hemispheres were destroyed but both were separated and successfully transplanted into separate bodies, each a perfect clone of your animal prior to brain division. If each transplanted half is psychologically continuous with and causally dependent upon the pre-division person, then it seems that you would be two people since the pre-fission person is psychologically continuous with both. But one cannot be both persons if they are not identical to each other. So it seems as if the Three-Dimensionalist has to modify the earlier stated psychological criterion and assert either: (a) the rather arbitrary claim that one of the two distinct persons is the original, or (b) that a person can have a divided mind and what looks like distinct persons is each just half of a person, or (c) maintain that the

person fissions out of existence like a cell when it divides, despite both products of fission consisting of sufficient psychology to be identical with the pre-fission person in the absence of the other. It is difficult to explain in (c) why the being that fissions out of existence would have earlier expressed prudence-like concern for two beings, neither of whom with which he is identical.

The Four-Dimensionalist avoids these three unappealing options. He can claim that there were two people all along sharing temporal parts prior to fission. They overlapped, just as two roads can overlap before forking off. Since no one goes out of existence with fission, it is unsurprising that prior to fission there is a prudence-like concern for the two persons separated by fission. And the post-fission persons are not in error when claiming to recollect and be responsible for actions they performed prior to fission.

The same notion of temporal part sharing offers a solution to the problem of embedded objects. Consider Tibbles and Tibs. Tibbles is a cat and Tibs is an entity embedded within Tibbles that includes every spatial part of Tibbles but the tail. If Tibbles later loses his tail, then it would seem that Tibbles and Tibs have come to have all their parts in common. If Tibbles is a cat, why then is not Tibs? Tibs can purr and meow and do whatever else Tibbles can. But then we would have two spatially coincident cats, i.e., each wholly occupying the same place at the same time. If there cannot be two objects in the same place at the same time, then they are identical. But Tibbles and Tibs cannot be identical for the former *had* a part that the latter did not.

The Three-Dimensionalist can instead claim there are not such things as undetached parts like tails and thus no problem of Tibs and Tibbles (van Inwagen, 1990). But if there are not undetached tails, why think there are such things as hands, head and brains? However, if Tibbles and Tibs are considered to share a temporal part, then there is no danger of two objects of the same kind being co-located. Tibbles and Tibs have some temporal parts in common after the former loses his tail. But before they came to share stages, they merely overlapped, Tibs's spatial parts being proper parts of Tibbles. The Four-Dimensionalist's acceptance of temporal parts allows him to avoid such unwelcome choices confronting the Three-Dimensionalist: the spatial coincidence of Tibs and Tibbles after the loss of the latter's tail, or denying the existence of undetached parts. Instead, the Four-Dimensionalist claims that Tibs and Tibbles are distinct entities that merely overlap for a time after the tail is detached. They are, in a sense, a temporal version of the familiar case of two distinct roads merging for a stretch before they fork apart.

Philosophers who affirm the existence of undetached parts and deny that there are spatially coincident objects are likely to accept unrestricted composition. We will see shortly why accepting unrestricted composition poses a problem for Three-Dimensional accounts of persons that fails to trouble Four-Dimensionalist accounts. One might try to reject unrestricted composition but it is not easy to put forth a principle of restricted composition that does not eliminate much of the ordinary world of folk ontology (brains, chairs, mountains etc.)

Hudson thinks a defense of unrestricted composition can be made through analyzing both the intuitions behind those who reject it, and showing the costs of doing so. He suspects that once people become clear about why they find unrestricted composition counterintuitive, as well as come to realize the costs of trying to restrict composition to allow just the objects of commonsense folk ontology, they will recognize that endeavor to be a failure (Hudson, 2001, p. 107). If composition is not unrestricted, then there will be either no composites (only elementary particles exist) or many of the objects posited by commonsense will be denied reality. For example, the best known account of restricted composition, van Inwagen's proposal (van Inwagen, 1990), denies the existence of every inanimate composite! Hudson maintains it is far preferable to keep chairs and mountains in one's ontology than abandon them to exclude something composed of the extant copies of the Guttenberg Bible, the ruins at Stonehenge, and all the silk in the world. There may be no name for the Bibles, ruins and silk, but that is because there is no human interest in them. Hudson insists that there is not a good reason to so anthropomorphically restrict an account of what exists. Not only do our interests change, but there could even be aliens with interests in what we take to be pointless objects.

If revealing people's motives for excluding the uninteresting from their ontology does not convince readers to accept unrestricted composition, Hudson's next strategy is to show that there is no principled way to allow some scattered objects and not others (Hudson, 2001, p. 108). Not only are atoms and molecules scattered objects, but we, as well as the entire animal kingdom, are swarms of scattered microscopic particles. And scattered at the macro level are Hawaii, the various planetary systems, the letter "i", and an hour glass with its sand falling. So once one admits such scattered objects, it becomes very hard to produce a principled distinction that denies the existence of something composed of the ruins, silk and Bible. Better to accept the latter than leave Hawaii and the Milky Way out of one's ontology.

A Three-Dimensionalist is not at liberty to accept unrestricted composition without some very unwelcome results (Olson, 2008, pp. 229-31). The reason is that the parts that compose you at any moment will soon be somewhat dispersed as you exhale, perspire etc. If you do not want to have to conceive of yourself as a scattered object – and one that exists as long as your component atoms do – then it is better to understand yourself as composed of temporal parts. The alternative is either: (a) to modify the principle of unrestricted composition so the same matter can compose

spatially coincident entities, thus allowing it to be so that you were never identical with the aggregate of your constituent atoms, or (b) to accept that you are in *human* form for only a brief part of your existence, somewhat more scattered before and after that time. The former renders it a mystery why two physically indistinguishable objects in the same place at the same time have different properties. The latter runs afoul of commonsense ontology which takes it that you survive a change of parts, rather than become somewhat scattered, your previous location occupied by a distinct object which appears as you did moments earlier.

