

©2015

Timothy Campbell

ALL RIGHTS RESERVED

# PERSONAL ONTOLOGY AND BIOETHICS

By

TIMOTHY CAMPBELL

A dissertation submitted to the  
Graduate School—New Brunswick  
Rutgers, The State University of New Jersey  
in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

Graduate Program in Philosophy

Written under the direction of

Jeff McMahan

and

Larry Temkin

and approved by

---

---

---

---

New Brunswick, New Jersey

October, 2015

# ABSTRACT OF THE DISSERTATION

## Personal Ontology and Bioethics

By TIMOTHY CAMPBELL

Dissertation Directors:

Jeff McMahan and Larry Temkin

This dissertation considers how we should treat human beings whose moral status is controversial, such as human embryos, fetuses, and the irreversibly comatose. My approach to this question sidesteps the traditional “personhood” debate, and instead explores the issue of moral status through the lens of what Judith Jarvis Thomson calls personal ontology—the branch of metaphysics that asks what kind of being we are and what the conditions of our identity over time are. Instead of asking what properties something must exemplify in order to count as a person, personal ontology is concerned

with the kinds of things that might exemplify those properties, such as immaterial souls, human animals, embodied minds, bundles of perceptions, etc. One of my central claims is that we are not human animals—beings that persist through time in virtue of the continuity of purely animal functions such as respiration, circulation, and digestion. Instead, I argue, we are conscious beings distinct from human animals, and our continued existence, under normal conditions, depends on the continuity of our neural capacities for consciousness. I argue that this view supports several important conclusions, including the conclusion that killing a human animal and/or using it as a biological resource is morally justified if this is the only reasonable way to achieve a certain significant good—the fulfillment of a sufficiently weighty self-interest that we have--and it does not violate anyone else's rights or produce any evil that nullifies or outweighs this good.

# Acknowledgements

Chapter 2 of this dissertation is based largely on my contribution to “Animalism and the Varieties of Conjoined Twinning,” co-authored with Jeff McMahan. The first published version of this paper, cited in Chapters 2 and 3 of this dissertation, appeared in the journal *Theoretical Medicine and Bioethics*. (For complete reference, see the bibliography.) An expanded version with the same title will appear in *Essays on Animalism*, an anthology published by Oxford University Press and edited by Paul Snowdon and Stephen Blatti.

I received helpful feedback from many people while writing this dissertation. I have benefited from conversations with Nick Beckstead, David Black, Simon Goldstein, Beth Henzel, David Hershenov, Hud Hudson, Mark Lee, Eric Olson, Melinda Roberts, Andrew Sepielli, Peter van Inwagen, and Tobias Wilsch.

For extremely helpful written feedback, I am grateful to Ruth Chang, Theron Pummer, Jacob Ross, David Shoemaker, and Dean Zimmerman.

Derek Parfit read two separate drafts of the dissertation and commented extensively on each. I have benefited tremendously from our exchanges, both in correspondence and in conversation.

I am grateful to my advisors, Jeff McMahan and Larry Temkin. I have learned an enormous amount from them. Since entering Rutgers, I have come a long way as an analytic thinker and writer due, in no small part, to Jeff’s mentorship. I am grateful to Larry for his constant encouragement, for pushing me when I needed to be pushed, and for being such an amazing teacher and friend.

I am grateful to my mother, Ruth Campbell, for her emotional and financial support, and for the hours she spent helping me with the formatting, grammar and syntax,

and diagrams in this dissertation. I am honored to be the second person in our family, after her, to receive the title of doctor. I am proud to be her son.

Finally, I am grateful to my wife, Dan, for the past seven years that we have spent together, and for never giving up on me. Without her love, this dissertation would not exist.

# Contents

Abstract of the Dissertation.....	ii
Acknowledgements.....	iv
<b>1. Introduction and Overview</b> .....	<b>1</b>
1.1 The Moral Issue.....	1
1.2 Personal Ontology.....	4
1.2.1 Two Examples: Animalism and the Psychological Approach.....	5
1.2.1.1 Animalism.....	5
1.2.1.2 The Psychological Approach.....	6
1.2.2 The Psychological Approach and the Relation between Psychological Continuers and Human Animals.....	7
1.2.2.1 Constitutionalism.....	11
1.2.2.2 The Embodied Parts View.....	12
1.2.2.3 Emergent Dualism.....	13
1.2.2.4 The Hybrid Approach.....	13
1.3 ‘We’, ‘People’, and ‘Human’.....	14
1.4 Self-Interest.....	16
1.5 The Personal Ontology Argument.....	19
1.6 How My Approach Differs from Others.....	21
1.6.1 Differentiating Questions of Personal Ontology from Questions of Personhood.....	21
1.6.2 The Moral Relevance of Personal Ontology.....	25

1.6.3	Neutrality and Theoretical Cost-Efficiency.....	27
1.6.4	Moral Uncertainty.....	29
1.7	A Brief Overview of the Remaining Chapters.....	30
<b>2.</b>	<b>Can Animalism Counter the Intuitive Appeal of the Psychological Approach?</b>	<b>31</b>
2.1	Introduction to Chapter 2.....	31
2.2	The Transplant Argument.....	32
2.2.1	Responding to Objections to the Transplant Argument.....	33
2.2.1.1	Are Psychological Traits Linked to Particular Bodies?.....	33
2.2.1.2	Can there be Continuous Consciousness between Different Human Animals? .....	36
2.2.1.3	Are Judgments about Identity corrupted by Judgments about “What Matters” in Survival? .....	41
2.2.1.4	Can Human Animals Think Remotely?.....	45
2.3	Another Problem for Animalism: The Severed Head Case.....	48
2.3.1	Appealing to the Importance of the Brainstem.....	48
2.3.2	Appealing to the Importance of Self-Maintenance.....	50
2.3.3	Animalism and the Severed Head Intuition: An Awkward Fit.....	51
2.3.4	Can there be Consciousness without a Subject? .....	52
2.4	The Dicephalus Argument.....	54
2.4.1	A Famous Case of Dicephalus: Abigail and Brittany Hensel.....	54
2.4.2	McMahan’s Dicephalus Argument.....	55
2.4.3	Blatti and Duplication Objections.....	57
2.4.4	Are there Two Human Animals in Dicephalus? .....	64



2.5	Cephalopagus.....	66
2.5.1	Considering Cephalopagus and Dicephalus Together.....	67
2.5.2	Cephalopagus and the Too Many Subjects Problem.....	70
2.6	Conclusion and Summary of Chapter 2.....	70
<b>3.</b>	<b>Do the Virtues of Animalism Outweigh those of the Psychological Approach?</b>	<b>73</b>
3.1	Introduction to Chapter 3.....	73
3.2	The Four Major Considerations that Allegedly Favor Animalism over the Psychological Approach.....	74
3.2.1	The Too Many Subjects Problem.....	74
3.2.2	Simplicity.....	76
3.2.3	Consistency with Common Sense.....	77
3.2.4	Composition and Unity.....	77
3.3	The Psychological Theorist's Response to the Too Many Subjects Problem and Why Animalism is not Immune to that Problem.....	78
3.3.1	Animalism and the Corpse Problem.....	78
3.3.2	The Derivatively Conscious Animal View.....	80
3.3.2.1	The Derivatively Conscious Animal View and Constitutionalism.....	80
3.3.2.2	The Derivatively Conscious Animal View and the Embodied Parts View.....	82
3.4	The First Response to the Too Many Subjects Problem on Behalf of the Biological Approach: Eliminativism.....	83
3.4.1	Eliminativism and the Special Composition Question.....	86

3.4.2	Eliminativism and The Psychological Approach: How the Biological Approach Loses the Advantage of Consistency with Common Sense.....	87
3.4.3	Can The Biological Approach Offer a More Plausible View of Composition and Unity? .....	89
3.4.4	If the Biological Approach and The Psychological Approach are Combined with Eliminativism, the Former Loses its Advantage.....	94
3.5	The Second Response on Behalf of the Biological Approach: Switcheroo.....	96
3.5.1	Combining Animalism and the Psychological Approach with Switcheroo.....	97
3.5.2	Why Defenders of the Biological Approach Cannot Appeal to the Impressiveness of Biological Death to Defend Switcheroo.....	101
3.6	The Bodily Approach and the Psychological Approach.....	105
3.6.1	Why the Bodily Approach Is No Better Than the Biological Approach.....	107
3.7	Why Animalism is no better than the Psychological Approach with respect to the Four Major Considerations.....	111
3.8	Against the Hybrid Approach.....	114
3.8.1	A Powerful Objection to the Hybrid Approach.....	115
3.8.2	Against Patrick Lee's Animalism.....	118
3.9	Cephalopagus Revisited.....	123
3.10	Conclusion and Summary of Chapter 3.....	126
<b>4.</b>	<b>The Psychological Approach and the Moral Status of Human Animals</b>	<b>135</b>
4.1	Introduction to Chapter 4.....	135

4.2	When Do We Begin? .....	137
4.3	The Moral Relevance of the Psychological Approach.....	140
4.3.1	Kind-based Arguments and Identity-based Arguments.....	142
4.3.2	Universalizability Arguments.....	144
4.3.2.1	Hare's Golden Rule Argument.....	144
4.3.3	Future-Like-Ours Arguments.....	149
4.3.3.1	The Derivative Deprivation Objection.....	150
4.3.3.2	A Response to the Derivative Deprivation Objection: Denying that Animals are Really Conscious.....	152
4.3.3.3	Can a Human Animal be deprived of a Future Like Ours While it is Associated with a Psychological Continuer? .....	157
4.3.4	Potential Person Arguments.....	174
4.4	Establishing the Conclusion of the Personal Ontology Argument.....	176
4.4.1	The Symmetry Argument.....	176
4.4.2	The Consent Objection.....	182
4.4.3	The Afterlife Objection.....	185
4.4.4	Using Human Animals.....	189
4.5	Weighing the Value of Self-Interest against the Value of Biological Life.....	191
4.6	Do Human Animals Have Rights? .....	196
4.7	Conclusion and Summary of Chapter 4.....	196
<b>5.</b>	<b>Do Reductionist Views Threaten the Personal Ontology Argument?</b>	<b>198</b>
5.1	Introduction to Chapter 5.....	198
5.2	The Question of Value-Bearers.....	199

5.3	Fission and Fusion.....	207
5.4	A Conflict Between Person-Centered Views of Aggregation, No Survivors, and Our Special Concern for Persons.....	213
5.5	Cohabitation and the Moral Significance of Temporal Wellbeing.....	219
5.6	Two Possible Solutions to the Reductionist Dilemma.....	228
5.6.1	Does the Reductionist Dilemma Rest on a Mistake in Quantifying Experience? .....	229
5.6.2	What If Both Personal and Temporal Dimensions Have Moral Significance? .....	230
5.7	Should Reductionists Accept the Personal Ontology Argument? .....	232
5.7.1	Does Double-Counting In Cases of Fission and Fusion Undermine the Personal Ontology Argument? .....	232
5.7.2	Double-Counting and the Distinction between Derivatively Conscious and Non-Derivatively Conscious Entities.....	232
5.7.3	Double-Counting and the Number of Conscious Perspectives.....	233
5.7.4	Double-Counting and the Claim that Human Animals are not really Conscious .....	234
5.7.5	Reductionists Who Accept Double-Counting in Cases of Fission and Fusion Can Accept the Personal Ontology Argument.....	235
5.7.6	Would a Radically Impersonal View of Aggregation Threaten the Personal Ontology Argument? .....	236
5.8	Conclusion and Summary of Chapter 5.....	237
<b>6.</b>	<b>Conclusion and Summary</b>	<b>240</b>

6.1	Summary of the Personal Ontology Argument and My Defense of It.....	240
6.2	The Challenge of Reductionist Views.....	242
6.3	Concluding Remarks.....	243
<b>Bibliography.....</b>		<b>244</b>

## List of Diagrams

Diagram 1.....	202
Diagram 2.....	216
Diagram 3.....	221

# 1. Introduction and Overview

## 1.1 The Moral Issue

Consider the following cases:

- (1) Gianna Jessen is the ostensible survivor of a failed saline abortion attempt, which is the cause of her cerebral palsy. Jessen became a pro-life advocate at age fourteen. She has told her unique story to several political bodies and governments. No one can deny Jessen's charisma as a speaker, or that her personal story is both fascinating and heart-warming. When telling her story, Jessen always uses the first-personal pronoun 'I' to refer to the fetus that was born on April 6, 1977—the date of the failed abortion. She and others assume that on this day *she* survived an attempt on her life. Her goal in life is to be a voice for “the voiceless”—for other human fetuses and embryos that are essentially no different from how *she* was the day that she was born.
- (2) When the body of Nancy Cruzan died in 1990, after eight years in a vegetative state, her family engraved on her tombstone: “DEPARTED JAN 11, 1983/ AT PEACE DEC 26, 1990.” The family had gone to the Supreme Court in an attempt to terminate the body's life support. The attempt was successful. In the late 1960s and early 1970s, there were a number of successful efforts to persuade medical bodies and legislatures that brain death marks the death of a human being. These efforts focused on the fact that brain death is sufficient for the irreversible loss of the capacity for consciousness. Thus Alan Shewmon, a pediatric neurologist says that “the introducers of the concept [of

brain death] intended to redefine death in terms of unconsciousness rather than diagnose it as the cessation of the biological life of the human organism.”<sup>1</sup>

- (3) In 1997 scientists created a headless frog by manipulating gene expression so that only certain parts of its body would grow. At the time there was some discussion of creating headless human organisms for the purpose of organ harvesting. This issue continues to be considered in the context of genetically engineering human embryos to create a biological resource for technological research and health development.
  - (4) On March 14, 1970, scientists from Case Western Reserve University School of Medicine led by the neurosurgeon Robert J. White, transplanted the head of a Rhesus monkey onto the body of another from which the head had already been removed. The resulting being, though paralyzed, showed unmistakable signs of consciousness and intentionality. It was able to smell, taste, hear, and see; it was able to visually track objects through space. The animal survived for some time after the operation, at times attempting to bite some of the staff.
- White, a Catholic, believed the brain to be the repository of the human soul and that the operation he had developed could be used to save paraplegics in danger of dying from organ failure. “People are dying today who, if they had body transplants, in the spinal injury community would remain alive,” White told an interviewer from the BBC.

---

<sup>1</sup> Alan D. Shewmon, “Brainstem death, Brain Death, and Death: A Critical Reevaluation of the Purported Equivalence.” *Issues in Law and Medicine* 14 (1998), 125-45. For original reference, see Jeff McMahan, *The Ethics of Killing* (New York: Oxford University Press), 401.

Not everyone was convinced, however. Stephen Rose, director of brain and behavioral research at the Open University, told the BBC that White's experiment was "medical technology run completely mad..." and that the question of who someone who had undergone a head transplant would "be" is extremely complicated. "Your person is largely embodied but not entirely in your brain," said Rose.

These and other similar cases highlight an important theoretical issue in *bioethics*—the branch of applied ethics that addresses normative questions in fields such as medicine and biotechnology. Each case raises a question familiar to bioethicists: How ought we to treat human beings whose moral status is uncertain, where such beings include human embryos, fetuses, and those in a permanent vegetative state? These cases also raise questions about identity. Did Gianna Jessen exist on the date of her mother's failed abortion? There is no doubt that if the fetus born on that date had died, Jessen wouldn't be here today. But is Jessen numerically identical to that fetus? Are the former and latter one and the same? What about brainless "spare parts" clones? Could you have come into existence as such a clone if the technology and laws had permitted "spare parts" cloning around the time of your birth?

Intuitively, questions such as these—questions about our identity—are relevant to questions about moral status. If Jessen was never a premature fetus, if she didn't actually survive a brush with death in a Los Angeles abortion clinic, then the force of her pro-life message seems somewhat diminished, even if her life story is no less captivating. When Jessen says "I was almost killed by a saline injection" we take her to be reporting a strict and literal truth; we do not think that she is speaking loosely or metaphorically. But if we



thought that what she is reporting is, strictly speaking, false, then our attitudes about the ethics of abortion might change drastically. We may be much less confident about the moral objectionableness of abortion; our feelings about it may become more unsettled. Similarly, if we think that we could continue to exist in a persistent vegetative state, or that any of us could have existed as a brainless clone, then our attitudes about “spare parts cloning” and certain kinds of euthanasia may be very different from what they would be if we were to reject these beliefs. If no brainless clone is or could be one of us (and vice versa), it is harder to see why the creation and use of such clones is morally objectionable.

## 1.2 Personal Ontology

All else equal, then, one’s view about identity over time and across possible worlds can inform one’s view of moral status. In this dissertation, I take this assumption as a starting point for a potentially fruitful moral and metaphysical inquiry. I explore the issue of moral status through the lens of what Judith Jarvis Thomson calls *Personal Ontology*.<sup>2</sup> This is the branch of metaphysics concerned with the question what *are* we most fundamentally, and what are our *persistence conditions*—what does it take for us to persist through time as one and the same being? I call this *the question of personal ontology*. This opening chapter has two aims—first, to differentiate the question of personal ontology from more familiar questions concerning personhood and to suggest why the former question is

---

<sup>2</sup> Judith Jarvis Thomson, “People and Their Bodies.” In *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne and Dean W. Zimmerman (Malden, MA: Blackwell Publishing, 2008), 155-76.

potentially morally important, and second, to lay out the central argument of this dissertation, which relies on an answer to this question.

## 1.2.1 Two Examples: Animalism and the Psychological Approach

### 1.2.1.1 Animalism

Following Eric Olson<sup>3</sup> I am inclined to think that the best way to understand the question of personal ontology is to see what would count as an answer to it. Here, I consider two popular answers—two views of personal ontology—that will concern us throughout the dissertation.

The first of these views is *Animalism*. This is the view that we are most fundamentally *animals*. Most who defend the view that we are animals—that is, most *Animalists*—assume that all animals, including human animals, have the same *basic nature* as other biological organisms. Although some animals have properties of great *practical and moral* importance—consciousness and rationality, for example—that less impressive organisms such as plants and fungi lack, all organisms have the same basic metaphysical nature.

Animalists disagree about what the fundamental properties and persistence conditions of animals are. Some hold that an animal has among its fundamental properties a kind of functional unity of its various parts that is lost when it dies, and hence, that the

---

<sup>3</sup> Eric Olson, *What Are We?* (New York: Oxford University Press, 2007), 6.

animal ceases to exist at death or when it has irreversibly lost the capacity for life.<sup>4</sup> According to these Animalists, what is necessary and sufficient for our identity over time is a kind of *biological continuity* that begins, perhaps, at conception and ends at biological death.<sup>5</sup> The relevant kind of continuity is an ongoing, self-organizing biological event—a *life*—which, according to these animalists, is necessary for our existence. When a biological event of this sort ends, a human animal, and hence, one of us, ceases to exist.

Other Animalists maintain that an animal can continue to exist after its death as a corpse. For these Animalists, each of us is identical to an animal that is alive only contingently.<sup>6</sup> Some who defend this version of Animalism use the term ‘body’ rather than ‘animal’ when explaining what kind of being they believe we are. For example, Fred Feldman writes:

I think I am my body. I think I formerly was a fetus. I think someday that I will be dead - just a corpse. When I refer to myself - I mean to be referring to this human body - the one that is writing this essay.<sup>7</sup>

We can distinguish two types of Animalism. One, *The Biological Approach*, states that we are animals that are alive *essentially*. We are most fundamentally living beings—those that persist through time by virtue of biological continuity. On this approach, if individual X

---

<sup>4</sup> Peter van Inwagen, *Material Beings* (Ithaca, NY: Cornell University Press, 1990), 146; Eric Olson, *The Human Animal: Personal Identity Without Psychology* (Oxford: Oxford University Press, 2007b), 111.

<sup>5</sup> I say “perhaps” because some writers have argued that the kind of biological continuity that constitutes our identity over time doesn’t begin until the cells of the embryo lose their totipotentiality, taking on highly specialized roles within the animal. This happens during the gastrulation phase of embryonic development, 14 to 16 days after fertilization. For example, see Barry Smith and Berit Brogaard, “Sixteen Days,” *Journal of Medicine and Philosophy* 28 (2003), 45-78. Others reject this view of the origins of the human organism, and defend the standard view that the organism originates at biological conception. For example, see S. Matthew Liao, “The Organism View Defended,” *The Monist* 89 (2006), 334.

<sup>6</sup> Fred Feldman, *Confrontations with the Reaper* (New York: Oxford University Press, 1992); Fred Feldman, “Basic Intrinsic Value,” *Philosophical Studies* 99 (2000), 319; W.R. Carter, “Will I Be a Dead Person?” *Philosophy and Phenomenological Research* 59 no.1 (1999), 167 – 71.

<sup>7</sup> Fred Feldman, “The Termination Thesis,” *Midwest Studies in Philosophy* 24 no.1 (2000), 98–115.

at time  $t$  is one of us, and individual  $Y$  at time  $t'$  is also one of us, then  $X$ -at- $t$  is numerically identical to  $Y$ -at- $t'$  if and only if  $X$ -at- $t$  and  $Y$ -at- $t'$  are appropriately *biologically continuous*. The other version of Animalism, *The Bodily Approach*, states that we are animals—or *bodies*—that are alive *only contingently*. We persist through time in virtue of *bodily continuity*. On this approach, if individual  $X$  at time  $t$  is one of us, and individual  $Y$  at time  $t'$  is also one of us, then  $X$ -at- $t$  is numerically identical to  $Y$ -at- $t'$  if and only if  $X$ -at- $t$  and  $Y$ -at- $t'$  have *the same body*. Part of the appeal of Animalism is that it seems to reflect “biological common sense.” Each of us *seems* to be identical with an animal, something that breathes the air we breathe, digests the food we eat, etc. When biologists tell us about the anatomy of animals, they seem to be telling us about *our* anatomy. Evolutionary biologists who say that human animals and chimpanzees evolved from a common ancestor would probably say that *we* and chimpanzees evolved from a common ancestor. This sort of talk suggests that *we* are animals.

### 1.2.1.2 The Psychological Approach

The second view that will concern us, *The Psychological Approach* (PA), states that our essential properties include *psychological properties* of some kind, and that our persistence conditions include the continuous instantiation of such properties. The PA is overwhelmingly popular. It has been the dominant view in the literature on personal identity theory for many years.<sup>8</sup> To see why many find the PA so attractive, consider the gruesome experiments involving head transplantation.<sup>9</sup> Now imagine that a similar kind

---

<sup>8</sup> Derek Parfit, *Reasons and Persons* (Oxford: Oxford University Press, 1984).

<sup>9</sup> Robert White, et al. “Cephalic Exchange Transplantation in the Monkey” *Surgery* Jul: 70 no. 1(1971), 135 – 39.

of experiment were to be done on *you*. Suppose that some (highly unethical) mad scientists were to detach your brain from the rest of your body and transplant it into the head of a different body from which the brain had already been removed, and that somehow they managed to connect that brain to the nervous system of the new body, so that the resulting person—the person with your old brain—had full control over the movements of this body. Suppose also that the mad scientists would be careful not to damage any part of your brain during the transplant, and that, therefore, the person with the new body and your old brain would think, talk, and act just like you, and would seem to have all of your memories.

When presented with such stories, many have what Olson calls *the Transplant Intuition*; they judge that if they were to undergo such an operation, *they* would survive as the person who inherits their brain and psychology.<sup>10</sup> This is what The Psychological Approach implies—*you* would be the person with the new body and your old brain. For this person is uniquely *psychologically continuous* with you; *its* psychological states are related to *yours* in the right way. (I will say more about what this means in a moment.)

Animalists reject the transplant intuition. On their view, if you were to undergo a brain transplant, you would be identical to the animal from which the brain was taken. If that animal's consciousness-generating parts were removed, it wouldn't go anywhere. It would simply lose its capacity for consciousness. It would be similar to someone in a permanent vegetative state. Animals cannot switch bodies. This should seem clear, especially if, as some animalists say, *we are* bodies. If *you* are an animal, and no animal would switch bodies if its brain were transplanted, then you would not switch bodies if

---

<sup>10</sup> Eric T. Olson. *The Human Animal: Personal Identity Without Psychology* (New York, Oxford University Press, 1997) Chapter 1.

your brain were transplanted. Instead, you would remain behind on the operating table, empty-headed.<sup>11</sup> Because Animalism implies this, it is incompatible with the Psychological Approach.

An important clarification is needed here. By ‘psychological continuity’ I mean any kind of continuity that is necessary for the continuous instantiation of some sort of mental property or the capacity to exemplify that property. Thus, the meanings of terms such as ‘psychology’ and ‘psychological’, are much broader than one might have thought. For example, the kind of psychological continuity that is necessary for our identity over time might be *minimal*—it might include only the continuity of those areas of the functional brain that are required for one to be *just barely sentient*. ‘Psychological continuity’ could be *any sort of mental continuity*. This is important because it implies that the Psychological Approach as I understand it encompasses a *very broad* range of views. It includes even some views on which we are immaterial beings, such as souls. Any such view according to which the immaterial beings we are persist through time only if they continue to exemplify some kind of mental property (or continue to have the capacity to exemplify that kind of property), is a version of the Psychological Approach.

### **1.2.2 The Psychological Approach and the Relation Between Psychological Continuers and Human Animals**

---

<sup>11</sup> Ibid. Chapter 3, Section 3.8.2, discusses a version of Animalism, defended by Patrick Lee, that accommodates the transplant intuition. On Lee’s view, you are essentially a being that has the capacity for rational thought and action. This view allows that in my hypothetical case, Transplant, the animal (or animal) that you are survives as the person who inherits your cerebrum, assuming as I am here that the parts of the animal which are most directly responsible for its capacity for rational thought and action are mainly parts of its cerebrum. Lee’s version of animalism is interesting and unorthodox.

As we saw, the Psychological Approach implies that I would be separated from the animal that is now sitting in my chair if my brain and psychology were transferred to a different animal. Since no object can be separated from itself, the Psychological Approach seems to imply that me and the animal that is now sitting in my chair are *different objects*. An often-cited analogy for the disagreement between animalists and defenders of the Psychological Approach concerning the relation between me and the animal is the disagreement between metaphysicians regarding the relation between an artifact and the lump of matter of which it is made.<sup>12</sup>

Suppose an artist molds a lump of clay, *Lump*, into a statue resembling Socrates, *Socco*. One view is that Lump and Socco are identical. When the artist molds Lump in the likeness of Socrates, no new thing comes into existence; rather, a certain thing that *already exists*, Lump, comes to look like Socrates. The names ‘Socco’ and ‘Lump’ refer to the same object. Similarly, if the artist were to squash Socco, nothing would cease to exist. Rather, Lump would lose the form of Socrates and would take on a different form. On this view, Socco and Lump are *numerically identical*—they are one and the same object. Similarly, according to Animalism, me and the animal that is now sitting in my chair are numerically identical. Because the animal was once a fetus, I was once a fetus. If a fetal animal does not start out as a *person* it can nevertheless be “worked up” into a person by its mother and genetic programming, the way lumps of clay can be worked up into statues by artists.

---

<sup>12</sup> Eric T. Olson, “Material Coincidence and the Indiscernibility Problem.” *Philosophical Quarterly* 51 no.204 (2001) 337-355.

Borrowing an example of Parfit's, we could remove 99% of the clay from Socco's interior, collecting it in a single portion, while leaving Socco's exterior—the remaining 1% of the clay—unaffected.<sup>13</sup> *Lump* would continue to exist as the thing that is made up of the extracted clay.