III. Why Four-Dimensional Human Animals Do Not Appear to be Persons

Let's now turn to the reason why if composition is unrestricted, Hudson insists the person cannot be identified with the animal. He maintains that animals are not persons for, at best, they would merely share *some* of the thinking stages composing a person. The normal human animal typically consists of thinking stages through most of its life and non-thinking stages during its embryonic months. There is also a distinct perduring creature, perhaps a large temporal part of the animal, whose temporal parts are all capable of thought. Which one is the person? Hudson finds it more plausible to identify the person with the entity consisting of *only* thinking stages rather than something like the animal which also has non-thinking temporal parts. However, such an unqualified principle would mean that your temporal part that exists for the duration that you are reading this sentence would be a person embedded within you. Since there are countless things that consist of only thinking temporal parts in a Four-Dimensional metaphysics, Hudson contends that the only non-arbitrary selection of stages to be given the label "person" are those thinking stages which are *not* embedded within a larger thinking being. Thus persons are *maximal* thinkers.

And it is not any kind of thinking, such as that produced by *merely* sentient stages, that is sufficient for composing a maximal person. Also needed are self-conscious thoughts appropriately related via psychological continuity and connectedness (Hudson, 2001, pp. 122, 130-31, 144). Even that is not enough because self-conscious Hannah could be vaporized in an atomic explosion and in an incredible cosmic coincidence, a psychological duplicate of her materializes on a distant planet (Hudson, 2001, p. 132). So there must also be the right kind of causal connection, an *immanent cause* involving earlier thinking stages bringing about later thinking stages (Hudson, 2001, pp. 134-36).

Hudson denies that the animal's *potential* to have a thinking stage will succeed in rendering it identical to the person. He acknowledges that might work for a Three-Dimensionalist metaphysics that restricts composition and denies that there exist spatially coincident entities (2001, pp. 125-26, 152-53). On such an account of persistence, it is better to maintain that the mindless embryo is identical to the later minimally sentient newborn, self-conscious adult, senile geriatric and irreversibly comatose patient. But the Four-Dimensionalist typically accepts unrestricted composition and so there will be countless objects that have thinking stages. There will even be an object that consists of an ancient Babylonian sandal that disintegrated in 8th century BC and President Obama. If mindless embryos are persons because they have later thinking parts, then there was an additional person present in ancient Babylonia in virtue of the sandal that was an early temporal part of an object composed of it and the later thinking Obama stages. But surely that object is not the earlier part of a person endowed with the moral status that typically accompanies personhood. When only the sandal was present centuries ago, if it had gone without repairs, surely nothing of moral status had been neglected. Or if Obama had been aborted, while *he* might have been deprived of a valuable future as Marquis (1989) famously argued, it seems wrong to claim that the ancient sandal/modern president *composite* was deprived of a valuable future. If not, every death will actually be a grave deprivation and rights violation of countless creatures.

IV. Little Persons

A. Natural Potential

Some readers might think that certain well-known conditions associated with accounts of diachronic identity such as (macro-level) spatial-temporal continuity or active and intrinsic potential could meet Hudson's challenge. I doubt it, though I would acknowledge that the sandal/Obama composite can be excluded from the ranks of persons if one insists that it is only the potential of *spatial-temporal* continuous entities that will allow something once mindless to be

considered a person. Nevertheless, the spatial-temporal continuous object composed of the unfertilized egg stages, the fertilized egg, and its later post-fertilization animal stages would still qualify as a person. It does not matter that the egg does not survive fertilization, for given unrestricted composition, there will be an object consisting of the stages of the unfertilized egg and then the later stages of the animal that begins with fertilization. One might instead deny that the unfertilized egg/animal composite is a person in virtue of its potential for it lacks the *active* potential to develop into a thinking being – the sperm being needed to give it the intrinsic, active power to so develop.³ However, such a criterion will not rule out the potential for thought (and thus personhood) of the object consisting of the stages of *both* gametes and then the fertilized zygote and its later animal stages. Since the entity consisting of the unfused gametes already exists in the supportive uterine environment that a new embryo needs to develop, it is hard to distinguish it from the fertilized egg being a potential person given that they will both develop into thinking beings in that same environment. It will not help to combine active potential and spatial-temporal continuity because there will be an entity that consists of the sperm and egg in contact, just prior to fertilization, and then the later animal stages.

So the dilemma Hudson presents us with is that if we want to deny that a person exists in ancient Babylonia due to the sandal existing there and it possessing the later Obama thinking stages, then we must also deny that the mindless embryo is a person in virtue of its future thinking stages. My response is to suggest there are grounds for claiming that some but not all potential thinkers are themselves persons even when they have not yet manifested that potential. There is a way to distinguish potential thinkers via the relations unifying the stages of a natural kind in order to then claim that the animal is also a person while other beings that have thinking stages for just some of their existence are not persons. The idea is roughly that the mindless embryonic stages are the same kind of stages of the latter thinking person – i.e., they are all animal stages. There are mindless animal stages linked by life processes to thinking animal stages. They are all living stages of an animal. Their diachronic (as well as synchronic) unity is due to their parts being caught up in the same life processes.⁴ They are stages of the same token of a natural kind, not parts of two things of distinct kinds cobbled together in virtue of the principle of unrestricted composition.⁵ The gerrymandered entity composed of the ancient sandal and President Obama does not have later *sandal* stages that happen to think. This suggests an explanation of why it seems much less plausible to ascribe the capacity of thought to the sandal-Obama entity when only the sandal is present than it is to so ascribe it to the mindless stages of Obama. The capacity is not found in the developmental telos of the sandal. It is not the nature of the earlier stages to give rise to later thinking stages. Compare the sandal/Obama composite with Obama himself. One finds a telos programmed into all the stages of Obama, even the mindless ones.

So the idea is that there's a principled distinction between things that have thinking parts at one time in their existence but not at another. The mindless animal stages that are part of a later thinking thing are stages of one and the same animal. The later thinking stages are also animal stages united by life processes. But the mindless sandal stages are not part of an entity that later thinks composed then of sandal stages. I suspect only the human animal will have its *mindless* and *thinking* stages bound by the same unity relation. And the reason there is no *animal* composed of you up to this moment and another reader after this moment is that there is not the appropriate *immanent causation* characteristic of life processes, the earlier stages of a life causing the successive stages of the same life. Likewise for the composite of the scattered gametes and the reader that resulted from their fusion. There are *three* lives involved. The *same* life does not link them diachronically or synchronically.

So we can grant that mindless human animals are persons without having to bestow that title on every object which has mindless stages preceding its thinking ones. However, there may also be a single relation, psychological continuity, unifying all of the thinking stages of the person. So Hudson could appeal to the existence of a non-gerrymandered, 'natural', psychological unity relation in order to apply the label 'person' to the perduring object consisting of only thinking stages. Thus my approach might seem to commit us to there being two kinds of persons – some that are mindless for a time, others that are always thinking. And that admission will run afoul of the maximality principle that persons are not to be found within larger persons (Hudson, 2001, p. 121). However, I shall put forth arguments in part V that reveal the best candidate for the label 'person' to be the one that was once a very little mindless animal. So it will not be, as Hudson claims, "arbitrary and unmotivated" to identify the human person and animal.

B. Persons are Composed of those Parts Contributing to the Production of Thought

I have so far ignored another Hudson-inspired argument regarding why we should not consider the mindless embryonic animal to be a person on the grounds that it has potential to develop into a thinker. Hudson makes the surprising claim that perduring persons are not temporal parts of animals. Rather, persons are "certain proper temporal parts of the brain

and central nervous system of living human organisms” (2001 p. 147). The basis for this claim is that the entire animal does not produce thought, merely part of it does. Hudson writes: “The best we can claim for such an object is that it has some further object, x, as a proper part, and that each of x’s parts plays some sort of role in furnishing x with the full range of features relevant to personhood” (2001, p. 143). This leads Hudson to declare that the person is to be found “within the lifespan and beneath the skin... of the human animal” (2007, p. 220). So if no mindless animal is ever going to develop to where it can *directly* produce thought, there is little reason to identify the person with the animal who will, at best, come to think only derivatively in virtue of some of its parts *really* doing the thinking.

Hudson insists that just as it is unprincipled to identify the person with a perduring animal that possesses non-thinking temporal parts, so is it to identify the person with any of the temporal parts of the animal since many of the animal’s spatial parts are uninvolved with thought production. Hudson comments: “Rather, once again, the only non-arbitrary choice would be an object each of whose parts plays a *contributory role* in supporting a psychological profile constitutive of personhood” (Hudson, 2007, p. 224). Even though Hudson cannot say exactly which parts are so involved, nevertheless, he claims that since he can rule out “such parts as one’s forearm (2001, p. 219)...some parts of the hand...” (2007, pp. 224-25), “finger nails and bone-marrow...” (2001, pp. 143-44) as making a contribution to thought, that is enough to sustain his thesis that persons are not temporal parts of animals.

I doubt that Hudson can rely upon the notion of “a contributory role” in supporting personhood to so shrink the size of the person. One reason has been presented by Eric Olson, who speaks of *direct involvement* with the production of thought instead of the near equivalent *contributory role*. Olson thinks the real problem with brain-size persons is that little sense can be made of the idea of “direct involvement in a being’s thinking” that motivates the position (Olson 2008, pp. 91-98). Olson wonders why if the respiratory and circulatory systems are not *directly* involved with thought, we should consider the oxygenated blood vessels in the brain to be so? Olson suggests that someone might maintain that the thought is really produced by the firing of neurons. However, Olson points out that not every part of the neuron is similarly involved in the sending of electrical or chemical messages to other neurons. Some serve other tasks like maintaining structural integrity of the cell or removal of its wastes. This, Olson claims, ought to make “the thinking minimalist uneasy” (2008, p. 92). Moreover, the neurons will not fire without these tasks being performed. Olson cautions that trying to determine what is *directly involved* in the production of thought is as hopeless as trying to determine which of the many workers, suppliers, managers, tools and materials is directly involved with the factory production of a knife, or which parts of the body are directly involved with walking. He insists that the problem is not even one of vagueness – it is not that we have a clear application and then boundary cases. Instead, the fault lies in the notion of *directly involved* being unprincipled.