Many philosophers reject this account of the relation between Lump and Socco for the same reason that many reject the Animalist's account of the relation between me and the animal sitting in my chair, namely Lump and Socco seem to have different persistence conditions. Since a lump of clay can persist through the loss of 1% of its constituent clay. But if Lump was then destroyed (burned up in the artist's kiln, for instance) Socco would continue to exist, but would be hollow. Physical separability of this kind seems to imply that Socco and Lump are different objects. But if Socco and Lump are not identical, how are they related? Similarly, and more important for our purposes, if I am not the animal that is now sitting in my chair, then how I am related to it?

### 1.2.2.1 Constitutionalism

*Constitutionalism* provides a unified answer to both of these questions—the question about the statue and the lump, and our central question about me and the animal that is now sitting in my chair. In each case, one object (a lump or an animal) constitutes another object (a statue or a psychological continuer). Thus, Constitutionalism implies that the animal sitting in my chair now *constitutes* me. Most versions of Constitutionalism imply that

---

<sup>13</sup> Derek Parfit, "Persons, Bodies, and Human Beings" In *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne and Dean W. Zimmerman. (Oxford: Blackwell Publishing, 2008), 18.



while the animal constitutes me it is spatially and materially *coincident* with me—in other words, the animal and I occupy the same space and are made of the same matter. (Henceforth, I will use the terms ‘coincident’ and ‘coincidence’ in place of the more cumbersome ‘spatially and materially coincident’ and ‘spatial and material coincidence’.) However, most Constitutionists also say that constitution implies more than mere coincidence. For example, they claim that constitution, unlike mere coincidence, is asymmetric and irreflexive (things can’t constitute each other and nothing can constitute itself).<sup>14</sup> (I will discuss Constitutionalism in greater detail in Chapters 3 and 4.)

### 1.2.2.2 The Embodied Parts View

Some defenders of the Psychological Approach claim that we are neither animals nor material things constituted by animals, but are rather *proper parts* of animals. (X is a proper part of Y if and only if X is a part of Y and X is not identical to Y. Henceforth, I shall use the term ‘part’ in place of ‘proper part’.) On what Parfit calls *the Embodied Parts View* (EPV),<sup>15</sup> we are spatial parts of animals, such as brains, cerebrums, or whole nervous systems. This view has a growing number of proponents.<sup>16</sup> One attractive feature of EPV is that it supports the conviction that I can survive the loss of most of my bodily parts (fingers, toes, etc.) but not my *consciousness-generating* part—that which is most directly

---

<sup>14</sup> Lynne Rudder Baker, *Persons and Bodies: A Constitution View* (Cambridge, M.A.: Cambridge University Press, 2000).

<sup>15</sup> Derek Parfit, “We Are Not Human Beings.” In “Persons, Bodies, and Human Beings,” *Philosophy* 87 no.1 (2012), 5-28.

<sup>16</sup> Jeff McMahan, *The Ethics of Killing: Problems at the Margins of Life* (Oxford: Oxford University Press, 2002); Ingmar Persson, “Our Identity and the Separability of Persons and Organisms” *Dialogue* 38 no.03 (1999) 519; Hud Hudson, *A Materialist Metaphysics of the Human Person* (Ithaca, N.Y.: Cornell University Press, 2001).

responsible for my consciousness. If I cannot persist after losing my consciousness-generating part, this gives me some reason to think that I am that part. Another potentially attractive feature of EPV is that it avoids Constitutionalism's implication that we are coincident with animals. We will see later (Chapter 3) that coincident entities are highly problematic. Finally, EPV seems to provide the most plausible interpretation of certain cases of conjoined twins. (I will discuss these cases at length in Chapter 2).

### 1.2.2.3 Emergent Dualism

Cartesian Dualism is not very popular these days. However, in contemporary philosophy of mind, certain versions of dualism have gained in popularity due to work on the metaphysics of consciousness by Thomas Nagel<sup>17</sup> and, more recently, David Chalmers.<sup>18</sup> The apparent hopelessness of explaining the subjective element of consciousness in terms of the physical has convinced some philosophers to accept a “naturalistic” dualism. For example, some contemporary writers defend *Emergentism*, according to which we are “emergent” minds—non-physical entities dependent upon our brains for conscious experience and thought. William Hasker, who defends a view of this kind, describes the relation between the emergent mind and the brain as follows:

As a consequence of certain configuration and function of the brain ... a new entity comes into being – namely, the mind or soul. This new thing is not merely a “configurational state” of the cells of the brain ... The mind, on this view, is “a thing in itself”; it is what some philosophers call a “substance.” It is not made of the chemical

---

<sup>17</sup> Thomas Nagel, “What is it Like to be a Bat?” *The Philosophical Review* 83 (1974 ) 435-50.

<sup>18</sup> David J. Chalmers, “The Problems of Consciousness.” In *Consciousness: At the Frontiers of Neuroscience*, edited by H. Jasper, L. Descarriers, V. Castellucci and S. Rossignol, 29 – 59, (New York: Lippincott-Raven, 1998).

stuff of which the brain is composed, though it crucially depends on that chemical stuff both for its origin and its continuance.<sup>19</sup>

According to Hasker, the brain generates the mind the way a magnet generates a magnetic field.

#### 1.2.2.4 The Hybrid Approach

There remains The Hybrid Approach, according to which we are animals with persistence conditions that are partly biological and partly psychological. On the Hybrid Approach, the kind of continuity that underpins the identity over time of animals according to the Biological (or Bodily) Approach—some kind of biological (or bodily) continuity—is sufficient for identity in the absence of psychological continuity, and psychological continuity (of an appropriate kind) is sufficient for identity in the absence of biological (or bodily) continuity. But when the two kinds of continuity diverge, identity follows psychological continuity. Proponents of this approach claim that each of us is identical to an animal, but they also accept the Transplant Intuition, as well as the claim that psychological continuity isn't necessary for the persistence of human animals. The Hybrid Approach has been defended by David Wiggins and John McDowell, among others.<sup>20</sup>

The obvious attraction of the Hybrid Approach is that it seems to capture more of what we want to say about ourselves than either The Biological/Bodily Approach or The Psychological Approach. It entails that each of us was once a fetus and that he could

---

<sup>19</sup> William Hasker, "On Behalf of Emergent Dualism," in *In Search of the Soul: Four Views of the Mind-Body Problem*, edited by Joel B. Green and Stuart L. Palmer, 78 (Downers Grove, IO: InterVarsity Press, 2005).

<sup>20</sup> David Wiggins, *Essays for David Wiggins, Identity, Truth and Value*, edited by D. Wiggins and S. Lovibond. (Oxford: Blackwell, 1996); McDowell, John, "Reductionism and the First Person," in *Reading Parfit*, edited by Jonathan Dancy (Oxford: Blackwell, 1997), 230-50.

continue to exist in a permanent vegetative state (PVS), but also that he would follow his transplanted brain instead of the brainless being that remains on the operating table—a being that is ostensibly alive in the biological sense.

### 1.3 ‘We’, ‘People’, and ‘Human’

The question of personal ontology is “What kind of thing are we most fundamentally?” I have given some examples of what I mean by ‘what kind of thing?’ What do I mean by ‘we’? An accurate but completely unhelpful answer is ‘beings of *our* fundamental kind.’ In this answer, the term ‘our’ refers to certain beings. But *which* beings? I need to specify the relevant class of beings without simply re-raising my question. I will start with the class *human persons*. Here ‘persons’ denotes beings whose status as persons is not in dispute—and ‘human’ denotes beings that are *genetically* human—those who possess human genes. I assume that all human persons belong to a common fundamental ontological kind. Beings of *that* fundamental kind, whatever it is, are what I refer to when I use terms such as ‘we’ and ‘our’. Something is *one of us* if and only if it belongs to this fundamental kind—it is a being *of our kind*. The question “What kind of thing are we most fundamentally?” is the question “What is the common fundamental kind to which *human persons* belong?”

We can approach the question of personal ontology by asking what a human person refers to when it says ‘I’. There is an intuitive albeit naive view of personal reference according to which, with certain rare exceptions, personal pronouns, proper names, and pure indexicals, such as ‘you’, refer invariantly to the same objects. On this view, which Olson and many others accept, I *am* what I refer to when I think and utter ‘I’—just as

Rutgers University *is* what one refers to when one thinks and utters ‘Rutgers University’; and I am what one refers to when one address me as ‘you’ or speaks of me as ‘he’ or ‘Campbell’. We can get a better grip on the question of personal ontology by asking “What do personal pronouns and proper names, as uttered by human persons, refer to?” Call this *the question of personal reference*. As Olson rightly observes, the question of personal ontology and the question of personal reference are not equivalent. Sometimes we use personal pronouns to refer to beings that are not of our kind. For example, we anthropomorphize (a child points to her favorite Transformers action figure and proclaims ‘He is so cool’). We refer elliptically (A man says ‘I am parked up a hill’. This is true because ‘I’ refers to the man’s car. The *man* isn’t parked, nor is he up a hill). Our personal pronouns sometimes refer to action figures and parked cars, but this does not suggest that each of us is most fundamentally an action figure or a parked car. In certain special cases, the question of personal reference and the question of personal ontology yield different answers. But according to Olson, if we set these special cases aside, the two questions amount to the same thing.

I assume that personal pronouns are referring expressions. As Olson observes,<sup>21</sup> some philosophers, such as Elizabeth Anscombe,<sup>22</sup> deny this. On Anscombe’s view, the word ‘I’ in ‘I am walking’ doesn’t even *purport* to refer to anything that is walking—just as the word ‘it’ in ‘it is raining’ doesn’t purport to refer to anything that is raining. If Anscombe is right, then I do not refer to anything when I say ‘I’. But we can still ask what

---

<sup>21</sup> Olson, *What Are We?*

<sup>22</sup> G.E.M. Anscombe, *Ethics, Religion and Politics. The Collected Philosophical Papers of G.E.M. Anscombe*, vol. III, (Oxford: Blackwell, 1981).

‘Campbell’ is, and what the conscious being now sitting in this chair is. Since we can still ask these questions, our inquiry can proceed even if ‘I’ isn’t a referring expression.

## 1.4 Self-Interest

Before presenting my central argument, I must introduce an important concept—the concept of *self-interest*. Following Holtug (2010), I will treat ‘self-interest’ as both a count noun and a mass noun.<sup>23</sup> It is correct to say that I have *a* self-interest, but also that the overall fulfillment of self-interest is positive.

A theory of self-interest should specify the individuals for whom existence may go better or worse, and it should say what makes existence go better or worse for them. A theory of self-interest should not be confused with what Parfit calls the Self-interest Theory, which is a theory about rationality.<sup>24</sup> Theories about self-interest specify what is involved in life going well, but they do not assign aims to persons.

The object of a self-interest is a state of affairs or outcome (I use the terms ‘state of affairs’ and ‘outcome’ interchangeably.) Furthermore, a self-interest can be fulfilled or frustrated. The fulfillment (or frustration) of individual self-interest is a value that can be promoted. To promote the fulfillment of individual self-interest is to cause it to be the case that the total amount of the fulfillment of individual self-interest is greater than it would otherwise be.

If individual S has a self-interest in outcome O, it does not follow that S *takes an interest* in O or *desires* that O obtain. On my conception of ‘self-interest’, that S has a self-

---

<sup>23</sup> Nils Holtug, *Philosophy and Ethics: New Research*, edited by Laura V. Siegal, (Hauppauge, New York: Nova Science Publishers, Inc., 2006), 60-110.

<sup>24</sup> Derek Parfit, *Reasons and Persons*, 4.

interest in O does not require that S *be interested* in O. In fact, it does not require that S be conscious at all, or that S have even the most rudimentary neural capacity for consciousness. Thus, the basic conception of self-interest that I am assuming is compatible with the view that non-conscious fetuses, embryos, and those in a permanent vegetative state (PVS) have self-interests. Indeed, given this very basic conception of self-interest, it is an open question whether even rocks and cars have self-interests. If rocks and cars are among those beings for which existence may go better or worse, it follows from my conception of self-interest that rocks and cars have self-interests. To deny that they have self-interests, on my view, we must deny that existence may go better or worse for them.

Theories of self-interest should be assessed on the basis of how plausible an account they give us of what is good for people. I shall assume that if an individual has a self-interest in something, she has a reason to bring it about, everything else being equal. This means that we may in part assess theories about self-interest in terms of the reasons with which they provide us.

*What are* the beings for whom existence might go better or worse? What are *the bearers* of self-interest? Naturally, one might assume, they are all and only beings of *our kind*. The conscious being that has typed this sentence is one, the conscious being now reading it is another. The bearers of self-interest are the referents of our personal pronouns. The bearers of self-interest that will concern us are just those beings of our kind. (This claim is problematic. Later in the dissertation (Chapters 4 and 5) I will consider some of the difficulties it faces.)

In addition to asking what the bearers of self-interest are, we can ask how the fulfillment and frustration of each bearer's self-interest is determined. There are many different views. For Chapters 1 – 4, I will assume

*The Identity View of Self-Interest:*

- (1) Individual S1 has a self-interest at time t1 in individual S2 benefiting at time(s) t2(-t3), if and only if (a) S1 exists at t1, and (b) S1 and S2 are numerically identical.
- (2) The strength of the self-interest depends only on the size of the benefit(s).

A simpler way of stating *the Identity View* is to say that identity is “what matters” in survival. I adopt *the Identity View* only as a starting point for my inquiry. There are powerful arguments against this view; I discuss these arguments in Chapters 4 and 5.

### **1.5 The Personal Ontology Argument**

The central argument of this dissertation establishes a connection between personal ontology and the ethics of killing and using human animals as a biological resource. I call it *the Personal Ontology Argument*:

Premise 1: We are psychological continuers.

Premise 2: If we are psychological continuers, then killing a human animal and/or using it as a biological resource is morally justified if this is necessary (or the only reasonable way) to achieve a certain good—the fulfillment of a weighty self-interest of some psychological continuer—and apart from its effects on any



human animal it does not violate anyone's rights or produce any evil that nullifies or outweighs this good.

Therefore,

Conclusion: Killing a human animal and/or using it as a biological resource is morally justified if this is necessary (or the only reasonable way) to achieve a certain good—the fulfillment of a weighty self-interest of some psychological continuer—and apart from its effects on any human animal it does not violate anyone's rights or produce any evil that nullifies or outweighs this good.

Three clarifications are in order. The first concerns moral justification. To say that an act, performed by an agent, is morally justified is to say both (a) that it is morally permissible, and (b) that there is sufficient moral reason to perform it rather than some other relevant alternative act. The relevant alternative acts are those an agent can perform in the context in which the act in question is being considered. (I leave the relevant notion of 'can' unspecified.) Moral *justification* does not imply moral *obligation*. For example, I think that one likely upshot of my argument is that certain agents with the right qualifications would be morally justified in cloning brainless human organisms for the purpose of expanding the pool of viable organs for transplant. These agents may have a moral obligation to do this, but I will not argue that they have such an obligation.

The second clarification concerns the phrase 'weighty self-interest'. The term 'weighty' is vague. I don't think *any* self-interest counts as weighty? I will have more to say about this issue in Chapter 4, but here I will provide a simple example to give the reader a rough idea of what I mean by 'weighty'. If an act would extend your life by a year, and if, during this year, you would be healthy and in high spirits, and if the act had no negative

effects, then it would promote a weighty self-interest. The self-interest that you would have in this additional happy life counts as *weighty*. Thus, if it were necessary to kill and/or use a human animal to promote this self-interest, that would be morally justified on my view.

The third clarification concerns the phrase ‘apart from its effects on any human animal’. Without this phrase, the conclusion of the argument would be much less interesting. With the phrase included, the conclusion should raise eyebrows. For then the conclusion implies that the effects on any human animal—including the human animal killed and/or used—make no difference to the moral justification of the act, as long as this act satisfies the other conditions that I have mentioned. Why should we believe this? One my central aims is to explain why we should believe it.

## **1.6 How My Approach Differs From Others**

### **1.6.1 Differentiating Questions of Personal Ontology from Questions of Personhood**

One familiar way of trying to determine the moral status of a being is to try to determine whether it counts as a *person*. Thus, the question “What is a person?” has taken center stage in debates about abortion, stem cell research, and euthanasia, over the last fifty years.

The question “What are we?” is importantly different from the question “What is a person?” An answer to the first question would tell us the *properties* that something must have in order for the term ‘person’ to be correctly applied to it, and for us to treat it in certain ways. An answer to the second question will tell us *what kind of thing* has those properties—for example, whether it is an animal, or a featureless Cartesian ego, or a psychological continuer, etc.

Ethicists who disagree whether a human embryo or fetus is a person could agree that *you and I* are animals. For example, Jonathan Glover writes

It is *obviously true* that the normal foetus is at least a potential person: it is an entity which will, barring abnormal circumstances or intervention, develop into something incontestably a person. The only question is what moral claim upon us this gives it.<sup>25</sup>

Glover goes on to argue that a fetus's moral claim upon us is weaker than what many pro-life advocates argue it is. However, both Glover and his pro-life interlocutors seem to agree that *what has the property of being a person* is a human animal—an organism—that was once a mindless fetus and that could become a mindless vegetable.

Here is another example. In the preface of his *A Defense of Abortion*, David Boonin considers the implications of his pro-choice views about abortion for his son, Eli. Boonin writes that on his desk are several pictures of Eli at different ages, and that “through all the remarkable changes that these pictures preserve, he remains unmistakably the same little boy.” Next he writes,

In the top drawer of my desk I keep another picture of Eli. This picture was taken on September 7, 1993, 24 weeks before he was born ... There's no doubt in my mind that this picture, too, shows *the same little boy* in a very early stage in his physical development. And there's no question that the position I defend ... entails that it would have been morally permissible to end his life at this point.<sup>26</sup>

Boonin goes on to argue that fetuses like the one depicted in Eli's sonogram aren't persons and therefore lack the right to life. Yet there is “no doubt” in his mind that *Eli* and, presumably, that each of *us* was once such a fetus. Again, this seems to be a point of agreement between those on each side of the abortion debate.

---

<sup>25</sup> Jonathan Glover, *Causing Death and Saving Lives* (Harmondsworth: Penguin, 1977), 122.

<sup>26</sup> David Boonin, *A Defense of Abortion* (New York: Cambridge University Press, 2001).

The assumptions of Boonin and Glover, and of their opponents, regarding our persistence conditions, are common in debates about abortion; in such debates it is often taken for granted that there is a single unified *thing*, whose moral status is being debated. Those on the “pro-life” side of the debate point to a certain cluster of properties associated with a prototypical human animal and tell us that those are the properties that make something a “person”, and that because human fetuses (i.e. human fetal animals) instantiate enough of these properties *they* are persons. Meanwhile, those on the “pro-choice” side point to a different cluster of properties associated with a prototypical human animal and tell us that *those* are the properties that make it a “person” and that since human fetuses *lack those* properties, they *aren’t* persons. But the central question that is being debated by those interested in the question “What are we?”—for example, Animalists and defenders of the Psychological Approach—isn’t whether some cluster of properties makes a human animal a person, but whether *we* are human animals.

Of course, beings of *any kind* will have certain *properties* in virtue of which they belong to that kind. For this reason, one might think, there isn’t much difference between the question “What makes something a person?” and the question “What are we?”. However, this overlooks an important Aristotelian distinction between substance and accident. As David Wiggins points out, everything that exists is a “this such,” and it is the concept under which a thing falls as a *this such* that tells us what that thing *most fundamentally* is.<sup>27</sup> This concept will differ from others which tell us what the thing’s accidental or contingent properties are—for example concepts which tells us what the thing is doing, or how tall it is, etc. To highlight the distinction between substance and accident,

---

<sup>27</sup> David Wiggins, *Sameness and Substance* (Oxford: Blackwell, 1980), 15.

Wiggins and others usefully employ sortal concepts or *sortals*, which are classificatory concepts. They distinguish two types of sortal— substance sortals and phase sortals. A *phase sortal* refers to a kind to which a particular thing may belong through only part of its history. ‘Infant’ is a phase sortal because something can *become* an infant and later *cease to be* an infant while continuing to exist as one and the same individual. If something ceases to exist *as an infant*, this does not imply that *it ceases to exist*. A *substance sortal* refers to a kind to which a particular thing necessarily belongs throughout its existence—the kind to which that thing *must* belong if it is to exist at all. ‘Electron’ is presumably a substance sortal. Something that is an electron cannot stop being an electron and continue to exist; and it could not have started out as anything but an electron. As Eric Olson writes, it is a thing’s *substance sortal* that “determines persistence conditions that necessarily apply to all (and perhaps only) things of that kind.”<sup>28</sup> Boonin and Glover are clearly treating ‘person’ as a phase sortal—personhood is a mere phase that human animals—including non-conscious fetuses—can enter into pass out of.

I think a common mistake in the abortion debate is to assume (implicitly, at least) that there is *only one* substance sortal that designates *both* fetuses—whose moral status is a matter of controversy—and ordinary adult persons, whose moral status is uncontroversial. If this were true, then there is *only one relevant substance* whose moral status is uncertain; and the task (or one important task) of participants in the abortion debate is to determine when (i.e. at what point) this substance acquires moral status. Similarly, in debates about euthanasia where the patient is in a permanently vegetative state, the assumption that there is only one relevant substance—something that was *at one time*

---

<sup>28</sup> Olson, *The Human Animal: Personal Identity Without Psychology*, 28.

awake and conscious but then *later* lost the capacity for consciousness—suggests that the task (or one important task) of participants in the debate is to determine whether *that substance* has lost the moral status it once had, and, if so, whether it retains *any kind* of moral status in virtue of the moral status it once had.

This assumption—that there is only one relevant substance whose moral status must be determined—is similar to the assumption that in the case of the statue and the lump considered earlier there is only one substance that first acquires and then loses the property of being a statue. But, as we saw, there may be good reason to think that there are in fact *two* relevant substances in that case, and that there may also be good reason to think that there are two relevant substances in debates about what we are. The same is true of debates about abortion and euthanasia. In these debates, there may be not one but two relevant substances.

In fact, my approach to determining the moral status of non-conscious human animals—including non-conscious fetuses—depends on such a claim. I claim that in debates about the moral status of non-conscious human animals there are two relevant kinds of substance that one must consider—human animals and psychological continuers—and I claim that only one kind of substance, the psychological continuer, has moral status, because it is the only one of the two different kinds of substance that *could* have a self-interest.

### **1.6.2 The Moral Relevance of Personal Ontology**

But how *exactly* would our being psychological continuers make a moral difference? One answer, pertaining to abortion, is suggested by Jeff McMahan.<sup>29</sup> It is that if we are psychological continuers then the line separating abortion and contraception—which is usually drawn at the moment at which the human ovum is fertilized by an egg—should be redrawn at the first moment at which there is something that possesses mental properties of a kind that is specified by whatever the correct version of the Psychological Approach is. For the Psychological Approach seems to entail that each of us came to exist no earlier than the onset of consciousness, and it is unlikely that this occurs until many weeks after the fertilization of the ovum. If one kills a human animal before the onset of consciousness, then this does not kill *one of us* but merely prevents one of us from coming into existence and, in that sense, is relevantly like contraception. On most versions of the Psychological Approach, there will likely be borderline cases—those in which it is neither determinately true nor determinately false that one of us exists when the fetus is killed. (When exactly the onset of consciousness occurs—and which cases involve indeterminacy—is an issue that I will consider in Chapter 4.) Because an act of killing is usually thought to be morally much more serious than an act of contraception, the Psychological Approach would seem to support the view that some abortions—those in which the fetus lacks mental properties—are morally akin to more ordinary forms of contraception.

For reasons I consider in Chapter 4, I think this does not adequately explain why our being psychological continuers makes a moral difference to abortion. I think a more likely explanation is that if we are psychological continuers then human animals are not bearers of self-interest—they are not beings for whom life may go better or worse. One

---

<sup>29</sup> McMahan, *The Ethics of Killing*, 267-68.

possible explanation for this is that human animals cannot be conscious. At first glance, the claim that human animals cannot be conscious may seem incredible, but it is implied by certain versions of the Psychological Approach. For instance, consider the example of Emergentism discussed earlier. On this view, each of us is a non-physical being—a mind—that begins to exist when the human brain becomes complex enough to support consciousness. According to Emergentism, this non-physical being is the subject of consciousness. The human animal whose brain sustains this being isn't a subject of consciousness. Strictly speaking, there is nothing that it is like to be the animal. Now consider a part of the animal's history during which it was gestating in its mother's womb, prior to the emergence of the conscious mind. If, during this part of the animal's history, one had pointed to it and said '*that* human animal will one day be conscious', one would have said something that was, strictly speaking, false. For the animal has never been conscious, and hence, relative to that earlier time in its history, it would never be conscious. If, strictly speaking, human animals cannot be conscious, then, plausibly, human animals are not bearers of self-interest—for if something cannot be conscious, it is hard to see how its existence can be better or worse for it.

Notice, I am not claiming that beings that lack the capacity for consciousness are not bearers of self-interest. Rather, I am suggesting that if a being *cannot be* (and could not have been) conscious, then it is not a bearer of self-interest. That human animals (and other material beings) cannot be conscious is an immediate consequence of Emergentism. Because Emergentism implies that human animals (and other material beings) cannot be conscious, it supports the view that human animals (and other material beings) are not bearers of self-interest.



Some may think this “non-conscious animal view” cannot be right if materialism is true—that is if there are no non-physical beings. However, some materialists hold that no human animal can be conscious (again, strictly and literally speaking).<sup>30</sup> I am not sure whether we should say that, strictly speaking, animals are not conscious. If we should say that, then I think we will need some way of explaining how it can be correct to attribute mental properties to animals. Whether or not this striking claim about animals and consciousness is true, I think the explanation of why some animals cannot have self-interests will have something to do with consciousness.

### 1.6.3 Neutrality and Theoretical Cost-Efficiency

My approach to evaluating the killing and/or use of non-conscious human animals is important for another reason. One of the most surprising and interesting features of this approach is that it does not directly challenge the moral assumptions of many arguments commonly thought to support the claim that the killing and/or use of non-conscious human animals is morally *unjustified*. To demonstrate this, in Chapter 4, I consider some of the most well-known *pro-life arguments*—arguments which, if sound, would seem to directly undermine the personal ontology argument—and show that in fact these pro-life arguments are *compatible* not only with the conclusion of the personal ontology argument but with each premise of that argument. If these pro-life arguments were sound, they would not directly threaten the personal ontology argument. The claim that many mainstream pro-life arguments are compatible with the personal ontology argument may seem dubious.

---

<sup>30</sup> Sydney Shoemaker, “Self, Body, and Coincidence,” *Proceedings of the Aristotelian Society* 73 (1999) 287–306; Sydney Shoemaker, “Persons, Animals and Identity” *Synthese* 162 no.3 (2008) 313 – 324.

But I think that once the personal ontology argument—especially the first premise—has been fully elaborated and defended, the compatibility of these arguments will seem obvious. The key to understanding why they are compatible is that, although no premise of the personal ontology argument directly challenges any premise of the pro-life arguments, the truth of the first premise of the former argument would greatly restrict *the scope* of certain key premises in the latter arguments—it would restrict the range of cases to which those premises apply. As I will explain in Chapter 4, restricting the scope of a premise in the way that I have in mind is not the same as rejecting it.

Because the personal ontology argument does not directly challenge the moral reasoning on which many mainstream pro-life arguments depend, it has great dialectical and theoretical *cost-efficiency*. Its practical implications are wide-reaching—it supports a very important and controversial conclusion. Yet, its normative commitments are very weak. It does not appeal to a view of personhood. It does not appeal to a view of the conditions of one's having a right to be helped, or of one's having a right not to be killed or used to promote the interests of others. It does not appeal to a theory of the badness of death or of the wrongness of killing. And the notion of *interests* on which it depends does not presuppose that a being has interests *only if* that being has the capacity for consciousness. That the personal ontology argument does so much work without appealing to such claims is very important in light of the diversity of moral issues that must be sorted out before one can have a full understanding of each of the different practices that involve the killing and/or use of human animals—practices such as those discussed in Section 1.1.