My suspicion is that Hudson is led by the truth that our thought could continue if our animal is reduced in size to assert the falsehood that such removals show that it is only some of the parts of the animal that produce thought (Hershenov, forthcoming). The animal needs to be alive to think. The fact that the event of someone’s biological life could configure less material than it does is irrelevant. While it is true that one’s animal can become smaller, that does not mean that the life processes which make thought possible were not earlier an event of a larger substance. Since one’s thoughts depend upon the contribution of such processes, wherever they are located, so is the thinker of those thoughts to be found. And that life is dispersed throughout the body.

The basis for denying that we thinkers are merely parts of animals does not just lie in the fact that the living body contributes the life support necessary for any brain activity. The neuroscientist Antonio Damasio argues that the brain’s constant monitoring of the body, its receiving and sending of the messages, is necessary for the working of the normal mind (Damasio, 1994, pp. 223-44). Even *partially* cutting off inputs to the brain in those suffering spinal chord injuries causes changes in the state of mind. Damasio’s acceptance of “the idea that the mind derives from the entire organism as an ensemble” (1994, p. 225) leads him to reject the assumptions underlying one of philosophy’s most famous thought experiments - the brain in the vat. He claims the disembodied brain floating in a vat of nutrients, without perfect duplication of the inputs and stimuli outputs, might not even be able to think. For similar reasons we should reject Hudson’s view which amounts to considering the person to be “a brain in a living vat.” Damasio explains:

In brief, neural circuits represent the organism continuously, as it is perturbed by stimuli from the physical and sociocultural environments, and as it acts on those environments. If the basic topic of those representations were not an organism anchored in the body, we might have some form of mind, but I doubt that it would be the mind we do have...the body contributes more than life support and modulatory effects to the brain. It contributes a *content* that is part and parcel of the workings of the normal mind (1994, p. 226).

V. The Human Animal is the Only Person

A. *The Collapse of Psychological Continuity into Biological Continuity*

I'll now offer a second line of reasoning for identifying the Four-Dimensional human animal and the human person. I'll show that the psychological continuity and connectedness criterion, some version of which is favored by Hudson (2001, p. 144) *collapses* into animal identity. What I mean by *collapses* is that there are cases which tend to elicit from us descriptions of one thinking entity being identified with another thinker that cannot be explained by a psychological criterion being satisfied. Such intuitions can only be accounted for by both thinkers being the same animal. So what we want to say are stages of a persisting person in cases involving the dreaming and the awake, the rational and the demented, divided and reunited minds, can only be construed as such if an appeal is made to the biological persistence conditions of animals.

The first problem for the psychological account of identity involves a twist on Reid's famous critique of Locke's memory criterion (Perry, 1975). Locke claimed that one's identity extended as far back in time as one's memories. Reid revealed a failure of transitivity by envisioning an old general who could remember his first military campaign as a young soldier, the young soldier could recall being a schoolboy flogged for stealing from an orchard, but the general could not remember being flogged. Therefore, the general *is not* identical to the boy, yet he's identical to the young soldier, who's identical to the boy. This absurdity could be avoided by appealing to psychological continuity, i.e., overlapping chains of psychological connections (Parfit, 1983, pp. 206-06). Psychological continuity involves the general being able to remember a time at which he could remember being flogged. So an overlap of memories will suffice in lieu of a direct memory connection. But the transitivity problem returns with a modified version of Reid's scenario that Perry named the *Senile General* case (1975, p. 19). The senile general could remember being flogged (or remember a time at which he could remember a time that he was flogged). So he is identical to the boy. The young soldier could remember being flogged, so he too is identical to the boy. But the general could not remember his more recent experience as a young soldier, nor could he remember any other time at which he then could remember his first military campaign. This renders the general identical to the boy but *not* identical to the young soldier, who is also identical to the boy. So if they are to be identified, as it intuitively seems they should, an appeal to their being the same animal can do what an appeal to psychological continuity cannot.

A second scenario where a psychological criterion of identity collapses into a biological one involves a temporary division of a mind. Consider Parfit's *My Physics Exam* scenario where there is just a short term loss of a unified consciousness due to cutting the corpus callosum so different cerebral hemispheres contribute to answering distinct problems (1983, pp. 246-48). The hemispheres are reunited after the dual work is done. As Parfit himself notes, the most plausible response is that there was one person temporarily cut off from himself. To account for that intuition, something other than a single causal chain of psychological continuity must be relied upon. Four-Dimensionalists usually qualify the criterion of psychological continuity for branching cases so the result is that there are two distinct persons continuous with the same earlier stage. They do so by insisting that psychologically continuous *x* and *y* are stages of the same person if there is no stage *z* that is psychologically continuous with *x* or *y* but simultaneous and distinct from either *y* or *x* (Brueckner and Buford, 2008). So during the exam there are two streams of thought that have stages that are simultaneous but distinct from each other, thus ensuring that there is not a single person despite their both being psychologically continuous with shared earlier stages. But this will deliver the counterintuitive result that there is not one person with the briefly divided mind but that there were two persons present at that time since they involve simultaneous but distinct stages. If the intuitive response is to be preserved, then it appears that we must appeal to a rather ad hoc modification of the psychological criterion or claim that it must be because the same animal is doing the thinking.

One can also undermine the psychological continuity criterion for identity by taking issue with Locke's account of Socrates awake and Socrates asleep (Locke, 1975, p. 343). Locke conjectured that if sleeping Socrates was psychologically cut off from waking Socrates then they would not be the same person. Imagine that your waking and dream states are not psychologically connected. You cannot recall your dreams and these dreams do not follow from your waking life. I suspect that few readers would follow Locke and deny that they were states of the same person, interpreting the psychological disconnect as evidence of two people sharing a body. Since there is not any psychological continuity between the waking and the sleeping, then what makes them the same person must be that they are the same living animal. It helps if readers imagine that medical technology reveals horrible nightmares occur when they are

asleep though they never recall them. I suspect that if they could prevent these nightmares by doing something when awake, they would, and for prudential reasons, since no moral concerns arise here about the suffering of others.

A fourth scenario undermining psychological continuity theories relies upon our reactions now to the possibility of future pain after the onset of amnesia or even more debilitating impairments (Unger, 2000). Consider the prudential concern many envision having for the individual with their brain after a stroke undermines the brain's capacities for rationality and self-consciousness, leaving a mere sentient child-like mind. If told earlier that the individual with our damaged brain will suffer horrific pains unless we take on almost as much physical pain before losing our memories and capacity for self-consciousness, most of us would consent to the lesser pain to ensure the greater does not transpire. Such a show of apparently prudential concern for an animal in the future, despite the absence of psychological continuity and the reflective capacities associated with personhood, suggests an adherence to an animalist/biological account of our identity.

What I have been hoping to get readers to recognize with the *Sleeping Socrates*, *Physics Exam*, *Senile General* and future pain scenarios is that there is a divergence between the psychological criterion and our intuitions about our survival. Only a reliance upon animalist identity conditions can accommodate our judgments of persistence. I suspect that some readers will offer an alternative interpretation. Their response is that it is not psychological continuity that matters to our persistence, but the capacity for mere sentience – minimal thought and feeling. As long as the same brain sustains sentience, then the individual survives despite memory loss and even some mental fragmentation.

Hudson contends that an individual suffering “profound senility” would not be a person (2007, p. 222). There would not be the requisite self-consciousness and psychological continuity. But our prudential concern in the well-known thought experiments suggests we would survive such mental incapacitation.⁶ So while I think this should lead Hudson to abandon his belief that we are essentially self-conscious persons (2007, p. 218), given unrestricted composition, it need not lead him to deny that there are beings that are essentially self-conscious with psychologically continuous stages. However, if anything deserves the title “person”, we do. So given Hudson's commitment to a maximality principle, he should accept that we persons are not essentially self-conscious psychologically continuous thinkers, merely self-conscious for just a period of our lives. We are persons because of our capacity for self-consciousness, but that capacity is not actualized during all of our stages. Of course, even if Hudson were to admit this, it still would not commit him to identifying the human person and the human animal. He could instead claim we persons are identical to a maximal being composed of all merely conscious stages rather than only self-conscious, psychologically continuous stages. Jeff McMahan and Peter Unger offer Three-Dimensional versions of this thesis (McMahan, 2002; Unger, 2000), claiming that we survive as long as the same brain produces sentience (consciousness). So what I propose to do in the next section is provide thought experiments which suggest that our prudential concern reveals that we persons believe ourselves to be not even essentially sentient. The thought experiments reveal that the future sentient states we are concerned with can be deemed ours only if they are united by a biological criterion.

B. The Collapse of Brain-Based Psychological Identity into Biological Identity

I want now to try to offer another argument to show that a Four-Dimensionalist theorist should claim the human animal is the only person by drawing upon our concern for our stages that are devoid of the traits that characterize personhood. I'll argue that our prudential concern towards our adult conscious animal in the future, including those times when it is without any psychological connections to the present, or even the same brain playing a role subserving our future mental life, suggests that we human people are animals essentially. I'll maintain that once it is established that we could survive certain brain injuries, we can resist the intuitive pull of two famous thought experiments that have provided considerable support to psychological accounts of personal identity. The first involves your brain being destroyed and replaced by a new brain. The second thought experiment involves you swapping brains with another person. Most people judge it to be that we would not survive in the first hypothetical scenario but would do so in the second, though in a different body. I shall try instead to elicit intuitions that in neither scenario do we cease to exist or obtain a new body.

Consider that you would care prudentially for the individual suffering from the results of a stroke that reduced your brain's capacities to realizing mere sentience. Many philosophers believe this shows that it is mere consciousness or sentience, not self-consciousness that is essential to our persistence (McMahan, 2002; Unger, 2000). I think instead that our prudential responses in such scenarios should actually be construed as showing that it is the criterion of biological identity across time that reveals our persistence conditions. Ask yourself whether your concern for your post-injury self with just a rudimentary mind really is due to your possessing the *same organ* that underlies consciousness or is it rather

that it is just the *same animal* that is conscious? I think it is the latter and this can be seen by pondering the following twist. Consider whether your reaction to the prospect of coming out of a stroke-induced coma with pain and pleasure sectors intact but no cognitive capabilities above this will be different if such sentience is a result of different parts of your cerebrum being rewired during the coma to realize pain and pleasure when you awaken?⁷ I suspect that most readers would have prudential concern despite different parts of the brain contributing to such sensations.