The personal ontology argument's most controversial premise is its first premise—that we are psychological continuers. This has been the dominant view of what we are for

many years. (A closely related view about the persistence conditions of *persons* has been enormously popular since Locke.) That the personal ontology argument's most controversial premise has enjoyed such popularity among such a large group of varied thinkers gives me confidence that the argument is sound.

#### **1.6.4 Moral Uncertainty**

However, the argument could be mistaken. For this reason, acting on its conclusion involves taking a potentially serious moral risk. If the risk is serious enough, then acting on the argument's conclusion may be irrational, or perhaps subjectively impermissible—impermissible in a sense that is most directly relevant for guiding actions and policy decisions. This raises the issue of what to do when one is uncertain of the moral reasons that one has to act. This is extremely important, but it is not an issue that I take up in this dissertation. Those who find my arguments persuasive may also find them action-guiding to some extent. For example, one might think that consideration of my arguments warrants some adjustment in one's views and attitudes toward others, including politicians that take a firm stance on the moral issues to which this dissertation is relevant. However, ultimately, the extent to which my arguments can appropriately guide one in decision-making depends on a view of how it is rational or subjectively permissible for one to behave under conditions of moral uncertainty, as well as one's degree of moral uncertainty, and the severity of the moral risk that one would take in allowing one's actions to be guided by these arguments.

#### **1.7 A Brief Overview of the Remaining Chapters**

Chapters 2 and 3 set out the case against the Biological, Bodily, and Hybrid Approaches. Chapter 2 argues that the Biological and Bodily Approaches cannot counter the intuitive appeal of the Psychological Approach. Chapter 3 argues that the virtues of the Biological and Bodily Approaches do not outweigh those of the Psychological Approach, and presents independent reasons to reject the Hybrid Approach. Together, Chapters 2 and 3 constitute my defense of Premise 1 of the Personal Ontology Argument. Chapter 4 establishes Premise 2 of that argument. It also shows that the Psychological Approach restricts the scope of application of many pro-life arguments without challenging their major moral assumptions. Chapter 5 considers a challenge to Premise 2 of the Personal Ontology Argument based on the possibility of branching psychological continuity. Chapter 6 contains a summary of the dissertation and some concluding remarks.

## **2. Can Animalism Counter the Intuitive Appeal of the Psychological Approach?**

### **2.1 Introduction to Chapter 2**

In this chapter I consider whether Animalism—i.e., the Biological and Bodily Approaches—can counter or explain away the intuitive appeal of The Psychological Approach. The chapter considers and reinforces two powerful arguments against Animalism. The first of these arguments—*the Transplant Argument*—appeals to thought

experiments involving brain transplantation, in which the brain of a human person is transplanted into the decerebrate head of a different human body, perhaps remaining conscious throughout the process. To many people, it seems obvious that the person whose brain is thus moved from one animal to another would continue to exist in the animal that received his brain, while the animal he animated prior to the transplant would remain behind, either as a corpse or as a living animal sustained by external life support. If, however, the person can survive separation from his animal, he cannot have been identical to it. Since Animalism entails that each of us is identical to a human animal, the possibility of such brain transplantation is a counterexample to Animalism. The second argument, which is due to Jeff McMahan<sup>31</sup> and Ingmar Persson<sup>32</sup> appeals to an actual case: dicephalus conjoined twinning (Dicephalus). In dicephalus, there appear to be two human persons—two of us—but only one human animal, and so it appears that, contrary to Animalism, at least one of us is not a human animal. I consider a variety of actual and possible responses to the Dicephalus Argument on behalf of Animalists. I conclude that none of these responses is successful, and that Animalism is severely undermined by this argument.

If, as I will argue, Animalists cannot plausibly explain away or accommodate the intuitions that drive the Transplant and Dicephalus Arguments, then in order to successfully defend their approach, Animalists must argue that the theoretical virtues of their approach outweigh its heavy intuitive costs. The subsequent chapter, Chapter 3, examines the major

---

<sup>31</sup> McMahan, *The Ethics of Killing*.

<sup>32</sup> Ingmar Persson, “Our Identity and the Separability of Persons and Organisms” *Dialogue* 38 (1999), 519.

advantages that Animalism is believed to have over the PA, and argues that these advantages do not outweigh the intuitive costs.

## 2.2 The Transplant Argument

In what I will call *The Transplant Case*, my brain is removed from my head and transplanted into a numerically distinct human animal that is a genetic clone of me. The person who wakes up after the operation has a body that is qualitatively just like mine, has my original brain, and is mentally indistinguishable from me. This person goes on to inherit my job and material possessions, as well as my relationships with friends and loved ones, etc. In some cases, my brain is never isolated but remains attached to at least one human animal (or human body) for the duration of the operation.<sup>33</sup> *The Transplant Argument* against Animalism then proceeds as follows:

Premise 1: I am identical to the person who inherits my original brain and mental characteristics. (The Transplant Intuition)

Premise 2: If Animalism is true, then it is not the case that I am identical to this person.

Therefore,

Animalism is false.

Most philosophers of personal identity have found this argument compelling. A few of the many examples are listed below.<sup>34</sup> Moreover, some empirical evidence suggests that The

---

<sup>33</sup> Parfit, "Persons, Bodies, and Human Beings," in *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne and Dean W. Zimmerman. (Oxford: Blackwell Publishing, 2008).

<sup>34</sup> H.P. Grice, "Personal Identity," *Mind* 50 (1941), 330-50; Derek Parfit, *Reasons and Persons* (Oxford, Oxford University Press, 1984); Peter Unger, *Identity, Consciousness, and Value* (New York, Oxford University Press, 1990); McMahan, *The Ethics of Killing*.

Transplant Intuition (Premise 1) is widely held.<sup>35</sup> What reason might there be to reject this argument? In what follows, I will consider several different attempts to block the Transplant Argument, and shall explain why each fails.

## 2.2.1 Responding to Objections to the Transplant Argument

### 2.2.1.1 Are Psychological Traits Linked to Particular Bodies?

Bernard Williams, in his discussion of body-swapping thought experiments, points out that there are limitations with regard to character and mannerisms, in our ability to imagine cases such as The Transplant Case:

The expression of some personal traits may be essentially connected with our particular bodies. In that case, how could they survive the “swap”? If they could not, then there is reason for questioning the intelligibility of these stories as involving the bodily exchange of persons.<sup>36</sup>

Those who agree with Williams might claim that we have insufficient reason to accept The Transplant Intuition (i.e. Premise 1 of The Transplant Argument), since we have insufficient reason to believe that there would be enough psychological continuity between me before the transplant and the person who inherits my brain.

However this response is unconvincing, since we can assume that the animal that starts out with my brain and the animal that ends up with my brain are genetic clones of each other. It is hard to believe that physiological differences between the two animals would give rise to *profound* psychological differences. Moreover, the plausibility of The Transplant Intuition does not seem to depend on the assumption that the person who

---

<sup>35</sup> S. Nichols and M. Bruno, “Intuitions about personal identity: an empirical study” *Philosophical Psychology* 23 no.3 (2010), 293-312; L.J. Rips, S. Blok, and G. Newman, “Tracing the Identity of Objects” *Psychological Review* 113 (2006), 1-30.

<sup>36</sup> Bernard Williams, “The Self and the Future,” *The Philosophical Review* 79 (1970), 166 – 167.

receives my brain has *all* of my psychological characteristics. I suspect that to generate the intuition, we need only assume that this person's experiences immediately after the transplant stand in a certain relation to my experiences immediately before the transplant; specifically, we need only assume that these experiences are parts of the same continuous stream of consciousness. For example, consider the momentary experience that you had when you began reading this sentence and the momentary experience that you are having *right now*. If your short-term phenomenal memory serves you correctly, you should recall the first of these two momentary experiences merging seamlessly with, or flowing into, the second. If this isn't obvious, think instead "of what it is like to suffer a continued toothache, or to hear an extended tone played on a flute, or to watch a balloon float slowly across the sky." As Barry Dainton and Tim Bayne point out, in cases like this, the transition between each brief phase of one's experience is so seamless that "the division of the experience into distinct phases can seem entirely arbitrary."<sup>37</sup> This seeming of earlier phases of one's experience as "flowing into" subsequent phases is a general feature of phenomenal consciousness. Let us say that two or more experiences are parts of a continuous stream of consciousness, or *phenomenally continuous*, if and only if they are related in this way.

If my pre-transplant experiences are phenomenally continuous with the post-transplant experiences of the person who inherits my brain then, plausibly, I am numerically identical to this person, even if many of *his* psychological features (e.g., memories, beliefs, desires, etc.) are very different from *mine*. Intuitively, one could survive even an enormous change in one's psychology, perhaps without noticing this change as it

---

<sup>37</sup> Barry Dainton and Tim Bayne, "Consciousness as a Guide to Personal Persistence" *Australasian Journal of Philosophy* 83 no.4 (2005): 554.



occurs, provided that one's experiences prior to the beginning of the change were phenomenally continuous with experiences that occurred during or after the change.

Williams himself presents a hypothetical case that supports this. He imagines that an evil neurosurgeon will soon tamper with his brain, causing both amnesia and intense pain.<sup>38</sup> According to Williams, knowing that he would lose the memories of his life up to the start of his anticipated pain would give him no less reason to fear this pain. He finds no reason to believe that his pain would suddenly cease with the onset of his amnesia. Critics of Williams have pointed out that this claim begs the question against views of personal identity that imply that sudden amnesia would bring one's existence to an end. Yet Williams's claim seems overwhelmingly plausible, especially if we grant that the throwing of the switch would leave intact the kind of *short-term* phenomenal memory that would be sufficient for Williams's experiences immediately before the throwing of the switch to be phenomenally continuous with experiences occurring immediately afterward. Rather than stating his view in a question-begging way, Williams could have stipulated that the sudden onset of the amnesia in his example would not disrupt phenomenal continuity, and hence, that the experiences before and after the sudden onset of the amnesia would be phenomenally continuous. He could then have claimed that because the experiences up to the start of the onset of amnesia would be phenomenally continuous with those immediately after the onset of amnesia, the stipulation that the former are had *by him* makes it overwhelmingly plausible to suppose that the latter would also be had *by him*. Had Williams stated his argument in terms of phenomenal continuity between different

---

<sup>38</sup> Bernard Williams, "The Self and the Future," 161-80.

experiences, he could have provided a non-circular argument for his overwhelmingly plausible judgment that *his* pain would not cease with the onset of sudden amnesia.

### **2.2.1.2 Can there be Continuous Consciousness between Two Different Human Animals?**

At least one animalist, Peter van Inwagen, would deny that my experiences are phenomenally continuous with those of the person who receives my brain in the transplant case. “To imagine whether a certain situation contains a continuous consciousness,” van Inwagen writes,

we have to find out first whether it contains a continuously existent [subject]. We can’t do things the other way around. We can’t find out whether the situation contains a continuously existent [subject] by first finding out whether it contains a continuous consciousness.<sup>39</sup>

According to van Inwagen, the claim that some subject remains conscious throughout the transplant operation cannot be inferred on the basis of the assumption that there is continuous consciousness throughout that operation. We cannot simply assume that experiences occurring at the beginning of the operation are phenomenally continuous with those occurring at the end of the operation. We must first provide independent reason for thinking that some subject *exists* throughout the operation, and then we must determine whether this continuously existing subject plausibly remains conscious throughout the operation. For van Inwagen, the only subjects in the transplant case are the two human animals, call them *O1* and *O2*. Because neither *O1* nor *O2* plausibly remains conscious

---

<sup>39</sup> Peter van Inwagen, *Material Beings*, 206. Here, I have substituted the term ‘subject’ for van Inwagen’s term, ‘thinker.’ I do not think that he would object to this substitution; moreover, my response to him does not hinge on it.

throughout the entire transplant operation (O1 loses its brain near the beginning of the operation and O2 has no brain until near the end of the operation), van Inwagen would insist that *no one* remains conscious throughout the operation, and that therefore, my pre-transplant experiences aren't phenomenally continuous with the post-transplant experiences of the person who inherits my brain.

It seems van Inwagen must say that there are *two distinct* streams of consciousness in the transplant case; one stream begins before the transplant operation and ends when O1's brain is removed, and the other stream begins when that brain becomes part of O2 and ends sometime after the transplant operation. This view of how the conscious experiences in the transplant case are related is more plausible on the assumption that consciousness is essentially an activity of the animal *as a whole*, not localized in any particular neural structure of the brain. But suppose that at each time during the transplant operation the brain is part of at least one of the two animals—O1 and O2. For example, suppose that rather than moving the brain from the head of O1 to the head of O2, the surgeon initially removes most of O2's head, (perhaps leaving the lower brain and brainstem intact), and then *conjoins* O1 and O2 so that O1's original head and brain are parts of both O1 and O2.<sup>40</sup> Suppose that O1 and O2 are then surgically separated so that O2 keeps the head and brain, while O1 becomes headless. In this case, the brain is never completely isolated, but is instead passed from one animal to the next, as a baton is passed between runners of a relay race.

Still, van Inwagen would probably deny that the pre- and post-transplant experiences are phenomenally continuous, since even in this version of the case, neither

---

<sup>40</sup> Parfit, "Persons, Bodies, and Human Beings."

O1 nor O2 remains conscious throughout the transplant operation. But in this case, denying that the pre- and post-transplant experiences are phenomenally continuous seems implausible. To see why, suppose that while the brain is part of both O1 and O2, the consciousness-generating parts are shared equally, so that we could not plausibly claim that experiences occurring during the sharing of the brain would belong either to O1 or O2 but not to the other. It would then seem that *each* animal was a subject of these experiences. Now suppose

- (1) Experiences that O2 has while O2 is sharing a brain with O1 are phenomenally continuous with experiences that O2 has *after* the transplant (when the brain is no longer being shared with O1 but belongs uniquely to O2).

Van Inwagen should allow that (1) is possible; it is consistent with his view that phenomenally continuous experiences belong to a continuously existing subject. In this case, O2 is the continuously existing subject of all of the experiences that occur during the relevant time period (i.e. starting while the brain is part of both O1 and O2, and ending after the transplant operation is complete). However, we are currently supposing that the experiences that O2 has while O2 shares a brain with O1 are *also* experiences that O1 has.

Thus, it seems, van Inwagen should agree that the following is also possible:

- (2) Experiences that *O1* has while O1 shares a brain with O2 are diachronically phenomenally continuous with experiences that O2 has after the transplant operation is complete, *even though O1 is not conscious then*.

Next, suppose

- (3) Experiences that O1 has before O1 shares a brain with O2 are phenomenally continuous with experiences that O1 has while O1 is conjoined with O2.

Again, it seems, van Inwagen should allow that (4) is possible, since O1 exists throughout the relevant period in which these experiences occur.

We now have the following result: An experience, E1, which occurs *before* O1 shares its brain with O2, is phenomenally continuous with another experience, E2, which occurs *while* O1 shares its brain with O2, and E2 is phenomenally continuous with a third experience, E3, which occurs *after* O2 shares its brain with O2 (when the transplant operation is finished). Now, setting aside aberrant cases of divided minds, diachronic phenomenal continuity as I have characterized it seems *transitive*. According to this characterization, two experiences are phenomenally continuous whenever they are both parts of the same continuous stream of consciousness. And, it seems, if E1 and E2 are parts of a single continuous stream of consciousness, and E2 and E3 are parts of a single continuous stream of consciousness, then E1 and E3 should also be parts of a single continuous stream of consciousness. But van Inwagen's view rules this out, since there is no continuously existing subject of both E1 and E3.

But if E1 is phenomenally continuous with E2, and E2 is phenomenally continuous with E3, it is hard to believe that E1 and E3 would *not* be phenomenally continuous, *at least in this case*. We have reason to believe *a priori* that phenomenal continuity is transitive whenever the relevant experiences are temporally ordered i.e., whenever we have experiences A, B, and C, such that A is phenomenally continuous with B, and B is phenomenally continuous with C, AND A occurs *before* B, which occurs *before* C. (Cases of divided minds involving, e.g., division and split-brains are special exceptions because

there is no temporal ordering of the experiences that seems to violate transitivity.) Moreover, if (3) is true, then the experiences of one subject can be phenomenally continuous with those of a numerically different subject; there can be continuous consciousness between numerically different subjects. This rules out one possible reason for denying that O1's pre-transplant experiences could be phenomenally continuous with O2's post-transplant experiences, namely that no subject could have experiences that were phenomenally continuous with the experiences of a numerically different subject. Once this possibility is ruled out, the claim that O1's pre-transplant experiences and O2's post-transplant experiences cannot be phenomenally continuous seems even more mysterious.

If we do not deny the transitivity of phenomenal continuity in this case, then it seems we are left with three options. First, we could deny van Inwagen's view that such phenomenal continuity presupposes a continuously conscious subject. In that case, it seems, we would have no reason to deny that there could be continuous consciousness in the transplant case. Second, we could accept van Inwagen's view but reject Animalism in favor of the PA. We could then say that there *is* a continuously conscious subject in the transplant case—a person who is associated first with O1 and then with O2. Finally, we could accept both van Inwagen's view and Animalism, but insist that any experiences occurring while O1 and O2 are sharing the same brain belong to O1 or O2, but not to both. On this view, experiences cannot be “shared” by two or more subjects; either (i) every experience that occurs during O1's and O2's sharing of the brain belongs uniquely to O1, (ii) every experience that occurs during the sharing of the brain belongs uniquely to O2, or (iii) there are two distinct but overlapping streams of consciousness during the sharing of

the brain, one belonging to O1, the other to O2. But (i) and (ii) are both completely arbitrary, and (iii) just seems incredible.

It is worth pointing out that van Inwagen tries very hard to avoid such a proliferation of conscious minds. For example, he denies that human animals have large undetached parts, such as heads and upper halves, on the grounds that such parts would appear to be capable of conscious experience. If one also assumes a view of the subject-experience relation on which the correspondence between subjects and experiences is necessarily one-to-one, then the number of conscious experiences occurring in the vicinity of the animal at any particular time would be equal to the number of its conscious parts at that time *plus one* (the animal itself). (Dean Zimmerman has discussed this problem at length.)<sup>41</sup> Therefore, it seems, we should reject either Animalism or the view that continuous consciousness presupposes a continually conscious subject. Those who are deeply committed to the first view should reject the second. But if they reject this view, then they cannot appeal to it, as van Inwagen does, in order to resist the transplant argument.

### **2.2.1.3 Are Judgments about Identity Corrupted by Judgments about “What Matters” in Survival?**

Another response to Premise 1 of the transplant argument has been offered by Eric Olson. We discussed the basis of Olson’s response briefly in Chapter 1. It is Derek Parfit’s thesis that psychological continuity—not identity—is what matters in survival. According to

---

<sup>41</sup> Dean Zimmerman, “The Constitution of Persons by Bodies: A Critique of Lynne Rudder Baker’s Theory of Material Constitution” *Philosophical Topics* 30 (2001) 295-338.

Olson, because in most ordinary cases of survival, what matters—psychological continuity—coincides with identity, we mistakenly infer from the observation that I am psychologically continuous with the person who receives my brain, that I am identical to this person. Thus, according to Olson, the transplant intuition is the result of a mistaken overgeneralization from ordinary cases of survival in which the relation of our identity over time, namely (non-branching) biological continuity, coincides with what matters in those cases, namely psychological continuity. When we consider the transplant case, we think that I am the person who inherits my brain and psychology because we think that I ought to be prudentially concerned about this person; or because this person would be morally responsible for my past actions; or because everyone would be justified in treating this person *as if* he were me. In other words, “the transplant intuition is based on practical concerns that, while perfectly valid, don’t necessarily coincide with numerical identity.”<sup>42</sup>

It is not clear that Olson can adopt Parfit’s thesis that identity is not what matters in order to explain away the transplant intuition. Parfit’s thesis depends mainly on the argument from fission,<sup>43</sup> which relies on three claims: First, it is possible that the relation in which I stand to each of two separate future persons, A and B, is such that either one of these persons, but for the existence of the other, would be numerically identical to me (and each is an equally good candidate for being identical to me, and there is no other better candidate for being identical to me); second, the correct (or best) description of this situation is that I am identical *not* to A, and *not* to B, but rather to *no one* (i.e. no future person) and that, therefore, I cease to exist; third, from a prudential standpoint, it is rational

---

<sup>42</sup> Olson, *The Human Animal*, 48.

<sup>43</sup> *Ibid.*, Chapter 3.



for me to regard this situation as being as good (or about as good) as a case of ordinary survival in which there *is* some future person who is identical to me—what matters in ordinary survival is preserved. In this case, what matters in ordinary survival is preserved, even though identity is lost. It would seem to follow from this that identity is not what matters. But if we accept this argument from fission, the most we can say is that identity, whatever that consists in, is *not* what matters. We cannot conclude straightaway that psychological continuity *is* what matters. If what underpins my identity over time in ordinary cases of survival is psychological continuity of a certain kind, then the argument from fission would support the claim that psychological continuity of this kind—not identity—is what matters in these ordinary cases of survival. But according to Animalists, what underpins my identity over time in ordinary survival isn't psychological continuity but *biological or bodily continuity*. On Olson's view, the relevant sort of biological continuity includes the continuity of life processes, such as respiration and digestion, for which psychological continuity is neither necessary nor sufficient.<sup>44</sup> It is therefore unclear whether Olson can plausibly adopt Parfit's thesis that psychological continuity is what matters in ordinary survival.

As Parfit argues, the claim that psychological continuity, and not identity, is what matters in survival<sup>45</sup> does not support Olson's claim that the transplant intuition is due to a mistaken overgeneralization from ordinary cases of survival in which psychological and biological continuity coincide.<sup>46</sup> Parfit refutes this claim by considering how a defender

---

<sup>44</sup> Ibid., 11.

<sup>45</sup> Parfit, "We Are Not Human Beings," 5-28.

<sup>46</sup> Olson, *The Human Animal*. Chapter 3. The claim that these relations almost always coincide is not, I believe, true, even on Olson's view that biological continuity is the criterion of identity over time. As Olson often mentions in his writings, we are not psychologically continuous with early fetuses, or with adult

of *the Finger Print Criterion*, according to which “some future person would be the same as some present person if and only if they have qualitatively identical finger prints,” might try to debunk the opposing intuition that one would survive the remolding of one’s fingertips, by claiming that “in nearly all actual cases, people with the same psychology also have the same finger prints” and that the opposing intuition is just the result of our mistaking *what matters* (psychological continuity) for identity (finger print continuity).<sup>47</sup> This debunking explanation offered by the defender of *The Finger Print Criterion* has no plausibility whatsoever because the mere fact that in nearly all actual cases of psychological continuity coincides with fingerprint continuity provides no reason for thinking that the intuition that one could survive the remolding of one’s fingertips is merely the result of a mistaken overgeneralization from these actual cases. Similarly, if psychological continuity coincides with biological continuity in most cases of ordinary survival this should not lead us to think that the transplant intuition is due to a mistaken overgeneralization.

Olson might object that Parfit’s example seems convincing only because *we know* that *the Finger Print Criterion* cannot be the right criterion of identity. However, this objection would miss the point: If, in hypothetical cases, theory A seems much less plausible than theory B, then the fact that A and B make the same identifications in most actual cases doesn’t support the claim that theory A is more plausible than these

---

human animals in a permanent vegetative state. Given that so many people judge that each of us was once an early fetus, and that those in a permanent vegetative state were once conscious, when it is clear that there is no psychological continuity in these cases, one might think that there is some pressure on Olson to explain how one could be so easily lured into thinking that in the transplant case one is identical to the person who inherits one’s brain and psychology. Why wouldn’t one feel more compelled to assume that one is the animal that remains behind on the operating table after losing its brain?

<sup>47</sup> Parfit, “We Are Not Human Beings,” 10.

hypothetical cases seem to show, or that people who thought theory B yielded the correct answer were merely overgeneralizing from usual cases. In short, the fundamental difficulty for Olson's debunking explanation is that there does not appear to be compelling evidence in support of his hypothesis that the transplant intuition is the result of a mistaken overgeneralization.

#### **2.2.1.4 Can Human Animals Think Remotely?**

A novel and ingenious response to the transplant argument has recently been proposed by the Animalist, Rory Madden. On Madden's view, when O1's brain is transferred to O2, O1 remains the thinker of what *seem* to be O2's thoughts; after the transplant, O1 thinks *remotely* with his former brain, even though that brain belongs to O2. Madden offers Animalists a novel way of accommodating the transplant intuition; they can say that the subject of one's thoughts prior to the transplantation of one's brain is identical to the subject of the thoughts occurring after the transplant. Therefore, they can accept Premise 1 of the transplant argument, while denying Premise 2, namely that if Animalism is true, I am not identical to the person who inherits my original brain and mental characteristics. According to Madden, I am identical to this person, not because I switch bodies when my brain is transplanted, but because I continue to use that brain even after I become empty-headed.

At first glance, the notion of "remote" thought seems incredible. However, Madden tries to warm readers up to this idea by discussing an analogous case of "remote" speech:

Tracy has had a tracheotomy. However, the efferent nerve endings of her speech centers have been connected by remote control radio links to the voice box, jaw and tongue of another person Victor. Victor now has no control over these parts of his body. As a

result of this arrangement Tracy's utterances issue from Victor's mouth, including first-person utterances such as 'I am Tracy' and perhaps 'My voice sounds uncannily deep and distant'. What are the truth-conditions of these first-person utterances? Token uses of the first-person pronoun 'I' refer to whichever person produced them - to the speaker of the linguistic context in question. So the utterance 'I am Tracy', for example, will be true if and only if the speaker in question is Tracy. And we are surely inclined to judge that, in the arrangement as described, this sentence will be true. Tracy speaks truly, albeit by remote means. Victor doesn't speak. That is to say, the tokens of 'I' issuing from Victor's mouth refer not to Victor but to Tracy.<sup>48</sup>

According to Madden, my thinking remotely in the transplant case is similar to Tracy's speaking remotely in this case. Madden explains that what makes Tracy the referent of I-utterances issuing from Victor's mouth is also what makes O1 the referent of I-thoughts generated by O2's brain. On Madden's proposal, the correct assignment of a particular referent to 'I' is *knowledge-maximizing*, where knowledge must meet a safety condition: for S to know p it must be the case that S could not *easily* have believed falsely that p. In the case of remote speech, the assignment of Tracy to 'I' maximizes knowledge, since the true utterances produced by Victor's mouth are justifiably believed by Tracy but not by Victor. When Tracy says, using Victor's mouth and voice box, "I have been to Venice," even if Tracy and Victor have both been to Venice, the utterance is *safe* for Tracy but not for Victor. This utterance could easily have issued from Victor's mouth if he had not been to Venice; thus, the belief expressed by the utterance on the assignment of Victor to 'I' could easily have been false. But "due to the dependence of the utterance on Tracy's memories, the utterance would not have issued from Victor's mouth had Tracy not gone to Venice." Thus, when Tracy is assigned to 'I' the utterance 'I went to Venice' is safely true. Similarly, assignments of O1 to 'I' will tend to maximize knowledge since most of the I-

---

<sup>48</sup> Roy Madden, "Externalism and Brain Transplants," In *Oxford Studies in Metaphysics* 6, edited by Karen Bennett and Dean Zimmerman (Oxford University Press: Oxford, 2011), 3.

thoughts occurring in Recipient's brain will be safely true when O1 is assigned to 'I' but not when Recipient is assigned to 'I'. Other writers have defended this safety condition<sup>49</sup> however, as Madden points out, what's crucial on his account isn't the safety condition, but the claim that knowledge requires reliability.