If you would have prudential concern for the same animal with different physical structures supporting sentience, then why should you react differently to your animal getting an entirely new cerebrum in the thought experiment in which your original cerebrum is destroyed and a new one imparted? Readers might respond that it matters that the different anatomical structures, without which there would be no sentience, are in the *same* cerebrum. If so, consider a second case where, early in someone's life, in the absence of injury and before a web of beliefs and desires arises, different parts of a developing brain play a role in receiving and processing painful and pleasurable signals. Imagine one is in the brainstem and the other is in the cerebrum. Would it be correct to say there were two thinking beings in one body? And if one is destroyed, is it correct to claim there remains then only one thinking being? My intuitions are that it isn't. And would readers say that there is a new thinking being produced by fusion if there is the later development of a self-conscious person who provides the respective pain or pleasure reports when either the brainstem or cerebrum sector is "stimulated"? I very much doubt it. And for all we know, this is roughly what happens in child development. The initially physically dispersed realization and thus psychologically unrelated fragmented mental states of the baby are only later psychologically united as the older child obtains reflective access to the different states. The child can come to say that "I'm in pain now and earlier had pleasant experiences," reflectively linking what before had been experienced without the capacity for reflection upon those experiences. Even if such conjectured development is not how we actually develop, our reactions to such a counterfactual assumption about ourselves does illuminate what we take ourselves to be: living human animals, rather than brain-unified thinkers.

I do not see any reason to identify ourselves with parts of the consciousness-producing central nervous system (Hudson, 2001; McMahan, 2002), nor with a larger being only if it *continuously* possesses the same functioning brain-like structure (Unger, 2000), rather than holding that these pains and pleasures would be mine because they are subserved by parts caught up in the same biological life and belong to the same animal. It seems arbitrary to insist upon the brain – understood to include brainstem, cerebellum, midbrain and cerebrum etc. – as providing us with ownership of our thoughts given that in the described scenario there are neither causal connections providing psychological connections between the pains and pleasures nor is it the same part of the brain that realizes both. Perhaps if the very same part of our brains served to multiply realize our pains and pleasures, then it would make a little more sense to insist upon that bit of anatomy as being essential to our persistence. But that is not the case. So to then insist that the thoughts are ours as long as *some* part of the brain produces them is unwarranted. The boundaries of the brain have been rather arbitrarily drawn by the authors of anatomy texts rather than determined by a unified function. The brain does many things, only some which involve the fore-mentioned neurological processing of pain and pleasure sensations. Since such sentient activities are not what unifies all the parts of the brain, there are not grounds for claiming we persist as long as somewhere in that brain are the vehicles of such sensations.

Despite the charge of arbitrariness, perhaps one will still insist upon arguing that you would survive with *any* parts of your existing brain contributing to the production of conscious states, but would perish if your brain ceased to exist. I believe opposing intuitions can be elicited. Imagine that now and after a debilitating stroke that your pain is received and realized (in some sense) by the upper spine while pleasure has a cerebral basis. I assume that pondering this prospect does not eliminate our now having prudential concern for the post-stroke creature in pain that lacks the capacity for self-conscious reflection. It seems that the best explanation of why these would be your pains and pleasures is that the parts involved with producing them are caught up in the same life, i.e., they belong to the same animal.

So it appears that the two most prominent psychological criteria of identity (*self-conscious psychological continuity* or a brain-based *mere* consciousness) cannot deliver the intuitive response – that there is but one and the same thinker in the stroke case. What can do so is the animalist account which identifies human persons and human animals. Thus it makes sense to claim that the only person in the stories is the animal. As long as our animal can have pleasures and pains into the future, we have some prudential reason to obtain the former and avoid the latter.

Our attitudes of prudential concern provide additional reasons to reject Hudson's idea that the person consists of that which beneath the skin directly produces thought. Thus even if some sense can be made of the proper part of the animal being what directly produces or contributes to thought, it does not seem to be the entity for which we have prudential concern. If different parts of our animal would later contribute to painful sensations, we would be prudentially concerned with preventing these feelings.

If you share my attitudes to the individuals with maimed or reduced brains, then why maintain that we would have no prudential reason to care about one's animal if it received a new cerebrum in a thought experiment after the old was

destroyed? And if you admit that you have some prudential concern for your animal with a new cerebrum, then you cannot also claim to have prudential concern for the being who would receive your cerebrum in a second thought experiment that involves a brain swap between you and your clone. This is not to deny that you can care about the recipient of your functioning cerebrum even though that person will not be you. I do not even have to endorse the claim that your commitment to the human animal with your original brain ought to be less than your concern for yourself with a new upper brain. My point is just that you cannot have *prudential* concern for both since prudence is *self* concern. So I do not have to claim you are irrational to care about the other person/animal who receives your functioning cerebrum in the transplant swap scenario where you stay behind as an animal with a new upper brain. Such concern would be no more irrational than caring more about your spouse or your child than yourself. Nor do I have to claim, as my fellow animalist Eric Olson once did, that it is not identity that matters but psychological continuity (Olson, 1997).⁸

So once readers see that thinkers are best individuated by life processes, it becomes arbitrary to claim only part of the animal is a person. One can still, on the basis of unrestricted composition, claim that the person consists of only scattered thinking stages of organisms before and after the stroke-induced coma and injury. But the stages of the animal do not have the right causal connections. Such a ‘person’ is an artificial gerrymandered product of the principle of unrestricted composition, not an entity possessing either a natural biological or psychological unity between its stages. Calling such an entity a person would be as suspect as claiming the first half of my life and the second half of your life would compose a person. There is no immanent *mental* causation between the thoughts of the person who suffers the stroke-induced brain damage and temporary coma, and the later pains and pleasures. Likewise for the other scenarios discussed. If immanent causation is needed, then it would be in the form of life processes unifying sleeping and waking Socrates, the senile general and the young thief, the later stroke victim and the earlier rational self, the merely sentient newborn and the later reflective child, or the divided and then reunified mind studying for Parfit’s physics exam. So we see that our prudential intuitions, our belief that we are persons if any entities are, and the maximality principle all serve to indicate that the human animal is the least arbitrary candidate for the persistence of the person in the above cases.

VI. Harming the Mindless

Hudson makes the conditional claim that if the embryonic animal is a person, then it is presumptively wrong to abort it. He denies the antecedent while I affirm it. Hudson just said he would presuppose without argument that persons and (merely) sentient beings have moral status (2001, p. 151). That strikes me as uncontroversial, given that his perduring sentient being and person possess only conscious stages. I do not want to help myself to his assumption since there will be readers who deny that the mindless can be harmed. So it is not enough to show that there is such a distinction between the potential of animal stages to give rise to thinking animal stages while there is no such potential of sandal stages to give rise to later thinking stages. The appropriate immanent causation unifying stages and the developmental telos might seem not to be enough to establish the possibility of harming or benefiting something with moral status. A further argument will have to be made why mindless fetal animals have interests and can be immorally harmed while the sandal/Obama entity cannot be harmed while the sandal stages are present.

Before encountering Hudson’s challenge, I had thought it was part of a sound argument that the mindless would be benefitted and harmed by being identical to a later being whose first mental experience could be good or bad for it. This stood in stark contrast to claims by McMahan (2002) that early abortion was not wrong because there would not be any psychological connections between the mindless and sentient. It seemed to me that if mindless X was identical to a later sentient Y, then if the first mental experiences of Y were good for it, then in virtue of the identity of X and Y, we should also say that it would be good for mindless X to develop and experience such experiences. And it could be a harm if its development was terminated and X never obtained the good experiences.

I did not worry about the sperm or egg, or their scattered composite because, like Marquis and Stone (Marquis, 2004; Stone, 1987), I was working with an ontology in which the latter did not exist and the first two went out of existence at fertilization. So potential mattered, but gametes and contraception did not provide a *reductio* of the view for as Marquis argued “Prior to conception there’s no individual that’s the same individual as the later human being that has, or would have had, a valuable life.... Individual identity doesn’t survive fusion...” (Marquis, 2004, p. 33). But this is just false if we accept unrestricted composition. Marquis and Stone thus cannot claim only embryos possess a morally significant potentiality on the grounds that there are no such creatures that were once gametes and later thinkers. Nor does it help to argue as Stone does that the *human animal* never existed as a gamete or pair of gametes (Stone, 1987, p. 17). There are still objects that have the gametes and the later human animal as proper parts even if none of them is identical to the human animal.

The line of reasoning that I earlier embraced might now seem to be especially dubious when we are discussing an entity that had a sandal stage earlier in its history and sentient stages much later. While it is true that it is one and the same entity that was a sandal for a period and sentient for a later period, it does not seem that the worm with sandal and sentient stages would earlier have been harmed if the sentient stages were prevented from arising. So it seems that if the mindless can be benefited or harmed, there necessarily must be more to account for this than the identity of the mindless with an entity that has a valuable future. We need an explanation for why when the sandal/Obama composite existed centuries ago in virtue of its sandal temporal parts being located at that time, it would not then have been bad for it that its later twenty-first century temporal parts were to have lost the election or to prematurely cease to be conscious.

Why should the mindless fetus have moral status and be capable of being harmed when it seems the ancient sandal/Obama person composite did not have moral status at the time that its temporal parts were mindless? I believe the answer starts with the recognition that mindless animals have *interests*: they have an interest in food and survival and flourishing of a sort. We can speak of things going well for mindless animals, their functioning as they should. They have a good. As creatures with a good, a later mental life can earlier be in their interest. That mental life will serve the animal's interests or telos. Just as other organ systems served to keep the organism alive and flourishing, so will its later cognitive systems. That is not true of the sandal that is part of a sandal/Obama entity. The sandal's functioning properly does not prepare the way for the Obama stages to flourish. The sandal stages do not serve some end or telos (be it survival, reproduction, flourishing, knowing God) that Obama's stages do as well. One cannot speak of such footwear as having any interests or good, so it cannot have an interest in the later well-being of its Obama stages. Any mention of something being good for the sandal, like polish preserving its leather, is parasitical upon serving its wearer's ends which can be furthered by the polish extending the sandal's longevity. But we can speak of thinking stages of Obama being interested in their later thinking stages and the whole they compose. I suggest that we can likewise speak of the mindless human animal having an interest and benefiting from its later minded states. It does not matter that it cannot be interested i.e., self-consciously desire that those later interests be fulfilled. Nor can the merely conscious late fetus or infant. They lack self-conscious concern for their future. Yet their surviving into the future is in their interest, just as broccoli is in the interest of youngsters who are not interested in it. It would be good for them to realize that future. So it still makes sense to speak of a telos due to an innate development pattern (or design) and an interest in that telos of mindless human animals.

This interest in the welfare of the human animal is lacking in the sandal stages as well as the stages of the gametes. This claim is less evident in the latter. Admittedly, if any gamete has a function or interest, it would be that it gives rise to an organism – though it does not matter which one. But neither an individual sperm or egg, nor the scattered pair of gametes whose chromosomes have yet to fuse, possess a genetic nature that determines the particular capacities whose actualization can make a life good. Contrast that with the living stages of the perduring human animal. We can describe their nature as being such that they immanently produce the goods of later stages. They are parts of the same life, governed by the same genetic constitution. So we can speak of the stages of the animal having interests in the later stages of the animal in a way that we cannot speak of the temporal parts of the gametes serving the interests of the later temporal parts of *the* animal, even though gametes and the human animal are all proper parts of larger worms. Thus I agree with Jim Stone, though he's assuming a Three-Dimensional metaphysics, when he writes:

What the fetus *is* finally, is something that makes *itself* self-aware: that good is the fetus's good – this is its nature. Anything benefits from the good which it is its nature to make for itself. I submit that we have a prima facie duty to all creatures not to deprive them of the conscious goods which it is their nature to realize (Stone, 1987, p. 821).⁹

¹ Advocates include Bertrand Russell, David Lewis, Alfred Whitehead, W.V.O. Quine, Robert Nozick, David Armstrong, W. D. Broad, Rudolph Carnap, Nelson Goodman, J.J.C. Smart, Mark Heller, Michael Jubien, Yuri Balashov, Robin LePoidevin, Ted Sider and Katherine Hawley.