I think we should reject Madden's proposal. Intuitively, what makes Tracy the speaker of utterances issuing from Victor's mouth in the case of remote speech has nothing to do with knowledge. Rather, there are certain *causal connections* between the intentions that Tracy forms and Victor's mouth and voice box, which, in this strange case, seem to suffice for Tracy's "using" Victor's mouth to speak her own words and thoughts. A related problem is that there are cases in which assignments of Victor to tokens of 'I' spoken by Tracy yield the same number of safely true beliefs as assignments of Tracy to those tokens, but in which it is nevertheless obvious that Tracy—not Victor—is the speaker. Suppose, for example, that in addition to attaching the efferent nerve endings of Tracy's speech centers to Victor's mouth, evil scientists sewed Victor's and Tracy's bodies together when they were born; as a result, they've gone everywhere together, experienced the same things, etc. If I-utterances emitting from Victor's mouth were general enough, it could be that every such utterance was safely true whether Victor or Tracy was assigned to 'I'. Alternatively, we might imagine Tracy as a compulsive liar, half of whose utterances are intended to deceive others. It could be that by some bizarre coincidence, assignments of Victor to I-tokens emitting from his mouth yield safely true beliefs because, although what Tracy intends to communicate to others with these I-utterances is false *of her*, it is true of

---

<sup>49</sup> Ernest Sosa, "Contextualism and Skepticism," in *Philosophical Issues*, edited by J. Tomberlin, Supplement to *Nous* 34 (2000), 1-18.

Victor, and couldn't easily have been false of him. But in all of these cases, Tracy is quite obviously the speaker.

### 2.3 Another Problem for Animalism: The Severed Head Case

Another problem for those who want to resist the transplant argument is the apparent possibility of an isolated human head or brain in a vat that is capable of generating consciousness, even though it is not a part of any human animal. Suppose that the head of a person with normal anatomy is severed, kept functional and conscious throughout the process of separation, and is then sustained through the provision of blood from an external source. Although the rest of the body is destroyed, the head continues to generate consciousness; it appears to be awake and responsive, communicative, etc. Let us call this *The Severed Head Case*.<sup>50</sup> Now, consider two conflicting claims:

- (i) The only conscious being in this case is one of us. This being is a human animal that loses most of its major parts, and is thereafter the size and shape of an ordinary human head; it is a *head-sized* human animal.
- (ii) The only conscious being in this case is one of us, and this being stands in some relation (e.g., constitution) to the severed head, but it is *not* a human animal.

We can accept only one of these claims. Which is more plausible?

#### 2.3.1 Appealing to the Importance of the Brainstem

---

<sup>50</sup> This case was originally discussed in Timothy Campbell and Jeff McMahan, "Animalism and the Varieties of Conjoined Twinning".

First, consider (i), the claim that the isolated head belongs to a human animal that has lost all of its other major parts. The reason that Animalists give for thinking that the animal survives with only a head is not (and cannot be) that it retains the capacity for consciousness. For, again, they claim that our identity conditions are not psychological but bodily or biological. There is apparently no good reason to suppose that the existence of an *organ* that generates consciousness is sufficient for the existence of an *animal*. Yet, some Animalists, such as Olson and van Inwagen, claim that a fully developed human animal dies if and only if its brain dies because the brain, and in particular the brain stem, is what regulates and coordinates the functions of the animal. The brain stem is the “biological core” of the animal, the internal integrator of the various life processes that jointly constitute the animal’s life, and hence, are jointly necessary for its continued existence (according to those who claim that the human animal is essentially alive). These Animalists therefore argue that an animal may be pared down to an isolated head or an isolated brain or even an isolated brain stem.

I think we should reject the claim that an animal could survive with nothing but a severed head that includes a functional brain stem. Neither the brain stem nor the brain as a whole is either necessary or sufficient for the integration of the function of bodily organs. That the brain and brain stem are not necessary is demonstrated by a case reported by the neurologist Alan Shewmon. In this case, a four-year-old boy was diagnosed as brain dead. Because his mother refused to remove him from life support, his body was transferred to her home where, with only ventilation, provision of nutrition and hydration, and basic nursing care, it continued to function normally apart from its being unable to generate

consciousness.<sup>51</sup> More than fourteen years later, Shewmon was permitted to perform an examination, which revealed that “the entire brain, including the stem, had been replaced by ghost-like tissues and disorganized proteinaceous fluids.” Yet, Shewmon observed that “while ‘brain dead,’ he has grown, overcome infections and healed wounds.”<sup>52</sup> What this case shows is that even in the absence of any brain at all, a human body can remain comprehensively functional for years with no more external life support than that required by many fully conscious human beings. The basic biological functions of the boy’s body remained internally coordinated; but the integration was decentralized, with organs responding to signals from other organs, rather than centralized in the activities of the brain stem.

### **2.3.2 Appealing to the Importance of Self-Maintenance**

Animalists might argue that what is necessary for a human animal to continue to exist is not that it retain a functional brain stem, but that it continue, with minimal external support, to be self-regulating and self-sustaining. According to van Inwagen, a living head satisfies that condition, while a headless body does not.

Give the severed head the proper environment and it will maintain itself, but the headless body will need a constant supply of “instructions” in the form of electrically transmitted information. Unlike the head, it will not be able to coordinate its activities. A life-support system for the head will be no more than an elaborate pump. A mechanical life-support system for the headless body must involve the functional equivalent of a computer.<sup>53</sup>

---

<sup>51</sup> Shewmon, “Brainstem Death, Brain Death, and Death,” 125-45.

<sup>52</sup> *Ibid.*, 36.

<sup>53</sup> van Inwagen, *Material Beings*, 177-78.



Neither of these empirical claims is true. As Shewmon's case illustrates, a headless body can remain functionally integrated with the assistance of a pump, albeit one that supplies oxygen rather than blood. A severed head, by contrast, requires more than a pump. The blood that the pump carries to the head has to be continuously renewed, a function that can be performed by the bone marrow in a headless body but not by the head itself. The blood must also be cleansed of toxins and supplied with immune cells and oxygen—functions that again can be carried out by a brainless body, such as the one described by Shewmon, but not by a head. Indeed, a living but isolated head has no internal regulation or integration, even of a *decentralized* sort. The regulatory capacities of the brain stem are idle since they are concerned with areas of the body to which the head is no longer connected. The constituent organs of the head, such as the brain and the eyes do not cooperate in maintaining biological homeostasis. Almost everything they need for survival must be externally supplied. If a sufficient degree of self-sustenance is necessary for being an animal and if, despite having a fully functional brain stem, an isolated head is neither self-regulating nor self-sustaining, then it would seem that having a fully functional brain stem is not sufficient for being an animal.

### **2.3.3 Animalism and the Severed Head Intuition: An Awkward Fit**

The only other reason for an Animalist to claim that having a functional brainstem is sufficient for being a human animal is that it just seems overwhelmingly plausible that where there is an isolated human head that is apparently capable of generating consciousness, and where there is apparently something that is awake and responsive, and capable of communication, etc., there is *one of us* that has these characteristics. Let us call

this *The Severed Head Intuition*. For the Animalist, accepting (i), the claim that the isolated head belongs to a human animal that has lost all of its other major parts, may be the only reasonable way of squaring her view with The Severed Head Intuition. But for the Animalist, this is not a sufficient reason for accepting (i). The Animalist must offer *independent evidence* for the hypothesis that a human animal can survive with a functional head (including a functional brain and brainstem) as its only major part. Otherwise, we ought to think that The Severed Head Intuition counts against Animalism. After all, we could easily accommodate this intuition by rejecting (i) in favor of (ii)—the claim that the isolated head belongs to a being that *is* one of us but *is not* a human animal.

The Animalist might respond as follows:

Our reasons to accept (i) include *both* The Severed Head Intuition *and* that the preponderance of evidence supports Animalism. Even if no independent evidence can be offered for the hypothesis that a human animal can survive with a functional head (including a functional brain and brainstem) as its only major part, we should nevertheless accept (i), since it is the most reasonable thing to believe *given both Animalism and The Severed Head Intuition*.

The first thing to note about this response is that it assumes that the case for Animalism is already sufficiently strong that we should accept that approach even without considering the metaphysical implications of The Severed Head Case. I will examine this claim in Chapter 3, where I discuss the arguments in favor of Animalism.

The second thing to note about this response is that those who accept it will be in the awkward position of claiming that, although Animalism cannot explain how a human animal can survive with a functional head (including a functional brain and brainstem) as

its only major part, we should nevertheless accept both Animalism and The Severed Head Intuition. This dialectical move seems unprincipled. Animalists should not accept The Severed Head Intuition unless they can accommodate that intuition in a principled way.

### **2.3.4 Can there be Consciousness without a Subject?**

If Animalists reject The Severed Head Intuition there is another interpretation of The Severed Head Case they might accept. According to this interpretation, in The Severed Head Case, although there *appears* to be a conscious being that is thinking and communicative, in fact there is no such being. After the head is severed, there is no human animal, and hence, no human entity of our kind, to whom the severed head belongs. The conscious experiences generated by the head before it is severed are experiences of a human animal; but the conscious experiences generated by the head after it has been severed—experiences that are phenomenally continuous with the experiences prior to the severing—are not experiences *of anyone*.

This view is considerably less plausible than the two alternatives that we have considered—(i) and (ii). Many philosophers find the notion of consciousness without an accompanying subject incoherent. Galen Strawson calls such views “crazy,” claiming that “it is an obvious conceptual truth that an experiencing is necessarily an experiencing by a subject of experience.”<sup>54</sup> Unlike Strawson, I don’t find the concept of subject-less experience incoherent. What I find incoherent is the view that some experiences within a particular stream of consciousness have a subject while others don’t. Fine if there are no

---

<sup>54</sup> Galen Strawson, “What is the Relation between an Experience, the Subject of the Experience, and the Content of the Experience?” unpublished manuscript, 20.

subjects of experience—but only bundles of experience. Not fine if, although experiences *do* belong to subjects, they can also exist as independent free-floating entities or events. This is too much of a stretch of my concept of the relation between experience and subject. To illustrate the problem with this, van Inwagen cites the example of the Cheshire Cat from Lewis Carroll's *Alice and Wonderland* who sometimes disappears, leaving his smile behind. Van Inwagen calls this “a metaphysical joke”; it's incoherent, he thinks, because modifications of substances, such as smiles and experiences, couldn't exist independently of the substances they modify. Although bundle theorists reject van Inwagen's account of the relation between subject and experience as between substance and accident, views on which experiences belong to some particular subject or other seem vulnerable to van Inwagen's charge. Experiences may be regarded as tropes that don't inhere strictly in the subjects that have them, but no one, to my knowledge, has argued that such experiences could exist without *any subject at all*.

## **2.4 The Dicephalus Argument against Animalism**

### **2.4.1 A Famous Case of Dicephalus: Abigail and Brittany Hensel**

Before returning to The Severed Head Case, I want to consider another problematic type of case for Animalism, which involves a rare kind of conjoined twinning. The first type of case that I will consider, Dicephalus, was originally proposed by Jeff McMahan<sup>55</sup> as a counterexample to Animalism. McMahan's “dicephalus” argument against Animalism has

---

<sup>55</sup> Jeff McMahan, “The Metaphysics of Brain Death.” *Bioethics* 9 (1995), 91–126.

since been extended and developed by others.<sup>56</sup> Dicephalus occurs when a human zygote divides incompletely, resulting in twins fused below the neck. In a case featured in *Life* magazine, Abigail and Brittany Hensel appear as two heads emerging from a single torso; yet they appear to be separate and distinct persons, each with her own private mental life and character. Each appears to feel sensations only on her own side of what is naturally described as a single human body, and each appears to have exclusive control over limbs on her side.

However, what appear to be *two separate* persons seem to share *a single* animal. Although there are two hearts, two esophagi, and two stomachs, there are three lungs, a single liver, a single small intestine, a single large intestine, and single urinary, circulatory, immunological, and reproductive systems (so that any child they might conceive would have three persons as biological parents: a father and two mothers). Their organs are contained in a single rib cage and function together in a harmonious, coordinated manner. The limited duplication of organs would appear to be contingent. Recorded cases of Dicephalus show varying degrees of duplication and it seems possible that there could be an even purer case than that of the Hensel twins, in which there would be virtually no duplication of organs below the neck (for example, only one heart and two esophagi leading to a single stomach). It is certainly plausible that there are two distinct animals in cases of superficially conjoined twins, in which there is only very limited sharing of non-vital organs but extensive duplication of others; but it is much less plausible in cases, such as

---

<sup>56</sup> Persson, "Our Identity and the Separability of Persons and Organisms," 519; McMahan, *The Ethics of Killing*, 35 – 39; Timothy Campbell and Jeff McMahan, "Animalism and the Varieties of Conjoined Twinning," 285–301.

that of the Hensel twins, in which there is only very limited duplication of organs and all the organs function together as a unit.

#### **2.4.2 McMahan's Dicephalus Argument**

Because Animalists claim that we are identical to human animals, they are committed to the claim that wherever there is one of us, there is precisely one animal identical to this individual, and wherever there is a human animal, there is one and only one of us identical with it. For this reason, McMahan<sup>57</sup> argues, Dicephalus is a counterexample to their theory. Each human person in Dicephalus—each one of us—is related to the animal in the same way; therefore, there is no basis for the claim that one is the animal and the other not. Given the transitivity of the identity relation, both persons cannot be identical to the animal without being identical to each other. Since they are not identical to each other, it seems that neither is identical to the animal. Thus, there are human persons that are not animals, and Animalism is false.

In fact, as McMahan suggests, a weaker argument against Animalism is available. Even the (arbitrary) claim that one dicephalic twin is identical to the animal and the other not implies that at least one human person—one of us—is not a human animal. Thus, a somewhat weaker argument against Animalism can be given:

#### ***The Dicephalus Argument:***

1. In cases of Dicephalus, there are two human people—two of us—but only one human animal.

---

<sup>57</sup> McMahan, *The Ethics of Killing*, 35-39.

Therefore,

2. At least one of us is not a human animal.
3. If Animalism is true, then each of us is a human animal (so, it is not the case that at least one of us is not a human animal).

Therefore,

4. Animalism is false.

In the next two sections (2.4.3 and 2.4.4) I will defend The Dicephalus Argument against existing objections. I will then consider how two other kinds of conjoined twinning raise similar problems for Animalism, and how all of these cases, when taken together, pose a formidable challenge to that view.

### **2.4.3 Blatti and Duplication Objections**

Stephen Blatti views the challenge of dicephalus (or the dicephalus objection, as he refers to it) as a version of what he calls a “duplication objection.” This kind of objection, he writes, “aims to demonstrate that a view—when correctly applied in a particular case—is committed to claiming of one thing that it is identical to each of two or more nonidentical things,” thus violating the transitivity of identity. “Any duplication objection that satisfies this condition has,” according to Blatti, “met its *prima facie* burden.”<sup>58</sup>

Blatti compares the dicephalus objection with an objection to the psychological continuity criterion of personal identity that appeals to cases in which psychological

---

<sup>58</sup> Stephan Blatti, “Animalism, Dicephalus, and Borderline Cases,” *Philosophical Psychology* 20 (2007), 595-597.

continuity takes a branching form. He discusses one such case in which the cerebral hemispheres of a particular person, Peter, are divided and transplanted into the decerebrate heads of two separate bodies. After the transplant, the recipients of Peter's cerebral hemispheres, Righty and Lefty, seem to be two numerically distinct persons, each psychologically continuous with Peter. If psychological continuity is the criterion of diachronic personal identity then Peter should be identical both to Righty and to Lefty. But because Righty and Lefty are not identical to each other, the psychological continuity criterion violates the transitivity of identity. While both the dicephalus objection and the double transplantation objection are construed by Blatti as duplication objections, he claims that only the double transplant objection meets its *prima facie* burden. Only this objection, Blatti argues, appeals to "a possible circumstance in which the psychological criterion would be committed, by its own lights, to a violation of transitivity."<sup>59</sup> The psychological criterion could avoid such a violation only if it were emended somehow, for example through the addition of a 'no branching' clause. But, as Blatti points out, Animalists are not committed by their own lights to an interpretation of dicephalus that violates transitivity. There are, in fact, several possible interpretations of dicephalus. The most natural interpretation is that there are two persons but only one animal. Call this view (a). Animalism would violate transitivity if (a) were true. The Animalist could, however, accept one of the following alternatives: (b) there is only one animal and therefore only one of us—a single person with two minds, or a single mind divided between two brains; (c) there are two of us, and therefore two overlapping or fused animals, each one a separate person; (d) dicephalus is what Blatti calls a "borderline case" in which the dicephalic twins

---

<sup>59</sup> Ibid.



instantiate enough of a certain cluster of properties to qualify as a candidate for being a single human animal, but not enough to qualify as a clear instance of a single human animal because they exhibit some properties that are not characteristic of a single human animal – “notably, the presence of two distinct subjects of experience, as well as the overabundance of various organs and appendages,”<sup>60</sup> Blatti expresses his view of dicephalus by stating that it involves more than one complete human animal, but fewer than two. There is also a fifth position that Blatti does not explicitly distinguish, namely, (e) that the number of human animals and therefore the number of entities of our kind is indeterminate. It is hard to tell from Blatti’s discussion of his view whether (e) and (d) are different views or merely different statements of the same view.

Blatti’s claim that the double transplant objection meets its burden while the dicephalus objection does not is mistaken; both objections fail to meet their respective burdens *as duplication objections*. However, I will argue, this does not diminish the challenge posed by either objection.

That the double transplant objection doesn’t meet its burden is shown by the fact that there are several different interpretations of the double transplant case, none of which is logically incoherent and some of which are compatible with the psychological continuity criterion. (1) The first possible interpretation is that it involves three numerically distinct persons—Peter, Righty, and Lefty—who are related to each other in a way that would violate transitivity if psychological continuity were the criterion of our identity over time. (2) The second is that Peter survives the transplant as a single oddly-shaped person with two disconnected cerebral hemispheres, each one controlling a separate body. The names

---

<sup>60</sup> Ibid.

“Righty” and “Lefty” co-refer to Peter, rather than referring to two separate persons or two disconnected parts of Peter, even though each name is associated with only one of his two bodies, just as the names “Hesperus” and “Phosphorus” are associated with different modes of presentation of the object to which they co-refer. On this view, despite appearances, Righty and Lefty *are* identical, and so there is no violation of transitivity. (3) A third alternative is that double transplant involves not three but two persons, each of whom exists both before and after the transplant. Prior to the transplant, these two temporally extended persons share a single person-segment, just as two overlapping roads might share a single segment of pavement before forking off in separate directions. (4) A fourth alternative is that double transplant is a borderline case in which there are fewer than two, but more than one, person. This parallels Blatti’s interpretation of dicephalus. (5) A fifth alternative (which may or may not be distinct from (4)) would be to claim that there is no determinate fact about how many entities of our kind are present in double transplant. (6) Finally, there is the view that after the transplant Peter is wholly present at two noncontiguous regions of space which appear to be occupied by two nonidentical persons. Peter is wholly present at Righty’s location, and wholly present at Lefty’s location, for example, as he might be wholly present today and wholly present tomorrow. Thus, as in view (2), “Righty”, “Lefty”, and “Peter” are actually three different names for the same person. Blatti<sup>61</sup> considers views (3) and (5) but treats them as ways of revising the psychological continuity criterion only after the double transplant objection has met its burden.

---

<sup>61</sup> Ibid., 598.

Blatti assumes that the psychological continuity criterion is initially committed to (1) and can avoid violating transitivity only if emended by, for example, the addition of a “no-branching” clause. Yet there is nothing about the psychological continuity criterion *per se* that initially commits it to (1). Rather, Blatti appears to assume that the psychological continuity criterion is initially committed to (1) simply because (1) is more plausible than any of the other interpretations. In this respect, the challenge of double transplant is like the challenge of dicephalus. For, although there are several possible characterizations of dicephalus, (a) is the most natural and plausible interpretation, and it is on this interpretation that Animalism violates transitivity. Yet Blatti’s comparison of the dicephalus and double transplant objections is illuminating because it reveals that neither objection is plausibly construed as a duplication objection. Rather, both objections challenge their respective targets by posing a dilemma: either the targeted view violates transitivity in a particular type of case, or its proponents must adopt an interpretation of that type of case that is less plausible than the interpretation on which their view violates transitivity. The challenge of dicephalus successfully undermines Animalism in this regard, since the most straightforward interpretation of dicephalus, (a), is more plausible than the alternative interpretations, (b) through (e).

Consider (b), the view that in dicephalus there is a single person with two minds, or one mind divided between two brains. The problem with this view is that it involves an objectionable distortion of our concept of a person. As Derek Parfit has recognized, (2) fails as a response to the challenge of double transplant for the same reason. In his well-known discussion of fission Parfit writes that

we ought to admit as possible that a person could have a divided mind. If this is possible, each half of my divided mind might control its own body. But though this description of the case cannot be rejected as inconceivable, it involves a great distortion in our concept of a person. ... If a mind was permanently divided, and its halves developed in different ways, it would become less plausible to claim that the case involves only one person.<sup>62</sup>

One way of highlighting the implausibility of views like (2) and (b) is to consider their moral implications. For example, if (b) were correct, the surgical removal and destruction of one head in the case of the Hensel twins would be relevantly like a hemispherectomy in a normal person—a grave diminishment but not as seriously wrong as the killing of a person. But intuitively, the removal of one of the heads in a case of dicephalus would clearly be the killing of one of us, and hence an instance of murder. Call this *The Decapitation Intuition*. Views (d) and (e) also violate The Decapitation Intuition. (I will consider (c) shortly.) If (d) were correct, removal of one head in a case of dicephalus would result either in a less obvious borderline case, or a nonparadigmatic case of one human animal, but not the killing of one of us, a person. And if (e) were correct, it would be neither determinately true nor determinately false that there had been a killing. Views (b), (d), and (e) blatantly deny the existence of a human person.

Animalists might reject my appeal to moral judgments as a way of deciding between the different interpretations of dicephalus. For example, Blatti explicitly warns against the use of such judgments as guides for shaping our metaphysical views about personal identity, suggesting that because our evaluative judgments derive from a broader moral framework, attributing

binding metaphysical significance to these judgments...invites analogous determinations regarding other hard cases—cases where our normative intuitions may be...altogether contrary. Consider...the critic who rejects my

---

<sup>62</sup> Parfit, *Reasons and Persons*, 256.

borderline case view of dicephalus on the grounds that dicephalic twins present exactly two loci of moral status, and the number of such loci correlates with the number of us. Since moral status is typically attributed to an entity in virtue of its psychological capacities (e.g., the capacity to suffer, the capacity for self-consciousness), this critic should also attribute an analogous moral standing to each of the multiple personalities belonging to those who suffer from dissociative identity disorder.<sup>63</sup>

But if an entity has moral status in virtue of its psychological capacities, this fact does not put pressure on Blatti's critic to attribute moral standing to the personalities exhibited by a patient with dissociative identity disorder. To attribute moral standing to an entity's *personality* would involve a rather egregious category mistake. Personalities are not themselves entities with psychological capacities, but are rather sets of dispositions *belonging to* entities with psychological capacities. It is the entity, not the personality, which has moral standing.

Interestingly, our attitudes about what counts as acceptable treatment of dissociative identity disorder tell more against views (b), (d) and (e) than they do against the view espoused by Blatti's critic. If, for example, (b) were the correct interpretation of dicephalus and if the memories, beliefs, and desires encoded in one of the two brains were very different from or in conflict with those encoded in the other, one might expect the death of one of these brains to be beneficial for the double-minded or split-minded person, just as the elimination of certain memories, beliefs, and desires associated with a particular personality is commonly considered beneficial for a patient with dissociative identity disorder. Yet this is clearly implausible. The death of one of the brains in a case of

---

<sup>63</sup> Blatti, "Animalism, Dicephalus, and Borderline Cases," 605.

dicephalus would constitute a loss precisely because it would be the ceasing to exist of a person.

However, I think Blatti may be right about our moral convictions not settling the matter about which interpretation of dicephalus is correct. In fact, for reasons that I will discuss in Chapter 5, it often seems inappropriate to reject a non-moral metaphysical claim about personal ontology on the basis of one's moral convictions about particular hypothetical and actual cases. However, it seems that in arguing against certain interpretations of dicephalus, I have appealed to moral convictions about such cases. Perhaps the moral convictions to which I have appealed in this instance reflect an antecedent metaphysical conviction that each dicephalic twin is a distinct person, and it is this antecedent metaphysical conviction that rebuts the interpretations of dicephalus that I believe we should reject. It is because I believe that the Hensel twins are two persons—two entities of our kind—that I would view the loss of one of their brains as a much more serious matter than the loss of one of the personalities in a case of dissociative identity disorder.

#### **2.4.4 Are there Two Human Animals in Dicephalus?**

I turn now to (c), the view that dicephalic twins are really two fused or overlapping animals. This is the most plausible and the most common response among Animalists to the challenge of dicephalus. Because, for example, the Hensel twins together have more organs than normally form a single complete set, and some of these organs belong to one twin rather than to the other, it is plausible to suppose that each of them is identical to a numerically distinct animal, but that each animal shares a large number of parts with the

other. Matthew Liao points out that “each twin has her own stomach and heart; they have distinct brainstems and distinct spines that are only joined at the hips; and they have partially distinct organs that are united. This suggests that in fact, there are two animals here although they are not fully independent animals.”<sup>64</sup> Like the previous response to the challenge of dicephalus, this one abandons the standard view that animals are individuated by lives in favor of a principle of individuation that focuses on the number of body parts instead. Unlike the previous objection, this one claims that the extent of the duplication in the case of dicephalus is sufficient for their being (determinately) two animals, each one identical to a separate person. Robert George and Patrick Lee share Liao’s view regarding the significance of the duplication of organs; in addition, they claim that each twin has organs that belong only to her, and that the twins’ organs cannot all be parts of a single animal. They argue that if dicephalus involved two persons but only one animal, then none of the organs in the twins could be assigned to one individual rather than the other. Each set of eyes, each set of ears, and so on, would not belong biologically more to one girl than to the other. Each of these organs would have to be a part of a single larger animal, subservient to the survival and functioning of this one animal. But this plainly is not the case. It is indisputable that each one biologically has not only her own brain, but also her own skull, eyes, ears, and many organs, while sharing many other organs.<sup>65</sup>

Yet, assuming there are two separate people in this case, it seems that most of the organs serve *both*. This seems true of all the duplicated organs below the neck, such as the

---

<sup>64</sup> Matthew Liao, “The Organism View Defended,” *The Monist* 89 (2006) 334–350.

<sup>65</sup> Robert George and Patrick Lee, *Body-Self Dualism in Contemporary Ethics and Politics*, (New York, Cambridge University Press, 2008), 47.

esophagi, stomachs, and hearts. McMahan<sup>66</sup> speculates that if the heart on one's side of the body (where "body" refers to the single organic mass and does not presuppose that this mass constitutes only one animal) were to die, the other heart might be sufficient to circulate blood e.g. to the head on the side with the non-functional heart, since there is one continuous circulatory system. Although McMahan's claim is speculative, it receives some support from a more recent case of dicephalus, in which the twins (i.e. each of two separate people) survived birth, despite sharing a single heart which sustained both.<sup>67</sup> If, as this more recent case shows, dicephalic twins can survive with only one heart and circulatory system, then it seems possible that this could be (or could have been) true in the case of the Hensels.

But George and Lee are right that some duplicated organs serve only one of the two persons. Unsurprisingly, these organs are on that person's side of the body (where, again, "body" refers to the single organic mass and does not presuppose that this mass constitutes only one animal). Thus, we say Abigail's eyes "belong" to *her* because they are the ones *she* sees with. That is exactly what one would expect when two persons are sustained by a single animal. One person's eyes transmit signals that are received only by her and not by the other person, but these eyes are still part of the animal that sustains both siblings; both sets of eyes are "caught up" in the animal's life-sustaining processes.

## 2.5 Cephalopagus

---

<sup>66</sup> McMahan, *The Ethics of Killing*.

<sup>67</sup> John Talty, "Two-headed reminds doctors of Abigail, Brittany Hensel," *International Business Times*, December 2011.



There is another form of conjoined twinning that raises trouble for Animalism: Cephalopagus. A brief and perhaps tendentious description of this phenomenon is that there is one head with two bodies—the reverse of dicephalus. In one case, reported in 2008, a single head contained a single brain but two cerebella and two brain stems. In addition to four arms and four legs, there were two spinal cords, two hearts, four lungs, two livers, two spleens, four kidneys, and so on. If the duplication of organs indicates the presence of two animals, there would seem to have been two in this case. There was, however, only one esophagus and one stomach, and the brain was formed from four fused cerebral hemispheres (so that there was also fusion of the faces) and so may not have been capable of supporting consciousness.<sup>68</sup>

In another case, there was a single head containing a single brain composed of two cerebral hemispheres, but two cerebella and two brain stems. Although there were two hearts, two spinal cords, two spleens, and two bladders, there were only two lungs, and only one esophagus, one stomach, one liver, and one pancreas.<sup>69</sup>

In both cases, there is no mention of whether there were signs of consciousness. So, neither of these actual cases is a clear instance of one subject of consciousness resident in the area of overlap between two distinct human animals. But these cases occupy positions on a continuum of actual and possible cases—a continuum that includes more possible than actual cases because cephalopagus is extremely rare. If the phenomenon

---

<sup>68</sup> Erin Digitale, “Conjoined Twin Toddlers Successfully Separated at Packard Children’s,” Stanford University, <http://med.stanford.edu/news/all-news/2011/11/conjoined-twin-toddlers-successfully-separated-at-packard-childrens.html>

<sup>69</sup> M. Hovorakova, Z. Likovsky, and M. Peterka, “A Case of Conjoined Twins: Cephalothoracopagus Janiceps Disymmetros,” *Reproductive Toxicology* 26 (2008), 178-82.

were more common, there might by now have been an actual case that, like the second case we cite, had a single normally formed brain with two cerebella and two brain stems, as well as a single face, mouth, and throat, but that, unlike that case, had two esophagi diverging from the single throat, each leading to a different stomach, as well as the normal complement of other organs and appendages in each half of the total bodily mass below the neck.

### **2.5.1 Considering Cephalopagus and Dicephalus Together**

If there were such a case, and if it, or they, were to survive long enough to experience not only consciousness but also self-consciousness, then it seems that this would be a clear case in which there would be only one self-conscious mind but two human animals. There is, as far as I can tell, no reason to suppose that such a case is physiologically impossible.

That there would be two animals is suggested by the fact that surgical “separation” could result in two clearly distinct animals, each of which, one could plausibly assume, existed prior to the beginning of the separation procedure. Such an operation may be possible even with technologies that already exist or will exist soon. The separation might be asymmetrical, with the cranium, the brain, and one each of the cerebella and brain stems going one way, and the cerebellum and brain stem going the other way. Plausibly, such asymmetrical separation could result in two self-sustaining animals: one relevantly like a normal person, the other, without a brain, relevantly like a patient in a persistent vegetative state whose brain had been destroyed but who could remain biologically alive with little external support other than nutrition and hydration. Alternatively, the cephalopagus twins might be divided symmetrically, with each taking part of the cranium, one cerebral

hemisphere, one cerebellum, and one brain stem. Each would then be relevantly like a patient who has received a hemispherectomy, although each would require a partial artificial cranium.

This case of cephalopagus combines with dicephalus to present a serious theoretical challenge to Animalism. Earlier we saw that there are several options available to Animalists with regard to dicephalus. Of these, the one that is clearly most plausible is to accept that because there are various duplicate organs and two wholly independent brain stems, there are two overlapping animals and therefore two of us. Next recall what options are available to animalists with regard to our hypothetical case of cephalopagus. They can accept that because there would be only one cerebrum, there would be only one mind and therefore only one person, or one of us. That would oblige them to say that there would be only one animal. But that seems impossible to reconcile with their most plausible interpretation of dicephalus, namely, that it involves two overlapping animals. Every reason that animalists have cited in favor of the view that there are two overlapping animals in a case of dicephalus—that there is considerable duplication of organs, that there are two functional brain stems, etc.—counts equally or even more strongly in favor of the view that there are two overlapping animals in a case of cephalopagus. There is apparently no principled reason for treating dicephalus as two animals but cephalopagus as one. If, therefore, animalists accept that there is only one person, and thus only one animal in cephalopagus, it seems they must accept that there can be only one animal, and thus only one person, in dicephalus. It is, however, very difficult to believe that the Hensel twins are only one person. So it seems that animalists must accept that there are two animals in cephalopagus.

That is highly plausible. What is implausible is what that view entails according to Animalism, namely, that in our hypothetical case of cephalopagus there are two separate persons—two of us. If the cerebrum has matured to a point at which the mind it sustains is fully self-conscious, there is undeniably a person present, someone who, with his single mouth, could engage with us in rational discourse while punctuating his assertions with multibrachiate gestures. But it is highly implausible to suppose that there are two persons present in this case. There is only one cerebrum—one consciousness-generating entity—and therefore a single unified mind, exactly as in the case of any ordinary person. How many limbs or organs there are below that single center of conscious experience seems irrelevant to how many persons, or individuals of our sort, there are.

Perhaps animalists could claim that in the case of cephalopagus although there are two animals and therefore two of us, only one of the two entities of our kind is a person *in the Lockean sense*—an entity that possesses reason and reflection and that thinks of itself as remaining one and the same entity over time. This is a viable response because according to Animalism we are not essentially persons in the Lockean sense; “Lockean personhood” is just a phase through which most of us contingently pass. But the problem with this response is that both animals are related to the cerebrum in exactly the same way. Anything that might give one animal a claim to be a person in the Lockean sense would be equally true of the other animal. There is no good reason to suppose that one animal rather than the other is the person. Animalists might try to argue that it is indeterminate which of the two animals the person is, but this is a strategy of desperation with little credibility.

### **2.5.2 Cephalopagus and the Too Many Subjects Problem**

Suppose the Animalist accepts that there are two of us in cephalopagus, and indeed two persons if the cerebrum has achieved self-consciousness, despite the fact that there is only one center of consciousness, just as there is in any ordinary person. Then for every conscious experience generated by the single cerebrum, Animalists must say there are two subjects. This gives rise to a version of a problem that some believe undermines most versions of the PA; I call it *the too many subjects problem*. I will discuss this problem further in Chapter 3. I mention it here in order to give the reader a glimpse of what is to come.

## 2.6 Conclusion and Summary of Chapter 2

The aim of this chapter is to reinforce and expand upon the two most powerful arguments against Animalism—The Transplant Argument and The Dicephalus Argument. In Section 2.1, I considered *The Transplant Argument*, which appeals to a hypothetical case—*the transplant case*—in which my brain is extracted from the rest of my body and transplanted into the decerebrate head of a different human body, while consciousness is maintained throughout the transplant operation. In this case, it seems, I survive the operation as the person who inherits my brain and psychology. If this is correct, then I cannot be identical to the animal, and so Animalism is false. After introducing the Transplant Argument, I considered the different responses to the argument that Animalists have proposed. These include denying that the brain transplant would preserve the right kind of psychological continuity, appealing to the view that identity is not what matters, and appealing to the possibility of ‘remote’ consciousness and thought. I argued that each of these responses fails to undermine the initial force of the Transplant Argument. To

reinforce this conclusion, I appealed to The Severed Head Case, in which a human person is apparently separated from a human animal and kept alive in isolation. I argued that while those who defend a Brain-based Approach can offer a plausible interpretation of this case, Animalists cannot.

Next, in Section 2.4, I considered another powerful argument against Animalism—*The Dicephalus Argument*. This argument appeals to both actual and hypothetical cases of dicephalus in which there appear to be two human persons—two of us—but only one human animal. Since the persons outnumber the animals, it seems at least one human person—one of us—is not a human animal, and so Animalism is false. In Sections 2.4.1 and 2.4.2, I considered some objections to the Dicephalus Argument, including the objection that the duplication of organs in a case of dicephalus indicates the presence of two distinct animals. In response to these objections, I appealed to a different case of conjoined twins—cephalopagus—in which there appear to be two human animals but only one person. I argued that those who appeal to the duplication of organs in order to defend the conclusion that there are two animals in the case of dicephalus have no principled way of avoiding the conclusion that there are two animals in the case of cephalopagus, since the duplication of organs is even more extensive in that case. However, I argued, the claim that there are two animals, when combined with Animalism, apparently leads to the conclusion that there are two human persons, though it seems clear that in the case of cephalopagus, there is only one person. I emphasized that for Animalists the only way to maintain that there is one person in the case of cephalopagus is to claim that there is only one human animal in that case. However, I claimed, if Animalists maintain that there is only one human animal in the case of cephalopagus, despite the extensive duplication of

organs in that case, then they will not be able to cite extensive duplication of organs as a reason for thinking that there are two animals in the case of dicephalus. In fact, it is difficult to identify any consideration that counts in favor of treating cephalopagus twins as a single animal that does not also count in favor of treating dicephalus twins as a single animal. However, as I emphasized, if Animalists insist that there is just one animal in the case of dicephalus, then they will likely be forced to say that there is only one *person* in that case, which is very difficult to believe.

In conclusion, the Transplant and Dicephalus Arguments pose huge problems for Animalism. Animalists seem unable to explain away or accommodate the intuitions that underlie these arguments. Thus, in order to successfully defend their approach against its broadly-categorized rival, the PA, they must show that the theoretical virtues of Animalism (relative to the PA) outweigh the heavy intuitive costs that I have emphasized in this chapter. In the next chapter, I consider whether they can show this.

### 3. Do the Virtues of Animalism Outweigh Those of the Psychological Approach?

#### 3.1 Introduction to Chapter 3

In Chapter 2, I introduced, defended, and offered weakened versions of the two most convincing arguments against Animalism and in favor of (some version of) The Psychological Approach. In this Chapter, I will consider the alleged advantages of Animalism over The Psychological Approach, and whether in fact Animalism has such advantages.

The main claim of this chapter is

**Main Claim:** For each Animalist theory, A, that has been proposed, there is a corresponding version of The Psychological Approach, P, such that A is no better than P with respect to those theoretical considerations which, Animalists claim, give their theory an advantage over The Psychological Approach.

I will begin in Section 3.1 by listing the four major considerations with respect to which Animalism is supposed to be a better theory than The Psychological Approach. In Sections 3.2 through 3.5, I will explain how, with respect to the first of these four considerations, Animalism is no better (that is, has no more theoretical plausibility) than The Psychological Approach. Specifically, I will show that Animalism and The Psychological Approach both face a common problem, to which proponents of each view may offer different solutions. As I show in Sections 3.3 through 3.5, the different solutions that Animalists have proposed lead to importantly different versions of their theory. I will argue that for each of these different versions of Animalism, there is a corresponding



version of The Psychological Approach such that the former is no better than the latter with respect to the four major considerations that allegedly make Animalism better than The Psychological Approach. In Section 3.7, I consider hybrid views and hybrid theories. In Section 3.8, I return to the example of conjoined cephalopagus twins introduced at the end of the previous chapter, in order to make a final point regarding the comparative plausibility of Animalism and The Psychological Approach. In Section 3.9, I summarize the main points and arguments of the chapter, and discuss their importance for the central argument of the dissertation.

### **3.2 The Four Major Theoretical Considerations that Allegedly Favor Animalism over the Psychological Approach**

Animalism is believed to be better than The Psychological Approach with respect to four important theoretical considerations. Here, I will lay out these four considerations and, for each consideration, I will explain why many believe Animalism to be better than The Psychological Approach with respect to it.

#### **3.2.1 The Too Many Subjects Problem**

According to The Psychological Approach, I began to exist no earlier than when the brain of the fetus that preceded me first became complex enough to support or generate consciousness. Defenders of The Psychological Approach typically deny that the fetus ceased to exist the moment I began to exist. They think it is absurd to suppose that the onset of consciousness destroyed the fetus. Rather, they claim, the fetus persisted through

the moment at which I began to exist, and, from that moment on, it co-existed with me, and continued to grow and develop into the adult animal that is currently sitting in my chair. However if, as The Psychological Approach states, I am a psychological continuer and not an animal, it follows that I am not the animal sitting in my chair. And this seems to imply that there must now be two entities sitting in my chair: me and the animal. Olson calls this *the fetus problem*. A similar problem arises when we consider what would happen if I were to irreversibly lose the neural capacity for consciousness. If this were to happen, then the human animal that is now sitting in my chair might persist in a permanent vegetative state. But according to the Psychological Approach, I would cease to exist. Again, it seems, we are led to conclude that there are at least two person-like entities sitting in my chair. We might call this *the vegetable problem*.

If I am not identical to this animal, what is the relation between me and it? As we saw in Chapter 1, there are several possibilities. Constitutionists who defend The Psychological Approach claim that we (psychological continuers) are constituted by animals; we coincide with animals as clay statues coincide with lumps of clay. On what Parfit<sup>70</sup> calls the Embodied Parts View (EPV), we are spatial parts of animals, such as brains, cerebrums, or perhaps whole nervous systems. On the Temporal Parts View (TPV), each of us is a temporal part of an animal—one that extends from the first time at which the animal's brain is complex enough to support consciousness to the time at which the animal's brain loses this complexity. Suppose that, as Constitutionists claim, I am spatially and materially coincident with the animal, that is, located in the same region of

---

<sup>70</sup> Parfit, "We Are Not Human Beings," 5-28.

space and constituted by the exact same matter. Then, if I am conscious and my mental properties supervene on certain of my physical properties, the animal must be conscious as well since it has all of my physical properties. But that is one subject of experience too many. It is absurd to suppose that there are really two subjects of each individual thought or experience occurring in my head.<sup>71</sup> This ‘too many subjects’ problem seems to arise for views which imply that the animal and I are not identical but have the same brain or the same consciousness-generating parts. Thus, the Embodied Parts View and the Temporal Parts View seem vulnerable to the problem even though, according to these views, I am not coincident with the animal. (I will say more about the too many subjects problem and the potential solutions to it in Section 3.3.)

### 3.2.2 Simplicity

A common assumption among philosophers is that, all else equal, one theory is better than another if it is simpler. Although it isn’t always obvious what makes a theory simple or complex, it is generally agreed that if one theory posits *fewer things* than another, then, all else equal, the former is simpler than the latter. Like Animalism, most versions of The Psychological Approach recognize the existence of animals. However, unlike Animalism, The Psychological Approach implies that we are not animals but psychological continuers. Thus, most versions of The Psychological Approach recognize the existence of both psychological continuers and animals. For this reason, other things being equal, these versions of The Psychological Approach are less simple than their Animalist rivals. They posit more entities and more kinds of entity.

---

<sup>71</sup> Olson, *What Are We?* 29-30.

### 3.2.3 Consistency with Common Sense

Animalists as well as many other philosophers assume that *all else equal* if one theory is more consistent with common sense than another theory, then the former theory is better. Animalism initially appears to be consistent with common sense to a greater extent than The Psychological Approach. Unlike The Psychological Approach, Animalism implies that each of us was once an embryo and that each of us could end up in a permanent vegetative state. Appealing to common sense, many non-philosophers would readily concede this. Moreover, we *seem* to be animals. Most biologists would affirm that we are animals of the species *Homo sapiens*. Strictly and literally speaking, The Psychological Approach is incompatible with these claims.

### 3.2.4 Composition and Unity

All else equal, one theory of personal ontology is more plausible than another if the former but not the latter can help explain, or is at least compatible with, the most plausible answer to what Peter van Inwagen calls *The Special Composition Question*: Under what conditions do things add up to or compose something? As van Inwagen puts it, “When does unity arise out of plurality?”<sup>72</sup>

---

<sup>72</sup> Van Inwagen, *Material Beings*, 31.

This consideration is the most obscure of the four. Regarding the first three considerations, one can immediately see not only how it could plausibly be considered a good-making feature of a theory, but also why Animalism at least appears to be better than The Psychological Approach with respect to it. That is not true of the fourth and final consideration. As the argument of this chapter proceeds, we will see why this consideration may be theoretically important, and whether Animalism is in fact better than The Psychological Approach with respect to it. I will say much more about these issues in Section 3.3.

### **3.3 The Psychological Theorist's Response to the Too Many Subjects Problem and why Animalism is not Immune to that Problem**

Let us start by considering the too many subjects problem. If I am not identical to the animal sitting in my chair, then there are at least two subjects of consciousness now sitting in my chair. One worry about the too many subjects problem is that it seems to introduce a 'too many persons' problem; if the animal experiences, thinks, and acts as I do, then it should be a person. (Here the term 'person' is used not as a placeholder for 'something that is one of us' but to refer to beings with robust psychologies.) If this is right, then persons come in at least two varieties: those that are animals, and those that are psychological continuers.

#### **3.3.1 Animalism and the Corpse Problem**

Olson<sup>73</sup> notes that his particular version of Animalism might appear to face the too-many-subjects problem. If you are an animal, it seems you could exist as a corpse after your death. But some Animalists, such as Olson and van Inwagen<sup>74</sup> deny that one survives death in this manner. They hold that animals have among their essential properties a kind of functional unity of their various parts that is lost at death. So they have claimed that the animal ceases to exist at death. Following Olson, I will call this version of Animalism according to which the animal is alive essentially *The Biological Approach*. The Biological Approach stands in contrast to what I will call *The Bodily Approach*, according to which the animal is alive only contingently, and hence, could continue to exist as a corpse. According to defenders of The Biological Approach, we are *essentially living animals*—animals that cannot exist without being alive in the biological sense.

If, as defenders of The Biological Approach claim, an animal ceases to exist when it dies, then, provided that its death is not the result of a violent explosion, there will seem to be something that *persists through* its death—something that is alive before the moment of death, and dead afterward. Let us call it *a body*. It seems obvious that there is a body associated with each one of us, and that this body will probably eventually become a corpse. For this reason, The Biological Approach faces something akin to *the vegetable problem* that arises for The Psychological Approach. Olson calls this particular problem *the corpse problem*. If the body and the essentially living animal are coincident before the latter ceases to exist, then there are two coincident entities where I am now sitting: the essentially living animal (which, according to The Biological Approach, is what *I* am) and

---

<sup>73</sup> Eric Olson, "Animalism and the Corpse Problem." *Australasian Journal of Philosophy* 82 no.2 (2004), 265-274.

<sup>74</sup> Ibid; van Inwagen, *Material Beings*, 81-97.

the body that would become my corpse if I were to die peacefully of natural causes. Since the essentially living animal is conscious, it seems the body is also conscious, since both are made of exactly the same matter. Too many subjects!

We have seen that The Biological Approach and the Psychological Approach appear to face a common problem. Olson and other animalists have argued that the solutions available to proponents of the Biological Approach for solving their version of the problem are different from those available to proponents of the Psychological Approach for solving their version of that problem. I will argue that this claim is false. The kinds of solutions available to proponents of each view are exactly the same.

### **3.3.2 The Derivatively Conscious Animal View**

The most common response to the too many subjects problem on behalf of the Psychological Approach appeals to the idea that the animal that is now sitting in my chair is conscious *only derivatively*—only in virtue of standing in a certain relation of dependence to a psychological continuer who is conscious independently or *nonderivatively*. Let us call this *the Derivatively Conscious Animal View* (DCAV).

#### **3.3.2.1 The Derivatively Conscious Animal View and Constitutionalism**

DCAV can be combined with different views about the relationship between the person and the animal. For example, Constitutionalists can claim that the animal is conscious derivatively, in virtue of its constituting the psychological continuer who is conscious

nonderderivatively. Lynne Rudder Baker defends such a view.<sup>75</sup> She defines ‘x is constitutionally related to y’ as ‘either x constitutes or x is constituted by y’. She then claims that

[t]he basic idea of having a property derivatively is [that] x has property H at time t derivatively if and only if x’s having H at t depends wholly on x’s being constitutionally related to something that has H at t independently of its being constitutionally related to x. If x constitutes y then x and y share many properties: The [psychological continuer] is over six feet tall and so is his body. But since the [psychological continuer] derives the property of being over six feet tall from the body that constitutes him, he has that property derivatively.<sup>76</sup>

Given Baker’s characterization of x’s having property H derivatively, a good test for determining whether x has H derivatively or nonderivatively is to determine whether x can have H without being constitutionally related to something else that has H. If it can, then x has H nonderivatively; if not, then assuming x has H at all, x has H only derivatively. For example, one reason for thinking that the constituted person has the property of being six feet tall derivatively, while the constituting body has that same property nonderivatively, is that the person cannot have the property of being six feet tall unless she is constituted by the body, whereas the body can have the property of being six feet tall without constituting the person. For example, the body might continue to be six feet tall after being separated from the person; this would happen if the person’s brain and psychology were transplanted into a numerically different body. The person depends on the body (or some body) for her being six feet tall, but the body does not depend on the person (or any person) for its being six feet tall.

---

<sup>75</sup> Baker, *Persons and Bodies: A Constitution View*; Lynn Rudder Baker, *The Metaphysics of Everyday Life* (Cambridge, MA: Cambridge University Press, 2007).

<sup>76</sup> Baker, *Persons and Bodies: A Constitution View*, 13.



Similarly, on Baker's view, the body cannot have the property of being self-conscious without constituting the person, but the person can have this property without being constituted by the body. Once again, the asymmetry of dependence (i.e. that of the body on the person) suggests that the body has the property of being self-conscious derivatively, whereas the person has this property nonderivatively, for example, if the person's brain were removed and sustained in a vat.

According to Baker, the fact that the animal has the property of being conscious derivatively while the person has this same property nonderivatively illuminates the sense in which there is only one person and one conscious being where I am now sitting. Baker asks rhetorically, "If I'm a person and my body is a person, why aren't there two persons here?" "The answer," she claims, "is that there is one instantiation of personhood—I am a person nonderivatively; my body, in virtue of constituting me is a person derivatively."<sup>77</sup> Thus, according to Baker, the metaphysically important sense in which there is only one person where I am now sitting is that there is only one instantiation of the property that I have of being a person. Since the person and the body *coinstantiate* the property of being a person, there is a sense in which they count as one person rather than two.

### **3.3.2.2 The Derivatively Conscious Animal View and the Embodied Parts View**

Some defenders of the Embodied Parts View (EPV) accept the Derivatively Conscious Animal View (DCAV) as well.<sup>78</sup> They claim that the animal is conscious derivatively in virtue of its having the person as a part, while the person is conscious nonderivatively, and

---

<sup>77</sup> Ibid.

<sup>78</sup> Parfit, "We Are Not Human Beings," 5-28; McMahan, *The Ethics of Killing*; Persson, "Our Identity and the Separability of Person and Organisms," 519.

hence, independent of its being a part of the animal. The plausibility of DCAV on this view derives from the apparent fact that the person could be conscious without being a part of the animal, while the animal could never be conscious without having the person, or some person, as a part.

Jeff McMahan, who defends both EPV and DCAV, illuminates the sense in which the animal derives the mental properties of the person by pointing to uncontroversial examples of a whole deriving certain properties from one of its parts:

Suppose ... that over a certain period of time the only part of a tree that grows is a particular limb. ... A property of the part—growth—is in this instance necessarily a property of the whole. There are thus two things that are growing: the limb and the tree of which it is a part. Similarly, when I blow the horn in my car, the horn makes a noise but so does the car. There are two things that have the property of emitting a noise: the horn and the car of which it is a part. ... There is only one noise; and there is a clear sense in which there is only one noisemaker: the horn. But we attribute the making of the noise not just to the horn but also, in a derivative way, to the larger whole [of which it is part]. It is clear, however, that the car is not some additional occult presence that mysteriously joins the horn in producing the honking noise. Nor is the animal as a whole involved in the experience of consciousness except by [having as a part] that which is conscious. In the same sense in which the tree grows because its limb does, and in which the car honks because its horn does, the animal may be said to think, feel, and perceive, because I do.<sup>79</sup>

According to McMahan, these examples help to clarify the sense in which, on his view, there are two conscious entities where I am. The animal is conscious only in the sense that one of its parts is conscious.

The too many subjects problem may seem much less problematic if there is at most one entity that is nonderivatively the subject of my consciousness. Thus, EPV appears to partially solve the too many subjects problem by positing a distinction which, once recognized by common sense, makes the problem seem less bad. Once the relevant

---

<sup>79</sup> McMahan, *The Ethics of Killing*, 92 – 93.

distinction is recognized, it may no longer seem objectionable that there are two subjects of my consciousness, provided that only one of these subjects (me) is conscious in a nonderivative sense.

### **3.4 The First Response to the Too Many Subjects Problem on Behalf of the Biological Approach: Eliminativism**

Inspired by Constitutionlists such as Baker, a Defender of The Biological Approach might claim that the essentially living animal coincides with and is constituted by the body during those times at which the former and the latter are both biologically alive. Next, this “Constitutionalist Animalist” might claim that the person is the only nonderivatively conscious entity, while the body is conscious only derivatively, in virtue of its being constitutionally related to the essentially living animal. With respect to the too many subjects problem, the Defender of The Biological Approach could claim, following Baker, that there is an important sense in which there is only one person now sitting in my chair, even though the essentially living animal that has the property of being a person and the body that has this same property (albeit derivatively) are not identical.

Alternatively, the defender of The Biological Approach could adopt the Temporal Parts View. She could claim that the person is an essentially living animal, which is a temporal part of a body. She could then claim that the body is conscious in virtue of its having the conscious essentially living animal as a temporal part. This version of Animalism parallels those versions of The Psychological Approach on which the person is a psychological continuer that is a temporal part of the animal.

However, perhaps understandably, Defenders of The Biological Approach reject such views. This is mainly because they are unconvinced that the strategy of treating one entity as the only nonderivatively conscious subject, while recognizing other person-like entities as conscious only derivatively, solves the too many subjects problem. In their view, the conclusion that there is more than one conscious subject where I am sitting is highly problematic, even if one of these subjects is conscious in a special sense. Thus, Defenders of The Biological Approach have felt compelled to offer a different kind of response to the too many subjects problem.

Olson and van Inwagen<sup>80</sup> respond by claiming that there simply are no bodies and hence no corpses. When an animal dies it ceases to exist. At the moment of death, assuming that the atoms of which the animal was previously composed continue to exist (an assumption that is highly questionable) these atoms remain in the shape of that animal but they do not compose anything. For example, suppose that at time  $t$  the animal dies of natural causes and therefore ceases to exist, and suppose that at  $t$  there are exactly  $n$  atoms within the space formerly occupied by the animal. Then, at  $t$ , there are exactly  $n$  things within this space. There are exactly  $n$  atoms and that's all; there is nothing that these  $n$  atoms compose. If there were something that they composed then, at  $t$ , the number of things within the space formerly occupied by the animal would be at least  $n + 1$ —not  $n$ . There would be  $n$  atoms plus the object that these atoms composed.