² 'Part' in the definition isn't to be construed (circularly) as *temporal* part. The Four-Dimensionalist can construe 'part'

as primitive, accept the existence of temporal and spatial parts, and even allow the same object to be both (Sider, 2008, p. 243).

³ See Jason Eberl (2005) for an account of active potential and its Aristotelian-Thomistic roots.

⁴ I'm assuming that animals go out of existence when life processes irreversibly cease rather than continue to persist as corpses. My reasoning has to do with the different ways that living organisms and corpses initially *assimilate* and then *retain* their parts (Hershenov, 2009). If the living and the dead were temporal parts of the same animal, it would possess (very non-natural kind-like) disjunctive mereological and persistence conditions.

⁵ Hudson admits that appealing to natural kinds is the best option for his rivals who want to identify persons and animals (2007, p. 233). But he thinks the notion of natural kind is "too obscure" to be effective.

⁶ Hudson operates on the methodological assumption that the reliance upon personal identity thought experiments ends in a stalemate (2007, p. 217). I find there to be more truth in his later "acknowledge(ment) that my dismissal of the fanciful thought experiment defense may have been uncharitable and over-hasty..." (2007, p. 233).

⁷ I'm not claiming these sectors produce thought, thus providing a piece of the answer to Hudson's question about which parts of the animal compose the person. I only mean that nerves send painful and pleasant signals to those parts of the brain which if knocked out would desensitize the person.

⁸ What I mean by *identity mattering* is that we must be identical to the future subject of our psychology if there's not to be some drop in prudence-like concern for that thinker. I'm not committed by this thesis to our caring prudentially about our later irreversibly comatose organism. But this thesis and my belief that the human person is identical to the human animal does commit me to caring about my animal's future psychology even if that thinking animal has a new brain and hence no psychological continuity to me now.

⁹ Thanks to Jason Eberl, Rose Hershenov, Stephen Napier, Adam Taylor, and an anonymous referee.

References

- Brueckner, Anthony and Buford, Chris. 2008. The Psychological Approach to Personal Identity: Non-branching and the Individuation of Person Stages. *Dialogue*. 47: 377-386.
- Damasio, Antonio. 1994. *Descartes' Error: Emotion, Reason and the Human Brain*. New York: Penguin Books.
- Eberl, Jason. 2005. Aquinas's Account of Human Embryogenesis and Recent Interpretations. *Journal of Medicine and Philosophy*. 30:4 379-94.
- Hershenov, David. Forthcoming. "Four-Dimensional Animalism," In *Essays on Animalism*. Eds. Paul Snowdon and Steve Blatti. Oxford: Oxford University Press.
- Hershenov, David. 2009. "Organisms and their Bodies." *Mind* 118:70, 803-809.
- Hudson, Hud. 1999. Temporal Parts and Moral Personhood. *Philosophical Studies* 93: 299-316.
- Hudson, Hud. 2001. *A Materialist Metaphysics of the Human Person*. Ithaca: Cornell University Press.
- Hudson, Hud. 2007. I am not an Animal! In *Persons: Human and Divine*, eds. Dean Zimmerman and Peter van Inwagen, 216-234. Oxford: Oxford University Press.
- Locke, John. 1975. *An Essay Concerning Human Understanding*. Ed. Peter Nidditch. Oxford: Oxford University Press.
- McMahan, Jeff. 2002. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press.
- Marquis, Don. 1989. Why Abortion is Immoral. *Journal of Philosophy*. 86:4 183-203.
- Marquis, Don. 2004. Korcz's Objections to the Future-of-Value Argument. *Journal of Social Philosophy* 35: 56-65.
- Olson, Eric. 1997. *The Human Animal: Identity without Psychology*. Oxford: Oxford University Press.
- Olson, Eric. 2008. *What are We? A Study in Personal Ontology*. Oxford: Oxford University Press.
- Parfit, Derek. 1983. *Reasons and Persons*. Oxford: Oxford University Press.
- Perry, John. 1975. In *Personal Identity*, ed. John Perry. Berkeley: University of California Press.

-
- Sider, Ted. 2008. "Temporal Parts." *Contemporary Debates in Metaphysics*. Eds Theodore Sider, John Hawthorne and Dean Zimmerman. London: Blackwell Publishing. 241-262.
- Stone, Jim. 1987. Why Potentiality Matters. *Canadian Journal of Philosophy*. 17:4 816-31.
- George, Robert P. and Tollefsen, Christopher. 2008, *Embryo: A Defense of Human Life*. New York: Doubleday Press.
- Unger, Peter. 2000. "The Survival of the Sentient." *Philosophical Perspectives*. Ed. James Tomberlin. 14: 328-345.
- Van Inwagen, Peter. 1990. *Material Beings*. Ithaca. Cornell University Press.