Not only is Olson inclined to deny the existence of bodies and mere hunks of matter, but he has also argued against the existence of large bodily parts such as brains, heads, and

---

<sup>80</sup> Eric Olson, "Replies," *Abstracta* 4 (Special Issue I)(2008), 32-42; van Inwagen, *Material Beings*.

upper halves. If animals had such large parts, Olson reasons, then every part of an animal that includes its brain would be able to think. According to Olson, this would be highly problematic for Animalism.

[I]t would make it hard to see how we could ever know that we are animals. If you think you're an animal, then your head, which thinks just as you do, ought to think, mistakenly but on the same grounds, that it is an animal. So for all you know, you might be your head. Why suppose, then, that you are an animal, rather than a head or a brain or some other thinking part of an animal?<sup>81</sup>

So Olson's strategy for avoiding the too many subjects problem is to eliminate (i.e. deny the existence of) (i) anything which, if it existed, would be coincident with the animal, and (ii) any part of the animal larger than an atom, including any part that would be able to think, feel, and perceive if and whenever the animal thought, felt, and perceived. I will call this strategy for avoiding the too many subjects problem Eliminativism. Defenders of The Bodily Approach like Feldman, who claim that an animal is a body, can adopt Eliminativism. That is, they can deny the existence of everything that is a possible (or plausible) candidate for being the subject of one's conscious experiences other than the body. This includes denying the existence of mere lumps of human matter, i.e., lumps of human matter that could exist even after the bodies that they apparently constitute have ceased to exist.

### 3.4.1 Eliminativism and the Special Composition Question

Eliminativism raises a question: When do some things, call them the xs, compose a further thing, y? This is the special composition question.<sup>82</sup> Many of us believe both that there

---

<sup>81</sup> Olson, *What Are We?* 216 – 219.

<sup>82</sup> van Inwagen, *Material Beings*, 22.

are at least some composite objects and that the conditions under which composition occurs are restricted. For example, while many of us believe that there are such things as hands and feet, with parts such as fingers and toes, we are inclined to deny that there is any such thing as the object composed of my left foot and your right hand. Thus, many of us are inclined to deny that for any two or more objects, there is a further object that they compose. Those who adopt Eliminativism, or Eliminativists, also believe that there are at least some composite objects. However, they believe that the conditions under which composition occurs are much more restricted than many of us are inclined to believe. Just as many of us would say that there is no object composed of my left foot and your right hand, eliminativists say that atoms do not compose anything when they are arranged in the shape of a body, or hunk of matter, or anything other than a living animal.

If Eliminativist Animalists are correct, then there is only one entity now sitting in my chair and thinking my thoughts—an animal that is essentially alive. There is no psychological continuer, no human body, and no mere hunk of matter coincident with the animal. The animal is the only person-like entity at its location. Hence, there is no too many subjects problem or any of the usual problems that arise for those who posit spatially coincident entities (or entities that share the same consciousness-generating parts).

### **3.4.2 Eliminativism and The Psychological Approach: How the Biological Approach Loses the Advantage of Consistency with Common Sense**

Defenders of The Psychological Approach can also appeal to Eliminativism. Whereas Animalists who accept Eliminativism deny the existence of every composite entity other

---

than the essentially living animal that is a plausible candidate for being me, defenders of The Psychological Approach who accept Eliminativism deny the existence of every such candidate other than the psychological continuer. For example, suppose that one's cerebral hemispheres irreversibly lose the capacity to generate consciousness but that the rest of one's body remains alive and functional while hooked up to a ventilator. The defender of The Psychological Approach who accepts Eliminativism will claim that in this case the atoms that previously composed a psychological continuer now compose nothing at all; she will say that there is no living animal hooked up to a ventilator, only atoms arranged in the shape of a living animal. If, as Olson is now inclined to believe, particles arranged in the shape of a body fail to compose a body (or anything else), why can't particles arranged in the shape of an essentially living animal fail to compose an essentially living animal?

Why spare essentially living animals from elimination? Why not spare psychological continuers or bodies? If there is a good answer to this question, it cannot be that common sense recognizes essentially living animals but does not recognize any other kind of composite entity. Common sense recognizes many different kinds of composite entities—heads and hands, tables and chairs, planets, lumps of matter, corpses, and many others. Eliminativist Animalism rules out the existence of almost every kind of thing that common sense recognizes. It blatantly defies common sense. If, in order to avoid the too many subjects problem, Animalism must blatantly defy common sense, then Animalists cannot justifiably appeal to consistency with common sense as a consideration in favor of accepting their view instead of The Psychological Approach. (I will say more in defense of this claim in Section 3.3.4.)

Another possible answer is suggested by Olson:

...[F]ew opponents of animalism deny the existence of animals. They have good reason not to: anything that would rule out the existence of animals would also rule out the existence of most of the things we might be if we weren't animals. If there are no animals then there are no beings constituted by animals, for instance, and no temporal or spatial parts of animals.<sup>83</sup>

But Olson's main claim in this passage is clearly false. As we just saw, defenders of The Psychological Approach can accept a version of Eliminativism that rules out the existence of animals but doesn't rule out the existence of psychological continuers. In support of his main claim, Olson makes the point that if there are no animals then there are no beings constituted by animals and no temporal or spatial parts of animals. This point is correct, but it does not support the conclusion that anything that would rule out the existence of animals would rule out the existence of most of the other things we might be. Of course, if there are no animals then there are no beings constituted by animals, and no beings that are temporal or spatial parts of animals. For this reason, if a defender of The Psychological Approach were to claim that there are no animals, she would be forced to admit that we (psychological continuers) are not constituted by animals, and that we are not temporal or spatial parts of animals. She would not be forced to admit that there are no psychological continuers.

### **3.4.3 Can The Biological Approach Offer a More Plausible View of Composition and Unity?**

---

<sup>83</sup> Olson, *What Are We?* 30 – 31.



Olson thinks that animals should be spared from elimination because he believes, following van Inwagen,<sup>84</sup> that the most plausible answer to the special composition question, “For any given objects, when do these objects compose something?” is: “Just when the activities of these objects jointly constitute a biological life.” Thus, Olson thinks, it is harder to deny the existence of essentially living animals than to deny the existence of any other composite object. In defense of this claim, Olson appeals to the ontological significance of the functional unity among the parts that compose an animal:

I claim that if there are any composite objects, there are animals. The particles that make up a live cat are unified if any particles are... If you don’t believe there are animals, you might as well say there are no composite objects at all. So I feel confident that there are animals if there are any composite objects. I’m a lot less confident about the existence of any of the rival candidates for being me.<sup>85</sup>

I think a defender of The Psychological Approach could reasonably respond that if there is any sort of activity that can be fully explained only by positing a composite object that engages in that activity, mental activity of some kind is the most plausible candidate. She could argue that the complexity of the unity among particles that together generate this mental activity is substantially greater than that among particles whose interactions constitute the life processes of what we take to be animals. If there are any composite objects, she might say, there are psychological continuers. If you do not believe there are psychological continuers, you might as well say there are no composite objects at all; we should be more confident in the existence of psychological continuers than we are in the existence of essentially living animals. This challenge seems especially forceful if one assumes that the kind of psychological continuity that is relevant for one’s identity over

---

<sup>84</sup> van Inwagen, *Material Beings*, 82.

<sup>85</sup> Olson, “Replies,” *Abstracta* 4, 41.

time involves phenomenal continuity, or the capacity for such continuity. The claim that there could be no phenomenal consciousness without an accompanying subject of that consciousness is plausible, though controversial. It is much harder to believe that we must posit an essentially living animal in order to explain the complex biological processes that Olson and DeGrazia find so ontologically impressive.

Perhaps Olson rejects non-Animalist answers to the special composition question because these answers would make us too small. He considers the following example (which no one has yet defended).<sup>86</sup>

**Subject Minimalism:** The *xs* compose a *y* at time *t* if and only if each of the *xs* exists at *t*, and there is a conscious experience, *E*, which occurs at *t*, and there is no conscious experience of which *E* is a part, and, at *t*, each of the *xs* is directly involved in producing *E*.

*E* is what I will call a maximal conscious experience—a conscious experience which isn't a part of any other conscious experience. (I assume that two distinct conscious experiences, each belonging to one of two distinct subjects, are not parts of any single conscious experience.)

One obvious objection to Subject Minimalism (Minimalism, for short) is that it implies that we are much smaller than we seem to be. None of us is six feet in height and none of us weighs 200 lbs. Instead, each of us is small enough that she could fit inside the skull of an animal (if there were any animals). In this regard, Minimalism is similar to certain versions of the Embodied Parts View we considered earlier.

---

<sup>86</sup> Olson, *What Are We?*

A more serious worry for Minimalism that Olson raises concerns the notion of something's being directly involved in producing consciousness. As Olson points out, even if one knew everything there was to know about the properties and activities of atoms, it would be hard for one to say which atoms were directly involved in producing consciousness and which were involved only indirectly. Consider some person S, who is now conscious. Presumably, one wants to say that S's nerve cells are among those parts of S that are now directly involved in her consciousness. Yet, as Olson points out, even this seems doubtful.

No one knows exactly how nerve cells produce thought and experience, but it appears to have something to do with electrical and chemical signals that they store and communicate to other nerve cells. And a good deal of what goes on within a nerve cell appears to be no more directly involved in this storage and transmission of information than the activities of the heart and other vital organs are. Many parts of a nerve cell are involved in acquiring nutrients, or in expelling waste, or in maintenance and repair, or in maintaining the cell's boundary. They don't seem to be directly involved in whatever it is that gives rise to [consciousness].<sup>87</sup>

Olson goes on to provide a sustained critique of the notion of something's being directly involved in consciousness.

If neither the whole of the brain nor even the whole of any nerve cell within the brain is directly involved in [consciousness], what is directly involved? This question gets harder if we think about how to distinguish in a principled way between direct involvement in a being's [consciousness] and indirect involvement. The point has nothing to do with [consciousness] in particular. Imagine a factory that makes knives—an old-fashioned factory where the work is done mainly by hand, with no robots. All the factory workers are involved in some way in the manufacture of knives. Some deliver the steel; others beat it with hammers, sharpen the blades, stoke the fires, repair the tools, sweep the floors, run the canteen, keep the accounts, and so on. Which workers are directly involved in making knives, and which only indirectly? I don't think we can say. There may be some sense in the idea that those who actually work the steel are more directly involved in the making of knives than those who sweep the floors. But is there really an absolute distinction—even an imprecise one—between those who are directly involved in making knives and those who are only indirectly involved? Or think of walking. Which parts of a human being—which atoms—are

---

<sup>87</sup> Ibid., 91 – 92.

directly involved in his walking? Those in his legs, surely. But are all the atoms in his legs directly involved in his walking? Suppose he has excess water in his legs owing to poor circulation, which hinders his walking. Are the atoms making up the excess water directly involved in his walking? (And what could determine which molecules are “excess” and which belong there?) What about his arms, or spinal cord, or heart? Is their involvement in his walking direct or indirect?<sup>88</sup>

As Olson points out, the problem of characterizing the notion of something’s being directly involved in activities such as experiencing and walking isn’t that we are ignorant of the mechanics of these activities. Nor is it that we are unable to draw the boundary between direct and indirect involvement in a way that does not admit of borderline cases—i.e., cases in which it is simply indeterminate whether a certain atom is involved directly or indirectly. In Olson’s view, asking which atoms are directly involved in walking or experiencing and which are involved only indirectly in these activities is not like asking which rocks are parts of Mount Everest and which aren’t. We seem to have a relatively clear notion of a rock’s being part of a mountain, even if this notion admits of borderline cases. The problem according to Olson is that the very idea of direct involvement appears to be devoid of content. Thus, according to Olson, any decision we make regarding which atoms are directly involved in a being’s experiencing is bound to be arbitrary. The most we can say is that some atoms are more directly involved than others. There is no principled way of drawing even a vague boundary between those that are directly involved and those that aren’t.

I am not convinced. The notion of direct involvement seems clearer if one thinks of conscious experience as a state that one is in, as opposed to an activity in which one engages. For example, consider the maximal conscious state that a person is in at a

---

<sup>88</sup> Ibid., 92.

particular time  $t$ . Intuitively, this maximal conscious state has a supervenience base, even if its boundaries are vague. Appealing to this intuition, the Minimalist could adopt the following definition of direct involvement in consciousness at a time:

Something,  $x$ , is directly involved in producing consciousness at time  $t$  =df.  $x$  exists at  $t$ , and there is a maximal conscious state  $C$  at  $t$ , and  $C$  has a (possibly vague) supervenience base,  $B$ , at  $t$ , and  $x$  is (possibly vaguely) a part of  $B$  at  $t$ .

This is not quite an answer to the special composition question, but it does seem to show that the notion of direct involvement in consciousness can be defined in a way that is both principled, and immune to Olson's worries.

Perhaps a promising answer to the special composition question for defenders of the Psychological Approach is that the  $x$ s compose a  $y$  at  $t$  *iff* the joint activities of the  $x$ s at  $t$  constitute a *mental life*, a complex event consisting of various mental phenomena, or perhaps some highly specific kind of mental phenomena. Why should this answer be thought inferior to the answer that is offered by animalists such as van Inwagen and Olson? Biological lives and mental lives both seem fairly well-individuated. Both are natural phenomena; biologists study biological lives, psychologists and neurobiologists study mental lives. How does animalism have an advantage here?

#### **3.4.4 If the Biological Approach and The Psychological Approach are combined with Eliminativism, the Former Loses its Advantage**

Suppose The Biological Approach is combined with Eliminativism. According to the resulting version of Animalism:

Atoms exist and they are simple (i.e. they have no parts).

The only composite objects that exist are essentially living animals.

Call this view Animalism 1.

Now consider a version of The Psychological Approach with the following features

Atoms exist and they are simple (i.e. they have no parts).

The only composite objects that exist are psychological continuers.

Call this view Psychological Approach 1 (PA1). Suppose that, apart from claims about the existence and persistence of psychological continuers and essentially living animals, Animalism1 and PA1 imply exactly the same ontological and metaphysical commitments. Let us compare Animalism1 and PA1 with respect to the four theoretical advantages Animalism is alleged to have over The Psychological Approach. These are: (1) avoidance of the too many subjects problem, (2) ontological and ideological simplicity, (3) providing a more plausible answer to the special composition question, and (4) greater consistency with common sense.

Neither Animalism1 nor PA1 is subject to the version of too many subjects problem that arises for views that countenance spatially coincident entities of different kinds. This is because on either view there is only one kind of entity that could plausibly be the subject of the person's conscious experiences. On Animalism1, this entity is the essentially living animal. On PA1, it is the psychological continuer. For this reason, it seems that neither view has an advantage over the other with respect to its ability to avoid the too many subjects problem. (Notice that defenders of either view could follow Olson in denying that the person has large parts—such as a head—that would apparently be conscious if the person was conscious.)

Animalism1 and PA1 both acknowledge the same number of person-like entities. (By ‘person-like entity’ I mean something that is a plausible candidate for being one of us. Animals and psychological continuers are both person-like entities.) Similarly, both Animalism1 and PA1 recognize the same number of kinds of entities. On Animalism1, the relevant kinds are atom and essentially living animal. On PA1, they are atom and psychological continuer. All else equal, it seems, neither view has the advantage of ontological simplicity.

Next, consider common sense. Both Animalism1 and PA1 deny the existence of most of the familiar objects that common sense says exist. However, one might argue that Animalism1 at least recognizes the existence of animals. For this reason, a defender of Animalism1 might argue, although both Animalism1 and PA1 are radically at odds with common sense Animalism1 is less at odds with common sense than PA1.

This is an interesting point, but I don’t think Animalists should find much comfort in it. Given that both Animalism1 and PA1 violate common sense, each view’s violation is so severe that, in relative terms, any advantage gained by consistency with common sense would be trivial. First, on either view, the deliverances of common sense are almost always mistaken. One might have thought that greater consistency with common sense provides a theoretical advantage because what is implied by commonsense is more likely to be true. But this assumption is thrown into doubt once one admits that common sense is dead wrong regarding almost every judgment about what exists. In this case, consistency with common sense may seem much less important as a desideratum for a theory. Not only would common sense favor Animalism1 only very slightly, but the theoretical importance of this advantage would seem greatly diminished.

Second, as I argued, Animalism's answer to the special composition question may imply that the boundaries of animals are radically different from what common sense suggests. (I think this is an inevitable result of any reasonable bottom-up account of animalistic unity.) If so, then the animals which Animalism posits will be very different from those common sense recognizes.

Finally, assuming that consistency with common sense gives Animalism an advantage over PA1, we need to consider whether this advantage outweighs the general advantage that all versions of The Psychological Approach (where this includes PA1) have over Animalism, namely the intuitive appeal that comes from considering brain transplantation and other hypothetical cases in which the person apparently outlives the animal. But this seems highly unlikely if, as I argued in the previous chapter, Animalists do not have a convincing way of countering this intuitive appeal of The Psychological Approach.

### **3.5 The Second Response for Defenders of The Biological Approach: Switcheroo**

Although Olson is now inclined to accept Eliminativism in order to avoid the too many subjects problem for The Biological Approach, he has also argued that this problem can be solved by claiming that although bodies exist they do not coincide with essentially living animals. On this proposal, the corpse problem is solved by claiming that the body does not begin to exist until the essentially living animal dies and ceases to exist. Thus, at the moment of death, the body (a substance of a certain kind) supplants the essentially living animal (a substance of different kind) in the region of space it previously occupied. I will call the view that avoids coincident entities by claiming that one entity supplants another



at the exact moment when the former ceases to exist—Switcheroo. It is as if, in a single instant, God snatches away one entity and deftly replaces it with a new entity of a different kind, without anyone noticing the switch.

Initially, Olson found this proposal attractive because he believed there was no analogous solution available to his opponents. That is, he believed that defenders of The Psychological Approach couldn't plausibly claim that the animal exists only after (and before) the psychological continuer exists, and that defenders of The Psychological Approach must therefore give a different response to the too-many-subjects problem that arises for their view. I will now argue against this assumption.

### **3.5.1 Combining Animalism and the Psychological Approach with Switcheroo**

Suppose Animalists accept Switcheroo. Call the resulting view Animalism2. This view includes the following claims:

1. Each of us is an essentially living animal.
2. There are no psychological continuers.
3. In addition to essentially living animals, bodies exist. However essentially living animals and bodies do not coincide. When an essentially living animal dies peacefully due to natural causes, it is supplanted by a body at the moment of its biological death.

Now consider a version of The Psychological Approach that includes the following claims:

4. Each of us is a psychological continuer.
5. There are no essentially living animals.
6. In addition to psychological continuers, bodies exist. If a body is alive, it is alive only contingently. The term 'animals' has its normal application; it refers to those entities

common sense recognizes as animals. However, no animal is essentially alive. Instead, every animal is a body that could be alive or dead. If it is dead, then it is a dead animal. Psychological continuers and bodies never coincide. When one of us (i.e. a psychological continuer) irreversibly loses the capacity for consciousness, a body supplants her at the moment at which this capacity is lost.

Call this version of The Psychological Approach PA2. Let us suppose that apart from the differences between Animalism2 and PA2 that are explicitly mentioned here, each view entails the same ontological commitments.

I will now argue that Animalism2 is no better than PA2 with respect to considerations (1) through (4).

Start with (1)—the too many subjects problem. Both Animalism2 and PA2 avoid the too many subjects problem. This problem arises for views that posit coincident entities or entities which, although not coincident, share the same consciousness-generating parts. Neither Animalism2 nor PA2 posits such entities. Each implies that two or more person-like entities can never coincide, and neither implies that two or more person-like entities share the same consciousness-generating parts.

Next, consider (2)—ontological simplicity. Both Animalism2 and PA2 posit the same number (two) of person-like entities. On Animalism2, these entities are the essentially living animal and the body. On PA2, they are the psychological continuer and the body. Since neither view posits more person-like entities (or more kinds of person-like entity) than the other, and since all of their other ontological commitments are the same, it seems that neither view could have an advantage of ontological simplicity.

Next, consider (3)—common sense. Both Animalism<sup>2</sup> and PA<sup>2</sup> are significantly at odds with our common sense judgments about the persistence of ordinary objects. The supplanting of a psychological continuer by a (living) body at the moment when one's capacity for consciousness is irreversibly lost seems no more implausible than the supplanting of an essentially living animal by a body at the moment of biological death. The supplanting of one entity by another seems equally implausible in each case. Fred Feldman, a Defender of The Bodily Approach, emphasizes the implausibility of the Defender of The Biological Approach's claim that the body supplants the essentially living animal at the moment of death:

Surely in every case in which a 150 pound person dies and leaves a 150 pound corpse, there are plenty of obvious reasons to suppose that a certain 150 pound object persists through the change from being alive to being dead...suppose a terminally ill 150-pound person is resting on a sensitive scale when he dies. Suppose he dies peacefully, so that the needle of the scale does not move. It pointed to '150' before he died, and it continued to point to '150' when and after he died. It did not even quiver at the moment of death. It would have been hard to remove a person and replace him with an equally heavy corpse. It would have been nearly impossible to do this without causing the needle on the scale to move. Since the needle did not move, there is at least some *prima facie* reason to suppose that some 150-pound object persisted through the change.<sup>89</sup>

Feldman's example illustrates the implausibility of Switcheroo as a defense of The Biological Approach. But the strategy also seems implausible as a defense of most versions of The Psychological Approach.

To see the absurdity of Switcheroo when it is combined with The Psychological Approach, suppose that the 150-pound person in Feldman's example is resting on the scale when she irreversibly loses the neural capacity for consciousness. (Suppose this was the inevitable result of a brain injury.) Suppose that when this happens, the arrow on the scale

---

<sup>89</sup> Feldman, "The Termination Thesis," 105.

continues to point to ‘150’. One who defends The Psychological Approach and who accepts Switcheroo will say that in this case, right when the capacity for consciousness is irreversibly lost, one of us—a psychological continuer—is supplanted by an animal. (This proponent of The Psychological Approach need not claim that the animal which supplants the psychological continuer is essentially alive. She could claim, following Defenders of The Bodily Approach, that this animal is a body that is alive only contingently.) In this modified version of Feldman’s example, as in his original example, there are plenty of reasons for thinking that a certain 150-pound object has persisted through an important change (in this case, the change is the irreversible loss of the capacity for consciousness).

It seems that if the Defender of The Biological Approach and the defender of PA both appeal to Switcheroo, their views will be implausible for the same kind of reason—both will violate what I will call

*Observable Continuity.* Without appealing to any principle of the identity of animals or bodies over time, we often recognize clear cases of the persistence of animals or bodies through change. For example, we recognize clear instances of an animal or body gaining (or losing) the neural capacity for consciousness while continuing to exist. We also recognize clear instances of an animal or body gaining or losing its biological life while continuing to exist. In these instances, we can just see that a certain animal or body has persisted through a certain change.

Animalism<sub>2</sub> and PA<sub>2</sub> both imply that many apparently observable instances of an animal or body persisting through change are in fact instances of that animal or body ceasing to exist or beginning to exist abruptly—in these instances, an animal or body either

springs into existence or blinks out of existence. I do not see how either view does more violence to common sense. Both views are radically at odds with common sense.

### **3.5.2 Why Defenders of the Biological Approach Cannot Appeal to the Impressiveness of Biological Death to Defend Switcheroo**

In defense of the claim that an animal ceases to exist when it dies, Olson and David DeGrazia both point to the radical changes that take place at the cellular level during the process of biological death.<sup>90</sup> All that unimaginably complex biological machinery breaks down into clumps of dead matter that appear to have little if any integrated biological function. Doesn't this remarkable stoppage of impressive biological machinery—this loss of unity—indicate that something has ceased to exist? And, if something has ceased to exist, doesn't this give us reason to say that the lifeless body that exists after the moment at which death occurs comes into existence at that moment?

I think our answer to the second question should be *no*. Suppose for the sake of argument that we have good reason to accept

- (a) Whenever something dies, it ceases to exist.

This claim is compatible with the claim for which Feldman argues, namely

- (b) Whenever something dies (peacefully) there is something that persists through the moment of death. There is some entity that is alive before the moment of death, and dead afterward.

---

<sup>90</sup> Olson, *The Human Animal*; David DeGrazia, *Human Identity and Bioethics* (Cambridge: Cambridge University Press, 2005).

In his defense of (b) Feldman does not offer a principle of bodily identity over time. Rather, he appeals to something like Observable Continuity. He assumes that we can identify clear instances of the persistence (or failure of persistence) of objects through change without appealing to a general view about the persistence conditions of objects. Feldman would say that whenever one considers an instance of biological death that occurs peacefully, one finds a clear instance of the persistence of an object through change. One sees that there is something which is alive one minute and dead the next. This claim about observable instances of persistence through change is compatible with the claim that Olson and DeGrazia accept, namely that whenever death occurs, something ceases to exist. Claim (a) is compatible with claim (b). Even if there are good reasons to accept (a), they do not seem to make the denial of (b) more plausible. The denial of (b) seems implausible even on the assumption that (a) is true.

The well-known case of the statue and lump provides a useful analogy here. Even if we have reason to say that a certain object—a statue—would cease to exist if it were squished, it also seems that we could just see that a certain thing—a lump—had persisted through the squishing of the statue. The squishing of the statue would be a clear instance of the persistence of a lump through change.

Now consider the first of the two questions we asked: Doesn't the stoppage of impressive biological machinery that occurs at death give us reason to suppose that something has ceased to exist? If we answer 'yes' then we attribute ontological significance to the stoppage of an animal's impressive biological machinery. But the defender of The Psychological Approach could reasonably appeal to the ontological significance of the stoppage of impressive *psychological* machinery in support of the claim that a

psychological continuer ceases to exist at the moment at which this machinery breaks down. In arguing for the view that an animal ceases to exist when it dies, Olson writes,

The changes that go on in an [animal] when it dies are really quite dramatic...If it looks like there isn't all that much difference between a living animal and a fresh corpse, that is because the most striking changes take place at the microscopic level.<sup>91</sup>

But something very similar is true of one's irreversibly losing the capacity for consciousness. When this happens, a neural network of unimaginable chemical complexity breaks down into a soup of disorganized protenacious fluids. No one observes these changes; all they observe is a seemingly smooth transition between what appear to be different states of one and the same being. A defender of The Psychological Approach could claim that if there does not appear to be much difference between one who possesses the neural capacity for consciousness and one who lacks this capacity, this is because the most striking changes that occur when one irreversibly loses this capacity take place at the microscopic level. It therefore seems to me that the Defender of The Biological Approach can gain no advantage over the defender of The Psychological Approach by pointing to physiological changes which they believe indicate that the loss of one's biological life brings one's existence to an end, since the defender of The Psychological Approach can respond by pointing to similar changes that occur in the brain which suggest that the loss of one's mental life brings one's existence to an end.

I should add that I am not convinced by these appeals to complexity, whether such appeals come from the Animalist or from the defender of The Psychological Approach. The reason there does not appear to be much difference between a living animal and a fresh corpse, or between a person with the neural capacity for consciousness and a living animal

---

<sup>91</sup> Olson, *The Human Animal*, 151 – 152.

that lacks this capacity, is that there appears to be an extremely high degree of continuity (for example, bodily continuity) between the former and the latter, despite any dramatic changes that may occur at the microscopic level in each case. But insofar as one finds such appeals to complexity convincing, I cannot think of a good reason one might have for accepting one of these appeals (say, the Animalist's) while rejecting the other.

Finally, (for reasons considered in Section 3.3.3) I think Animalism<sub>2</sub> derives no advantage from the fact that it recognizes the existence of essentially living animals, while PA<sub>2</sub> doesn't. Recognizing the existence of essentially living animals would give Animalism<sub>2</sub> a theoretical advantage over PA<sub>2</sub> only if it was harder to deny the existence of psychological continuers than to deny the existence of essentially living animals. However, as I claimed in Section 3.3.3, I can find no good reason to believe this. In fact, it seems to me even more difficult to deny the existence of psychological continuers than to deny the existence of essentially living animals.

To sum up, Animalism<sub>2</sub> is no better than PA<sub>2</sub> with respect to theoretical considerations (1) through (4). Now let us combine this conclusion with Section 3.3's conclusion that Animalism<sub>1</sub> is no better than PA<sub>1</sub> with respect to (1) through (4). The conclusion that we should now draw is that if The Biological Approach and The Psychological Approach are combined either with Eliminativism or with Switcheroo, the resulting version of The Biological Approach is no better than the corresponding version of The Psychological Approach with respect to considerations (1) through (4). It therefore seems that those who have been led to accept The Biological Approach on the basis of theoretical considerations (1) through (4) do not have good reason for accepting their view



rather than The Psychological Approach. These defenders of The Biological Approach have no good reason to defend The Biological Approach.

### **3.6 The Bodily Approach and the Psychological Approach**

Perhaps the most plausible way to avoid the bizarre supplanting of one entity by another that is implied by Switcheroo, is to follow the lead of writers such as Feldman, Mackie, Ayers, Carter, and Williams, and accept The Bodily Approach, according to which each of us is a body that has the capacity for consciousness only contingently, and that is alive only contingently. On this view, if one were to irreversibly lose all brain function while one's biological life continued, one would continue to exist in a permanent vegetative state; and if one's biological life were to later cease, one would continue to exist as a corpse. With each change in one's physical state, there is no person-like entity that ceases to exist. There is only one person-like entity—a body—which persists through the change from being awake and conscious to being alive but permanently unconscious, and then from being alive to being dead. Although one who defends this view might use the terms 'animal' and 'animal' to refer to the body, on this view there are no essentially living animals—there is nothing that must cease to exist at the moment of its biological death. It is open to a Defender of The Bodily Approach to use the term 'animal' to refer to something that has died. Thus, a defender of the Biological Approach can say, plausibly, that there are dead animals.

The Bodily Approach appears to have several nice features: First, it does not countenance spatially coincident entities of different kinds. Hence, it seems to avoid the too many subjects problem. Second, unlike Animalism<sub>2</sub> and PA<sub>2</sub>, it is consistent with

Observable Continuity. Apparently observable instances of the persistence of objects through change are just that—*observable*. They do not involve the abrupt and bizarre supplanting of one person-like entity by another. Third, it seems the bodily approach can have these features while remaining compatible with commonsense ontology. In order to get the desired result that there is only one conscious entity and one person where I am now sitting, defenders of The Bodily Approach need not appeal to Eliminativism; they need not adopt a radical eliminativist view of what exists. They can claim that the one and only conscious entity and person sitting in my chair is a body, which exists alongside many other objects, such as the chair in which I am sitting, the desk at which I am writing, the building in which this chair and desk are located, etc.

Finally, and most important, when we consider the comparison of The Biological Approach and The Psychological Approach, developed in Sections 3.3 and 3.4, we see no obvious parallel comparison between The Bodily Approach and The Psychological Approach. In comparing The Biological Approach and The Psychological Approach, I first considered how a defender of The Psychological Approach could respond if a defender of The Biological Approach were to appeal to Eliminativism to avoid the too many subjects problem for her view, and then I considered how a defender of The Psychological Approach could respond if a defender of The Biological Approach were to appeal to Switcheroo to avoid this problem. In each case, I argued, the defender of The Psychological Approach can point to a version of her view that is no worse than the proposed version of The Biological Approach, at least with respect to considerations (1) through (4). However, as we have just seen, The Bodily Approach seems to avoid the too many subjects problem without relying on Eliminativism or Switcheroo. Thus, there is no

obvious reason for the defender of The Bodily Approach to appeal to either of these implausible views, and hence, there is no obvious way to extend the arguments of Sections 3.3 and 3.4 to The Bodily Approach. Therefore, The Bodily Approach may appear to be the best alternative that is available to Animalists.

### **3.6.1 Why the Bodily Approach Is No Better Than the Biological Approach**

However, upon reflection, I find The Bodily Approach to be no better than The Biological Approach. Most Animalists who accept The Bodily Approach find it attractive because they believe that it straightforwardly avoids many of the problems for The Biological Approach. But few defenders of The Bodily Approach offer a detailed account of the body—of its most fundamental properties and its persistence conditions. I suspect that if we were to try to give such an account, we would find either that our account was vulnerable to the same problems that arise for The Biological Approach, or that it was clearly worse than any version of The Biological Approach that we have considered.

To see this, recall Feldman's example in which a certain 150 lb. body that is resting on a scale is apparently alive at one moment and dead the next. As Feldman argues, and as I have suggested, this would seem to be a clear instance of the persistence of a body through change. This example makes trouble for The Biological Approach because on that approach a certain substance—an essentially living animal—ceases to exist at the moment of death. It seems defenders of The Biological Approach must either deny the existence of a body that persists through the death of the animal, or recognize the existence of such a body and explain how it is related to the essentially living animal. As we saw, both options are problematic.

But is The Bodily Approach immune to this kind of problem? Consider the following illustrative, if not somewhat extreme, case. In the epic tale The Mahabharata, as a means of protecting his beloved wife from the unwanted advances of a womanizing general, one of the major protagonists of the story, Bhima, disguises himself as a beautiful woman and lures the general into his private tent. Then, in a gruesome display of Herculean strength, he transforms the man into something one would rightly hesitate to call a human body.

And that foremost of mighty persons ... again attacked his adversary and thrust his arms and legs and neck and head into his body ... reducing into a shapeless mass the deer, which form Sacrifice had assumed in order to escape his ire. And having crushed all his limbs and reduced him into a ball of flesh, the mighty Bhima showed him unto Krishna.<sup>92</sup>

Bhima's squishing of the general seems analogous in certain respects to the squishing of the clay statue in the familiar example of the statue and lump. If, as The Biological Approach implies, the general is an *essentially living animal*, then certainly he ceases to exist at some point during his attack at the hands of Bhima. But even if, as The Bodily Approach implies, the general is a body that is alive *only contingently*, still we ought to say that there is something, for example, a lump of flesh, that persists through the gruesome change that occurs during the attack, as well as something, for example, a distinctively human body—one that has at least some of its distinctive human features essentially—does not persist through that change.

Whether The Bodily Approach implies that the general persists through the change depends on how the sortal *body* is defined. And here The Bodily Approach faces a

---

<sup>92</sup> Krishna Vyasa, *Mahabharata*, English translation by Ganguli. Retrieved from <http://books.google.com/>

dilemma: Suppose that *body* is defined narrowly so it refers only to those objects that are sufficiently similar in form and likeness to a prototypical human body—in other words—only to distinctively human bodies. Then The Bodily Approach should imply that the general does not persist through the attack, since what remains after the attack does not have enough of those features that make a body distinctively human. In that case, The Bodily Approach faces a problem which parallels the corpse problem for The Biological Approach and the vegetable problem for The Psychological Approach. As we saw, it is apparent that something—a lump of flesh—persists through the attack. Before the attack, this lump has most of the features of a prototypical human body. During the attack, it is transformed into a grotesque ball that has almost none of those features. If the lump persists but the distinctively human body ceases to exist, was the lump coincident with the distinctively human body prior to the attack? If so, did the lump constitute the distinctively human body? Is the distinctively human body a temporal part of the lump? Or, assuming that the lump and the distinctively human body weren't coincident before the attack, did the lump begin to exist right when the body ceased to exist? (We could get this result by combining The Bodily Approach and Switcheroo.) Or, could it be that despite appearances there are no lumps of flesh, although there are distinctively human bodies? (We could get this result by combining The Bodily Approach and Eliminativism.) In short, if *body* is defined too narrowly, The Bodily Approach faces the same general problem as The Biological Approach and The Psychological Approach. We might call the version of this problem that arises specifically for The Bodily Approach *the lump problem*. This is the first horn of the dilemma for The Bodily Approach.

The second horn is encountered if *body* is defined broadly enough to avoid the first horn. In that case, The Bodily Approach becomes so implausible that it is no longer a live hypothesis. For example, suppose that body is defined so that it implies that the bodies we are could persist after being molded into something that lacks most of our distinctively human features. In that case, it is clear that in the example we considered, the general “survives” his attack—in other words, he persists through that attack. (Notice that this claim is consistent with the language the author uses, e.g., “After ... reducing *him* to a ball of flesh, the mighty Bhima showed *him* unto Krishna.”) The Defender of The Bodily Approach could then deny that there is any person-like entity—such as a distinctively human body—that ceases to exist in the attack. Instead, she could claim, there is only one person-like entity in the story, and this person-like entity has a distinctively human form prior to the attack and a distinctively spherical form after the attack. This is similar to how some philosophers claim that in the case of the statue and the lump nothing ceases to exist when the statue is squished, but a single entity—the lump—persists through the squishing. To make the parallel even more exact, the Defender of The Bodily Approach could claim that each of us just is a mere lump of flesh, or, perhaps, a mere lump of human “stuff.”

However, this claim is massively implausible. Our persistence conditions are not those of mere lumps. Suppose that I weigh 350 lbs. and that, tired of dragging around the extra weight, decide to undergo liposuction. During the liposuction procedure, 200 lbs. of human matter are vacuumed out of me and placed in a receptacle for biohazardous human waste. At the end of the procedure, there is a 150 lb. object lying on the operating table, who is both biologically and psychologically continuous with (and looks at least somewhat like) the 350 lb. person whom the plastic surgeons wheeled into the operating room.

Intuitively, the 350 lb. being survives the procedure as the 150 lb. being. But there is no reason to think that a 350 lb. lump of human stuff survives the procedure as the 150 lb. being. There is no reason at all to assume that a mere lump of stuff could persist as something that has less than 50% of its constituent matter. What happens to the 350 lb. lump of human stuff in this case? I think that the lump either (a) ceases to exist during liposuction, (b) is reduced to 200 lbs. and then placed in a sealed container for biohazardous human waste, or (c) becomes scattered and continues to exist with one of its major parts sealed inside the container and the other resting on the operating table. Clearly, neither (a), (b), nor (c) is true of *me* in this case. I am not a lump!

### **3.7 Why Animalism is no Better than the Psychological Approach with respect to the Four Major Considerations**

I have argued that The Bodily Approach has no more theoretical plausibility than The Biological Approach. If ‘body’ is defined narrowly so that it refers to the distinctively human body, then The Bodily Approach and The Biological Approach are in the same boat. Defenders of both views must either posit spatially coincident entities, resulting in the too many subjects problem, or they must deny spatial coincidence by appealing to Eliminativism or Switcheroo. On the other hand, if ‘body’ is defined broadly so that it refers only to mere lumps of stuff, then The Bodily Approach will be completely implausible—worse than any view that one could reasonably consider a live hypothesis. This is important because it allows us to finally establish one of the major claims of this chapter, which I stated in the introduction of this chapter, namely

For each Animalist theory, TA, that we have considered, there is a corresponding version of The Psychological Approach, T<sub>psych</sub>, such that either (i) TA is no better than T<sub>psych</sub> with respect to the four major considerations that are supposed to make Animalism better than The Psychological Approach, or (ii) TA is all-things-considered worse than T<sub>psych</sub>.

For reasons I gave in Sections 3.4 and 3.5, this claim is true if ‘Animalism’ refers only to The Biological Approach. For, as we saw, in order to avoid the too many subjects problem, defenders of The Biological Approach appeal either to Eliminativism or to Switcheroo. And, as I showed in Sections 3.3 and 3.4, when The Biological Approach is combined with either of these two views, the resulting version of Animalism is no better than the corresponding version of The Psychological Approach that one gets by combining it with that very same view (either Eliminativism or Switcheroo). If the argument of the previous section (3.6) is correct, then the above claim is also true of The Bodily Approach. In that case, I have established both of the following claims.

1. For every proposed version of The Biological Approach, T<sub>bio</sub>, there is some version of The Psychological Approach, T<sub>psych</sub>, such that T<sub>bio</sub> is no better than T<sub>psych</sub> with respect to good-making features (1) through (4).
2. For every proposed version of The Bodily Approach, T<sub>body</sub>, there is a version of The Biological Approach, T<sub>bio</sub>, such that either (i) T<sub>body</sub> is no better than T<sub>bio</sub> with respect to good-making features (1) through (4), or (ii) T<sub>body</sub> is all-things-considered worse than T<sub>bio</sub>.

If both 1 and 2 are true, then the following claim should also be true.



3. For every proposed version of The Bodily Approach,  $T_{\text{body}}$ , there is some version of The Psychological Approach,  $T_{\text{psych}}$ , such that either (i)  $T_{\text{body}}$  is no better than  $T_{\text{psych}}$  with respect to alleged good-making features (1) through (4), or (ii)  $T_{\text{body}}$  is all-things-considered worse than  $T_{\text{psych}}$ .

(I will say more about this step from 1 & 2 to 3 in a moment.)

Finally, since The Biological Approach and The Bodily Approach jointly exhaust Animalism, if 3 is true, then Supporting Claim 1 should also be true. That is, it should be true that

For any Animalist theory (i.e. any version of Animalism), TA, there is at least one version of The Psychological Approach,  $T_{\text{psych}}$ , such that either (i) TA is no better than  $T_{\text{psych}}$  with respect to considerations (1) through (4), or (ii) TA is all-things-considered worse than  $T_{\text{psych}}$ .

An important clarification is in order. That 3 can be inferred on the basis of 1 & 2 presupposes that the relation ‘is no better than’ is transitive—in other words, for all a, b, and c, if a is no better than b, and b is no better than c, then a is no better than c. Call this claim *transitivity*. To see that the inference from 1 & 2 to 3 assumes transitivity, consider a simple example. Suppose we have three different groups of objects—the xs, the ys, and the zs—and that all of these objects can be meaningfully evaluated and compared. Now suppose we find that at least one of the xs is such that none of the ys is better than it, and suppose we also find that at least one of the ys is such that none of the zs is better than it. It seems we should now conclude that at least one of the xs is such that none of the zs is better than it. It is exactly this sort of reasoning which leads one to conclude 3 on the basis of 1 & 2.

If (1) through (4) exhaust all possible theoretical considerations with respect to which Animalism could be better than The Psychological Approach, and if The Biological Approach and The Bodily Approach jointly exhaust all possible versions of Animalism then, it seems, we have yet to see a good theoretical reason to accept Animalism.

### **3.8 Against the Hybrid Approach**

However, as I stated in Chapter 1, some defend the Hybrid Approach. I have argued that at least some versions of the Psychological Approach are more plausible than the Biological and Bodily Approaches. These views are undermined by the Transplant and Dicephalus Arguments, and they appear to have few if any compensating virtues that are not also virtues of some version of the Psychological Approach. In this section I will consider the Hybrid Approach. Recall that on this approach, the kind of continuity that underpins the identity over time of animals according to the Biological (or Bodily) Approach—some kind of biological (or bodily) continuity—is sufficient for identity in the absence of psychological continuity, and psychological continuity (of an appropriate kind) is sufficient for identity in the absence of biological (or bodily) continuity. But when the two kinds of continuity diverge, identity follows psychological continuity. The Hybrid Approach is less counterintuitive than the Biological and Bodily Approaches because the former, unlike the latter, implies that I once existed as a non-conscious fetus, and that I could end up in a permanent vegetative state, but also that if my brain were to be transplanted into a different body, I would go with my brain.

#### **3.8.1 A Powerful Objection to the Hybrid Approach**

There is a powerful objection to the Hybrid Approach. Consider two numerically distinct adult human persons, A and B, with completely different psychologies. Suppose that A's brain is destroyed and that A continues to exist in a vegetative state (biological continuity is preserved). B's brain and psychology are then transplanted into A's brainless but living body, resulting in a certain individual with B's original brain and psychology and A's original living body. Call this individual C.

The Hybrid Approach implies that B is identical to C, since B is psychologically continuous with C, and since, on the Hybrid Approach, psychological continuity is sufficient for identity. But the Hybrid Approach also implies that A is identical to C, since, on this approach, it is also the case that biological continuity is sufficient for identity in the absence of psychological continuity, and A is biologically continuous (but not psychologically continuous) with C. A and C satisfy the Hybrid Approach's criteria of identity over time; so do B and C. But identity is a one-to-one relation—and so this result cannot be correct. Instead, the Hybrid Approach must be committed to one of the following results. Either (i) C is identical to A, who continues to exist with B's former brain, while B either (a) ceases to exist, (b) continues to exist in a vegetative state, or (c) becomes spatially coincident with A; (ii) C is identical to B, who continues to exist with A's former body, while A either (a) ceases to exist, or (b) becomes spatially coincident with B; or (iii) C isn't identical to A or B, but is an entirely new individual, while A ceases to exist, and B either (a) ceases to exist or (b) continues to exist in a vegetative state.

I will discuss (i) (c) and (ii) (b) in the subsequent paragraph. Here, I will focus on the other possibilities listed above. The problem for the Hybrid Approach is that all of these epistemic possibilities are both implausible and incompatible with this approach; (i)

(a) and (iii) (a) are implausible because one cannot destroy an adult human person simply by transplanting her brain into a different headless body. (This also fails to capture the transplant intuition.) These views are also incompatible with the Hybrid Approach, as are (i) (b) and (iii) (b); the Hybrid Approach states that psychological continuity is sufficient for identity when psychological continuity and biological continuity come apart. Thus, according to this approach, it cannot be the case that B ceases to exist or continues to exist in a vegetative state. (ii) (a) is implausible because one cannot destroy a brainless human animal simply by giving it a new brain. This is also incompatible with the Hybrid Approach, which states that biological continuity is sufficient for identity in the absence of psychological continuity. According to this approach, it cannot be the case that A ceases to exist.

That just leaves (i) (c) and (ii) (b), according to which A and B become spatially coincident after the transplant. That gives us two spatially coincident entities of the same kind—both human animals with hybrid identity conditions. None of the other approaches that we have considered has this consequence. This gives us too many subjects, which is usually cited as one of the main reasons for rejecting the Psychological Approach. But since these two subjects are *of the same kind*, I do not think that we can plausibly claim that only one of these animals is the non-derivative thinker.

Furthermore, defenders of the Hybrid Approach have said nothing about dicephalus. Are there two entities of our kind in a case of dicephalus, or one entity of our kind with two separate brains and psychologies? If defenders of the Hybrid Approach say that there is only one entity of our kind, then this approach is subject to the same criticisms as the Biological and Bodily Approaches that we discussed in Chapter 2. If they say that

there are two entities of our kind, then because defenders of the Hybrid Approach say that we are human animals, they must explain how the two human animals in this case are related. Are they spatially and materially coincident? If not, then are they parts of the larger organic mass that sustains both of them? If so, why? And why think that these entities are human animals? (If anything in this case deserves to be called a human animal, surely it is the biologically living being with two heads.) Furthermore, if each human animal is a part of the organic mass, why should we say that these human animals existed during the embryonic stage, before the formation of the head and brain, when anyone looking would have said that there was at most *one* human animal—a biologically living human being? And if the human animals in a case of dicephalus are but parts of what appears to be a single human animal, why shouldn't we think that most or all human animals are parts of what appears to be a single human animal?

In short, it seems to me that either defenders of the Hybrid Approach must say that we have spatially coincident entities of the same kind in cases in which it seems obvious that we do not, or they must give up part of the appeal of their approach, namely the part that comes from intuitions about the identity and persistence conditions of human animals relating to biological continuity and biological life. In either case, it becomes unclear why the Hybrid Approach is any better than the Psychological Approach.

### **3.8.2 Against Patrick Lee's Animalism**

A different view has been defended by Patrick Lee.<sup>93</sup> According to Lee, each of us is essentially and most fundamentally a rational animal—something that has the internal capacity for rational thought and action. We have this capacity from the moment of conception. Even embryos have a remote capacity for rational thought and action, since they have the capacity to develop into something that thinks and acts rationally. Once the rational animal loses this capacity, it ceases to exist. Thus, for Lee, under normal conditions, a rational animal that is one of us ceases to exist when it loses the neural capacity for consciousness. For example, if I lapse into an irreversible vegetative state, I thereby cease to exist. In this case, the living but now permanently unconscious thing that remains behind when I cease to exist, which Animalists such as Olson and van Inwagen would call an animal is, on Lee’s view, a “mere heap” of organic tissue; similarly, when my brain is transplanted into a different body, the empty-headed thing that remains behind, which standard Animalism identifies as an animal, is a mere heap of organic tissue. In this case I go where my brain goes, since my capacity for rational thought and action are realized my brain. Thus, Lee’s version of Animalism is consistent with The Transplant Intuition—the intuition that in the transplant case, I am identical to the person who inherits my brain.

It is not clear whether Lee’s view can avoid the problems of conjoined twinning considered in Chapter 2. His official view about dicephalus, in which there appears to be two heads emerging from a single body—is that there are two rational animals, and hence,

---

<sup>93</sup> Patrick Lee and Germain Grisez, “Total Brain Death: A Reply to Alan Shewmon,” *Bioethics* 26 (2012), 275-284.

two of us in that case.<sup>94</sup> He argues for this view by pointing to the duplication of organs in dicephalus, emphasizing the fact that there are two separate brains, and, especially, two separate brainstems. I am uncertain what he would say about the case of cephalopagus, in which there appear to be two separate animals, each with its own separate brainstem, but only one brain (one consciousness-generating part) shared equally between them. Suppose, that on Lee's view, the number of rational animals in a case of conjoined twinning is determined by the number of separate brains. In that case, Lee could say that there are two of us in dicephalus but only one of us in cephalopagus. And, as I have argued, both of these claims appear to be correct. What we have, then, is a version of Animalism that seems to avoid the main objections to the standard versions of that view. It can get what is intuitively the right answer in both the transplant case and the conjoined twinning cases. Moreover, unlike The Psychological Approach, this version of Animalism is consistent with the commonsense view that each of us was once a mindless fetus. What should we make of this view?

One might be inclined to reject the view on the grounds that it is simply unbelievable that a human embryo could have the capacity for rational thought and action. However, my hunch is that those who disagree with Lee on this point will be engaging with him in a terminological dispute. It is strange to say that the embryo has the capacity for rational thought and action while it is still undeveloped; but Lee can draw a distinction between having a certain capacity immediately and having that same capacity remotely. He can then claim that although an embryo doesn't have the capacity for rational thought

---

<sup>94</sup> George and Lee, *Body-Self Dualism*, 44 – 49.

and action immediately, it does have that capacity remotely, in virtue of the fact that, assuming it develops normally, it will have this capacity immediately.

However, I think we should reject Lee's view. To see why, consider the following case

My brain is transplanted into a separate body from which the brain has already been removed. The rest of my body remains alive on external life support. Prior to the transplant, I was injected with a special gene that allows one to spontaneously regenerate lost limbs and organs, just like some invertebrates in the animal kingdom can regenerate certain bodily parts that have been cut off. Unless the living body from which my brain was originally taken is destroyed, it will spontaneously grow a new brain, whose owner will then be capable of rational thought and action.

On Lee's view, in the original version of the brain transplant case, when my brain is transplanted, the rest of my body which is sustained on life support is not a rational animal, and hence, not one of us; it is instead a mere heap of organic tissues, since it completely lacks the capacity for rational thought and action. In the case that we have just considered, this is not true. Thanks to the new gene that was given to me prior to the transplant operation, what is kept alive on external life support after the operation is a being which, on Lee's view, has the capacity (albeit remotely) for rational thought and action. Metaphysically speaking, it is no different from an embryo or fetus that has the genetic capacity to grow a brain, even though it currently lacks one. Someone might claim that the genetically modified brainless being in my example does not have the natural capacity for rational thought and action. But this is not Lee's view, and it seems to have no bearing on whether we should say that this genetically modified being has the remote capacity for



rational thought and action, and hence, no bearing on whether it is what Lee would call a rational animal, as opposed to a mere heap of tissue.

Suppose that this being is (what Lee would call) a rational animal. What should Lee say happens to me in this case? In the original transplant case, Lee's view is that I go with my brain, since it is what realizes my capacity for rational thought and action. But in the case we are currently considering, after I am injected with the regenerative gene, this capacity (which need not be an immediately exercisable capacity) is realized not only by my brain, but also by the rest of my genetically modified body. The brain therefore becomes redundant; just as each of my two kidneys is redundant with respect to my body's capacity to filter out toxins. Thus, one might think Lee should treat my case as an instance of fission, since, after the transplant operation, the capacity for rational thought and action is realized by each of two different physical entities—the brain and the rest of the body from which it is separated. This would be a case of asymmetrical fission in which the two entities resulting from the fission have noticeably different physical qualities, and so are not equally good candidates for being me. Assuming that this is a case of asymmetrical fission, and assuming that I am essentially and most fundamentally (what Lee would call) a rational animal, do I survive in this case? If so, where am I after the transplant operation?

It is not plausible to claim that I cease to exist. Nor is it plausible to claim that I am identical to the genetically modified but brainless being that is kept alive after the transplant operation. Both of these claims are at odds with The Transplant Intuition—the intuition that in the transplant case I survive as the person who inherits my brain. Suppose then that I go with my brain. In that case, I am the rational animal that is pared down to its brain when the brain is separated from the rest of the body, and the genetically modified

being that is kept on life support is a new entity that comes into existence when my brain is separated from the rest of the body. Call this being A1.

Now suppose we add the following detail to the case: The body into which my brain is to be transplanted, call it A2, has been injected with the same regenerative gene that A1 now has, and A2 is qualitatively no different from A1. Thus, although A2 is currently brainless, it will spontaneously regenerate its brain unless given a new one. Clearly, if A1 is a rational animal, A2 is also a rational animal, since A1 and A2 possess the same physical characteristics. Now consider me, the rational animal that is pared down to its brain during the operation. What happens when I am transplanted into the head of A2? One possibility is that this would be an instance of asymmetrical fusion, in which at most one of us (A2 and I) would survive. Again, it is not plausible to suppose that I would cease to exist. Then, assuming A2 and I undergo asymmetrical fusion, A2 must cease to exist. I find this almost impossible to believe. How can a rational animal that would otherwise have spontaneously regrown its brain, cease to exist just by getting a new brain from elsewhere? For a similar reason, I find it almost impossible to believe that A1 comes into existence when my brain is separated from the rest of my body. How can merely separating someone's brain from the rest of their body cause a new rational animal to exist?

Suppose that instead of undergoing fusion, A2 and I coexist in the same body after the transplant operation. In that case, after the operation A2 and I are co-located. We are spatially coincident entities of the same kind. Because we are of the same kind, Lee cannot accept a view akin to what I called The Derivatively Conscious Animal View (DCAV) (Section 3.1.2), according to which I am the non-derivative subject of my experiences while the animal associated with me is conscious only derivatively—i.e., only in virtue of

standing in a certain relation to me, for example by constituting me or having me as a part. I doubt that there is a principled way of claiming that one of two spatially coincident rational animals is the non-derivative subject of experience while the other rational animal experiences only derivatively. Those who defend DCAV can plausibly claim this because on their view the derivative and non-derivative subjects of experience are of different kinds.

### **3.9 Cephalopagus Revisited**

At the end of Chapter 2, I considered a case of conjoined twins involving (what appears to be) two distinct animals that have a single brain as a common part. I suggested that this would give rise to yet another form of the too many subject problem. If indeed there are two animals in this case, Animalism seems to imply that there are two of us, and indeed two human persons if the cerebrum has reached the point at which self-consciousness is possible, despite the fact that there is only one center of consciousness, just as there is in any ordinary human person. For every experience generated by the single cerebrum, Animalism seems to imply that there are two subjects. These are not subjects of different kinds—for example, an animal and a person—but two subjects of the same kind: animals. They are also not spatially coincident but are animals that are physically overlapping in the consciousness-generating regions.

While it seems that Animalists must accept that cephalopagus is a case of two persons who share the same mind, I think, and have argued elsewhere, that it is a case in which the operations of a single cerebrum sustain a single unified and self-conscious mind,

which is, as in all other cases that satisfy this description, the mind of a single person.<sup>95</sup> There is no one else who has this one person's experiences, and no one else who thinks its thoughts. However, this one person is sustained by two overlapping animals. Yet one person cannot be identical with each of two non-identical animals. Since anything that might suggest that one animal was identical with the person would be equally true of the other, there is no reason to suppose that one of the animals but not the other is the person. It seems, therefore, that neither animal is identical with the person. In our hypothetical case of cephalopagus, there are three distinct or non-identical individuals: an individual of our kind—a person who is, we are supposing, made up of the matter in the consciousness-generating areas of the brain—and two animals. If this is the correct description, Animalism is false.

Animalists have shown that theories that imply that we are not identical to but are spatially coincident with entities such as animals or brains face a difficult too-many-subjects problem, and we have now seen that Animalism faces a version of that problem regardless of whether it can avoid the version that arises from countenancing spatially coincident entities of different kinds. But defenders of psychological accounts of our identity have a variety of resources for addressing their version of the too-many-subjects problem. As we saw, they can argue that entities of our kind have mental properties, while animals do not,<sup>96</sup> that only entities of our kind are capable of self-reference,<sup>97</sup> that we and

---

<sup>95</sup> Timothy Campbell and Jeff McMahan, "Animalism and the Varieties of Conjoined Twinning."

<sup>96</sup> Sydney Shoemaker, "Self, Body, and Coincidence."; Sydney Shoemaker, "Persons, Animals and Identity."

<sup>97</sup> Harold Noonan, *Personal Identity* (New York: Routledge, 2003).

our constituting animals are non-identical yet not numerically distinct<sup>98</sup> and that we are parts of our Animals.<sup>99</sup> It is worth noting that there are no parallels to these responses that are available to Animalists in addressing the too-many-subjects problem that arises as a result of certain possible cases of cephalopagus. This is because Animalism's too many subjects are of the same kind rather than spatially coincident entities of different kinds. It does not help, therefore, to claim that only biological continuers can have psychological properties or refer to themselves, since both animals in cephalopagus have biological identity conditions. Nor is it plausible to claim that the two animals in cephalopagus are non-identical but numerically one, or that one, which thinks nonderivatively, is a part of the other, which thinks only derivatively; for each has parts that the other lacks. Eliminativist strategies are also unavailing because the two alleged subjects in cephalopagus are of the same kind; hence, any proposal that would eliminate one would also eliminate the other.

It seems, therefore, that the too-many-subjects problem to which Animalism is vulnerable is more intractable than the one that faces the rival Psychological Approaches. Those who have thought that Animalism is the frontrunner in the debate about our identity on the ground that it is uniquely exempt from this objection appear to be mistaken.

### 3.10 Conclusion and Summary

At the beginning of this chapter (Section 3.1) I stated the four major considerations with respect to which Animalism is supposed to be a better theory than the PA: (1) the too many

---

<sup>98</sup> Baker, *Persons and Bodies: A Constitution View*.

<sup>99</sup> Persson, "Our Identity and the Separability of Persons and Organisms"; McMahan; Parfit, "We Are Not Human Beings."

subjects problem, (2) simplicity, (3) consistency with common sense, and (4) consistency with the most plausible account of unity and composition. The central claim of this chapter is that for each major version of Animalism, TA, that has been proposed, there is a corresponding version of The Psychological Approach, TPsych, such that TA is no better than TPsych with respect to the conjunction of (1) through (4). In this concluding section of the chapter, I will briefly summarize my arguments for this claim, and explain its importance for establishing The Personal Ontology Argument.

I began by drawing attention to a problem that Olson explicitly acknowledges—namely that if The Psychological Approach faces what he calls the vegetable problem, then his version of Animalism—which is also the view of Animalists such as Peter van Inwagen and David DeGrazia—seems to face a parallel problem. According to Olson’s version of Animalism, which I call The Biological Approach —animals have among their essential properties a kind of functional unity of their various parts that is lost at death. If this is true, then, just as a psychological continuer could irreversibly lose its capacity for consciousness, leaving behind a non-conscious but living animal, so an animal could irreversibly lose its capacity for life, leaving behind a corpse. Where The Psychological Approach seems to face the vegetable problem, The Biological Approach seems to face the corpse problem. In both cases, there is significant pressure to say that there are two person-like entities, and that in each case—the vegetable case and the corpse case—the entity that continues to exist was at one time spatially coincident with the entity that ceases to exist. In either case, spatial coincidence leads to the too many subjects problem—there are at least two subjects of consciousness, (two thinkers, two agents, etc.) now sitting in your chair, where common sense implies that there is only one—you. In fact, the too many

subjects problem seems to arise even if one entity is a spatial part or a temporal part of the other since these view seem to imply that the two entities share the same consciousness-generating parts.

Next, in Section 3.2 (3.2.1), I pointed out that, as Olson admits, one particular version of Animalism, which I call The Biological Approach, is not immune to the too many subjects problem. If you are an animal, it seems you could exist as a corpse after your death (i.e. as a dead animal). But according to The Biological Approach this is not the case, since animals have among their essential properties a kind of functional unity of their various parts that is lost at death. For this reason, The Biological Approach implies, animals are essentially alive, and hence, cease to exist at the moment of biological death.

Also in Section 3.2 (3.2.2), I discussed the most common response to the too many subjects problem on behalf of The Psychological Approach, which appeals to the what I call the Derivatively Conscious Animal View—you (the psychological continuer) are the only being now sitting in your chair that has the property of being conscious nonderivatively, while the animal that is intimately associated with you has this very same property only derivatively. You are the only nonderivative subject of your consciousness; the animal is merely the derivative subject of that consciousness. I suggested that perhaps, since common sense recognizes only nonderivative subjects of consciousness, it interprets the claim that there are at least two subjects of consciousness sitting in your chair as the claim that there are at least two nonderivative subjects of consciousness sitting in your chair, which seems absurd, but once we see that there is only one nonderivative subject of consciousness sitting in your chair, the sense in which there are two subjects of

consciousness sitting in your chair may seem not to conflict with common sense, and may seem less absurd.

In Section 3.3, I pointed out that although Defenders of The Biological Approach might follow defenders of The Psychological Approach and claim that the essentially living animal (which, according to them, is what I am) is the only nonderivative subject of consciousness, while the body, which is alive only contingently, is conscious only derivatively in virtue of the relation in which it stands to the essentially living animal. However, as I then pointed out, this is not how Animalists have typically responded to the too many subjects problem that arises for their view. Instead, they have opted for one of two alternative views. One view, Eliminativism, states that we are the only composite entities, and hence, the only person-like entities, that exist. When The Biological Approach is combined with Eliminativism, the resulting view avoids the too many subjects problem. On this view, the essentially living animal now sitting in my chair is the only composite entity, and hence, the only person-like entity now sitting in my chair. The too many subjects problem is avoided by eliminating everything that is plausibly a candidate for being me, with the exception of the essentially living animal. The second view Animalists have proposed, Switcheroo, states that while essentially living animals and bodies exist, they do not coincide spatially or materially; instead, a body supplants an essentially living animal when the latter dies and ceases to exist. Since this view does not posit spatially coincident entities, or entities that share the same consciousness-generating parts, it avoids the too many subjects problem.

In Sections 3.3 and 3.4, I argued that whether Defenders of The Biological Approach accept Eliminativism (as Olson and van Inwagen do) or Switcheroo (as



DeGrazia does), the resulting version of Animalism is no better, with respect to the conjunction of considerations (1) through (4), than some corresponding version of The Psychological Approach. This is because Eliminativism and Switcheroo can be combined not only with Animalism but also with The Psychological Approach. If The Psychological Approach is combined with either of these views, it avoids the too many subjects problem (although, as I pointed out, defenders of The Psychological Approach tend to reject both of these views in favor of the Derivatively Conscious Animal View). I argued that if The Biological Approach is combined with Eliminativism, then with respect to the conjunction of (1) through (4), the resulting version of Animalism (Animalism1) is no better than the view that one gets by combining The Psychological Approach with Eliminativism (PA1). Similarly, I argued, if Animalism is combined with Switcheroo, then with respect to the conjunction of (1) through (4), the resulting version of Animalism (Animalism2) is no better than the view that one gets by combining The Psychological Approach with Switcheroo (PA2). In either case, there is no good reason to accept the version of Animalism that has been proposed.

For example, in order to avoid positing coincident entities, a defender of The Psychological Approach can say either that when one of us (a psychological continuer) irreversibly loses the neural capacity for consciousness, an animal in a permanent vegetative state supplants him in the space he previously occupied (as Switcheroo and The Psychological Approach jointly imply), or that although we (psychological continuers) exist, animals do not. Wherever there appears to be an animal there are only atoms arranged animal-wise (as Eliminativism and The Psychological Approach jointly imply). Both Eliminativism and Switcheroo are radically at odds with common sense, no matter

which personal ontology they are combined with (that is, The Biological Approach or the PA). For this reason, it seems, if each view is combined with Eliminativism or Switcheroo, neither will have an advantage with respect to consideration (3)—consistency with common sense.

Next, since each view—whether it is combined with Eliminativism or with Switcheroo—avoids positing spatially coincident entities of different kinds, or entities of different kinds that share the same consciousness-generating parts, each view avoids the too many subjects problem that arises when one posits such entities. In that case, neither view has an advantage with respect to consideration (1)—the too many subjects problem.

Next, all else equal, neither view will have an advantage with respect to (2)—ontological simplicity. Each will posit the same number of person-like entities. If each view is combined with Eliminativism, then the resulting version of The Biological Approach (Animalism1) posits animals and denies the existence of every other person-like entity that is a plausible candidate for being one of us; and the resulting version of The Psychological Approach (PA1) posits psychological continuers and denies the existence of every other person-like entity that is a plausible candidate for being one of us. Each view recognizes exactly one kind of person-like entity.

If, instead, both The Biological Approach and The Psychological Approach are combined with Switcheroo, then the resulting version of The Biological Approach (Animalism2) posits animals (which, on this view, is what we are) as well as bodies, and it implies that no animal, and hence, none of us, could ever be spatially coincident with a body, and that any animal, and hence, any one of us, would be supplanted at the moment of her natural death by a body. On the other hand, the resulting version of The

Psychological Approach (PA2) posits psychological continuers (which, on this view, is what we are) as well as bodies, and it implies that no psychological continuer, and hence, none of us, could ever be spatially coincident with a body, and that any psychological continuer, and hence, any one of us, would be supplanted at the moment of her natural brain death by a body (this body might be biologically alive for a short time if it is given sufficient external life-support). Now recall, I stipulated that all other ontological commitments of either view (i.e. commitments other than those concerning what we are) are the same. Therefore, overall, each view posits the same number of entities, as well as the same number of kinds of entity.

I also argued (3.3.3) that whether The Biological Approach and The Psychological Approach are combined with Eliminativism or with Switcheroo, the view of composition and unity that The Biological Approach can offer is no better than what The Psychological Approach can offer. I argued that the answer to the special composition question that Olson and van Inwagen favor, namely that the *xs* compose a *y* if and only if the activities of the *xs* jointly constitute a biological life, isn't more plausible than the corresponding answer that a defender of The Psychological Approach might give, namely that the *xs* compose a *y* if and only if the activities of the *xs* jointly constitute a mental life. Rather than thinking "If there are any composite objects at all, then surely there are animals," defenders of The Psychological Approach can claim that we ought to think "If there are any composite objects at all, then surely there are psychological continuers." I doubt whether the Animalist will be able to produce a cogent argument that her answer is the right one. This is one reason to think that considerations of composition and unity do not favor the Biological Approach more than the Psychological Approach.

Next, in Section 3.5 (3.5.1), I showed how the arguments of Sections 3.3 and 3.4 apply to The Bodily Approach as well as to The Biological Approach. Defenders of The Bodily Approach, (such as Feldman, Carter, Mackie, and Ayers, and Williams) claim that I am a body that is alive only contingently, and that could continue to exist as a corpse after my biological death. I argued that these Defenders of The Bodily Approach do not clearly avoid the too many subjects problem. If they claim that the body that I am has certain distinctively human features essentially—that is, if they claim that I am a distinctively human body—then they face a problem which parallels the vegetable problem for The Psychological Approach (if I am a psychological continuer, then apparently I could be outlasted by an animal that is now sitting in my chair) , as well as the corpse problem for The Biological Approach (If I am an animal that is essentially alive, then apparently I could be outlasted by a body that is now sitting in my chair—the body that will one day be my corpse). I called this problem for The Bodily Approach the lump problem. The problem is that there are possible cases in which, apparently something—such as a lump of flesh—persists through a change that causes the distinctively human body to cease to exist. If this is correct, then if I am a distinctively human body, apparently I could be outlasted by a lump of flesh that is now sitting in my chair. (We should have no problem seeing how a lump of flesh could take on a distinctively human form, just as a lump of clay takes on the distinctive form of Socrates in the familiar case of the statue and the lump.) Defenders of The Bodily Approach must therefore struggle, alongside Defenders of The Biological Approach and defenders of The Psychological Approach, in order to avoid the conclusion that there are now at least two subjects of consciousness sitting in my chair—a distinctively

human body (which is what I am) and a mere lump of flesh that could persist after I (the distinctively human body) have ceased to exist.

I argued (Section 3.5.1) that the alternatives available to defenders of the Biological Approach for avoiding the lump problem are very similar to those available to Defenders of The Biological Approach for avoiding the corpse problem, as well as those available to defenders of The Psychological Approach for avoiding the vegetable problem, and that if Defenders of The Bodily Approach adopt one of these alternatives, then for reasons of the sort that I gave in Sections 3.3 and 3.4, the resulting version of their view will be no better than some corresponding version of The Biological Approach, which, as I also argued in those sections, is no better than some corresponding version of The Psychological Approach.

However, I pointed out, there is another alternative available to Defenders of The Bodily Approach. They can claim that the body is a mere lump (i.e., is essentially and most fundamentally a lump) which is not coincident with a substance of any other kind. Since the lump has distinctively human features only contingently, it would not cease to exist if it lost these features. This gets around the lump problem; I cannot be outlived by the lump because I am the lump. However, as I then showed, this isn't a live option for Defenders of The Bodily Approach, since the identity conditions of mere lumps of stuff are clearly different from our identity conditions. A personal ontology that identifies us with mere lumps of stuff is clearly worse, all things considered, than any other version of Animalism (or the Psychological Approach) that we have considered.

I argued against The Hybrid Approach and Patrick Lee's Animalist theory, and that cephalopagus raises another version of the too many subjects problem for Animalism that

does not apply to the Psychological Approach. All things considered, the case against Animalism seems much stronger than the case against the Psychological Approach. We should reject the former for the latter.

In the subsequent chapter, I will consider the moral implications of the Psychological Approach.

## 4. The Psychological Approach and the Moral Status of Human Animals

### 4.1 Introduction to Chapter 4

In Chapters 2 and 3 I argued that the Biological, Bodily, and Hybrid Approaches are false. Henceforth, I shall assume that this conclusion is correct, and that the Psychological Approach is true. As this approach implies, each of us is a psychological continuer of some kind—not an animal. In this chapter, I will argue that the view that we are psychological continuers, when combined with plausible empirical claims about the timing of the onset of consciousness, has two important implications: First, it greatly limits the number of actual cases in which the conclusions of most familiar pro-life arguments apply. The practical significance of these arguments depends on the assumption that each of us began at (or near) the time of the fertilization of the ovum.<sup>100</sup> But The Psychological Approach implies that none of us could have existed prior to the time at which she first acquired the neural capacity for consciousness; and our best empirical evidence indicates that this occurs well after conception.<sup>101</sup> Second, and more important, the Psychological Approach supports the conclusion of the Personal Ontology Argument, which states

---

<sup>100</sup> I say at *or near* the time of conception because some defenders of the BA have argued that a human animal does not begin to exist until gastrulation, which occurs during the fourteenth to the sixteenth day after the fertilization of the egg by the sperm. See, for example, Smith and Brogaard, “Sixteen Days,” 45-78. An argument against Smith and Brogaard’s position is given in Liao, “The Organism View Defended.”

<sup>101</sup> For specific details, see e.g., Julius Korein, “Ontogenesis of the Brain in the Human Animal: Definitions of Life and Death of the Human Being and Person,” *Advances in Bioethics* 2 (1997), 1-74.

Killing a human animal and/or using it as a biological resource is morally justified if this is necessary (or the only reasonable way) to achieve a certain good, the fulfillment of a weighty self-interest of some psychological continuer, and apart from its impact on any human animal, it does not violate anyone's rights or produce any evil that nullifies or outweighs this good.

By appealing to the Psychological Approach, one avoids most of the practical implications of familiar pro-life arguments without directly challenging the moral assumptions on which those arguments depend. The particular stage of human development in which we begin to exist varies depending on which version of the Psychological Approach one adopts. On some versions, the mental properties necessary for one's existence could likely be had only after the birth of the human animal. For example, some defenders of the Psychological Approach have claimed that our essential and most fundamental properties include the capacity to think about ourselves in the first-person.<sup>102</sup> It seems likely that we could develop such a capacity only after birth. Thus, on some versions of the Psychological Approach, it seems likely that one of us does not begin to exist until after the birth of the human organism that precedes this individual. It is irrelevant to the present argument whether any such version of the Psychological Approach is true. However, for obvious reasons, whether or not this is true has important moral implications regarding infanticide. What is relevant is that on *any version* of the Psychological Approach none of us could have existed prior to having the neural capacity for the most rudimentary forms of consciousness.

---

<sup>102</sup> See, e.g., Lynne Rudder Baker, "When does a Person Begin?" *Social Philosophy and Policy* 22 (2005): 25-48.



I will begin by examining the moral implications of the Psychological Approach for certain pro-life arguments against abortion, embryonic stem cell research, and euthanasia. I will show that the Psychological Approach restricts the number of actual cases in which the conclusions of most familiar pro-life arguments apply. In some cases, the Psychological Approach restricts the scope of the argument without directly challenging any of the argument's major moral assumptions. In other cases, it restricts the scope of the argument without challenging *any* of the argument's premises. For each argument, it seems, the range of cases to which the conclusion of the argument applies is miniscule.

Following this, I will present a positive argument, which I call the Symmetry Argument, for the conclusion of the Personal Ontology Argument. The Symmetry Argument seems to depend on the falsity of Animalism and the Hybrid Approach, and therefore seems to establish a link between the Psychological Approach and the conclusion of the Personal Ontology Argument. I then defend the Symmetry Argument against objections. Finally, I make several remarks about how the value of the self-interest of psychological continuers might be weighed against the value of the biological life of human animals.

## **4.2 When do We Begin?**

On any version of the psychological approach, we do not begin to exist until our organisms develop the neural capacity to generate consciousness. When does this happen? Most neurologists accept that the earliest this happens is around the twentieth week of pregnancy,

and that consciousness isn't clearly present until after the twenty-eighth week. Neurologist Julius Korein describes the development of the neural capacity for consciousness:

Neurons in the cortical plate first begin to form cortical synapses at about 20 weeks. These neurons then form synaptic connections between other intracerebral structures such as the thalamus and the brain stem, resulting in sensory reception and more patterned and spontaneous motor activity. Cortical EEG activity can be first recorded at about 21-22 weeks after fertilization; the blink-startle response, with eyes opening, to auditory stimuli can be demonstrated at 24 weeks; and cortical sensory evoked potentials appear at about 25 – 27 weeks. ... Major components of cerebral function including aspects of consciousness (sentience) are unequivocally present in the fetus after 28 weeks of fetal age. The onset of the fundamental core of brain function ... can be identified between the limits of about 20 to 28 weeks.<sup>103</sup>

For now, let us suppose the neural capacity to generate consciousness is formed at some point between the twentieth and twenty-eighth week of gestation; furthermore, let us suppose, somewhat implausibly, that the formation of this capacity is perfectly determinate—*not* vague.

Before proceeding, I should emphasize my use of the term 'neural capacity'. As we saw in the previous chapter, one can argue that even young embryos have the capacity to develop the physical basis for consciousness, albeit remotely, and that the defender of the Psychological Approach must explain why, given her emphasis on the moral importance of the *capacity* as opposed to the *actual manifestation* of consciousness, this isn't sufficient for each of us having once been an embryo. The explanation that I provided appealed to a hypothetical case which, I argued, is most plausibly interpreted as a case in which my functioning brain, A, is removed from a certain human organism, B, while B retains the remote capacity for consciousness, in virtue of B's having the ability to spontaneously regenerate any amount of lost brain tissue. I argued that in this case,

---

<sup>103</sup> Korein, *Ontogenesis of the Brain in the Human Organism*, 25-26.

although B has the remote capacity for consciousness, *I* go with my original brain, A; and, I claimed, this fact gives us at least some reason to believe that whatever sort of continuity is necessary and sufficient for my identity over time includes continuity of the neural basis of my consciousness—that is, neural continuity—and excludes certain kinds of continuity, such as biological continuity, for which neural continuity is neither necessary nor sufficient. I acknowledged that this hypothetical case involving spontaneous regeneration of the brain does not refute versions of The Hybrid Approach on which neural continuity is *sufficient* but not *necessary* for one’s identity over time. However, I went on to present a separate argument, which, I claimed, shows that the Hybrid Approach is less plausible than the Psychological Approach. Unlike The Psychological Approach, The Hybrid Approach must either embrace an especially bad version of the “too many subjects” problem, or embrace an implausible view of the persistence conditions of human animals.

This refutation of the Hybrid Approach is important; it shows that my emphasis of *the neural capacity* for consciousness is not arbitrary, as Patrick Lee and others might be tempted to claim. My use of the term ‘neural capacity’—rather than ‘remote capacity’ or ‘capacity’—isn’t merely a ploy on my part to exclude views of our identity over time that imply that each of us was once a brainless fetus with *some kind of capacity* for consciousness. As we have seen, there is good reason to believe that *the neural capacity* for consciousness is necessary and sufficient for our identity over time—at least in certain cases, including both ordinary run-of-the-mill cases that we encounter in daily life (cases ranging from conception to death) and certain hypothetical but clearly possible cases of brain isolation and transplantation.

### 4.3 The Moral Relevance of the Psychological Approach

If the neural capacity to generate consciousness is formed at some point between the twentieth and twenty-eighth week of gestation, then in more than 99% of actual cases in which a prenatal human organism is killed (or allowed to die), no psychological continuer—none of *us*—is killed (or allowed to die). Consider abortion. In the United States, less than 2% of actual abortions are performed after the 20<sup>th</sup> week of pregnancy.<sup>104</sup> In most of the major European countries, less than 1% of abortions are performed after the 20<sup>th</sup> week. Furthermore, any human animal that is killed for the purpose of destructive embryonic stem cell research is killed while it is still in the embryonic stage—and therefore cannot possess the neural capacity for consciousness. These cases involve killing, but the same point applies to the vast majority of cases in which a prenatal human animal dies without being killed. For example, consider the phenomenon of natural embryo loss or spontaneous abortion, in which a prenatal human animal does not survive to birth. In most cases of spontaneous abortion, death occurs around one week after conception, prior to the time at which the embryo usually implants in the uterine wall.<sup>105</sup> During this early stage, the proportion of surviving embryos drops rapidly. Only about half successfully implant. Spontaneous abortion is an everyday phenomenon. As Toby Ord notes “a mother of three children could be expected to have also had approximately five spontaneous abortions. An embryo’s survival to term is [therefore] the exception rather than the norm.”<sup>106</sup> If these

---

<sup>104</sup> L.T. Strauss, S.B. Gamble, W.Y. Parker, D.A. Cook, S.B. Zane and S. Hamdan, “Abortion Surveillance – United States, 2003,” *Morbidity and Mortality Weekly Report*, 55 (2006): 1-32.

<sup>105</sup> Wilcox et al. 1999.

<sup>106</sup> Toby Ord, “The Scourge: Moral Implications of Natural Embryo Loss,” *The American Journal of Bioethics* 8.7 (2008): 12-19.

dramatic death rates for early embryos remain unknown to the general public, that is likely because most embryo loss occurs before the pregnancy has been detected, and the woman is unaware that anything out of the ordinary has happened. The embryo simply passes out of the uterus with the next menses. On the Psychological Approach, the death of any such embryo would not be the death of one of us. However, there are some cases in which a life-or-death decision for a prenatal human animal may determine whether one of us—a psychological continuer—dies. These are cases in which the human animal in question has formed the neural capacity for consciousness. I will discuss these cases later in the chapter.

Some claim that the non-identity of the psychological continuer and the animal, which is implied by the Psychological Approach, undermines support for a pro-life position. For example, Jeff McMahan claims that on his version of the Psychological Approach, killing a nonconscious prenatal human animal is relevantly like contraception—it does not kill someone but merely prevents someone from coming into existence.<sup>107</sup> Lynne Rudder Baker claims that her favorite version of The Psychological Approach “clears the field of misleading claims about an [early fetus’s] right to life,” since “there can be no right to life unless there is *someone* to be a subject of that right”,<sup>108</sup> and Derek Parfit claims that the Catechism of the Catholic Church, which states that “from the first moment of his existence, a human being must be recognized as having the rights of a person” doesn’t apply to *us*, since, on his favorite version of the Psychological Approach *we* are not human beings (read “human animals”) but Lockean persons—entities that can think

---

<sup>107</sup> McMahan, *The Ethics of Killing*, 267.

<sup>108</sup> Baker, “When does a Person Begin?” 45.

about themselves, and whose continued existence essentially involves psychological continuity.<sup>109</sup>

However, as I will show, the argument from the Psychological Approach to a substantive conclusion about the moral status of nonconscious human animals is not straightforward. For example, as I show in Section 4.3.3, the inference from the premise that none of us was ever an early fetus to the conclusion that abortion is morally similar to contraception, or that no early fetus has a right to life, or that the Catechism of the Catholic Church does not apply to early fetuses (or embryos), can be challenged. Responding to this challenge requires saying more about the relationship between the Psychological Approach and the moral status of nonconscious human animals. As it turns out, the non-identity of the person and the human animal is not sufficient for the conclusion that the human animal lacks the moral status of the person. An additional claim about the relationship between the person and the human animal is needed.

#### 4.3.1 Kind-based Arguments and Identity-based Arguments

The moral relevance of the Psychological Approach is perhaps best demonstrated by considering how it undermines what I will call *kind-based arguments*. Many pro-life arguments rest a great deal of weight on a claim about our identity. For example, many pro-life advocates explicitly argue that it is at least *prima facie* seriously morally wrong to kill *one of us*.<sup>110</sup> These writers typically oppose practices such as abortion, embryonic stem

---

<sup>109</sup> Parfit, "We Are Not Human Beings," 5-28.

<sup>110</sup> Robert George and Christopher Tollefsen, *Embryo: A Defense of Human Life* (New York: Knopf Doubleday Publishing Group, 2008); Francis Beckwith, *Defending Life: A Moral and Legal Case Against Abortion Choice* (New York: Cambridge University Press, 2007).

cell research, IVF procedures, and others that involve the killing of human animals, because they believe that *we* are most fundamentally human animals. But if we are psychological continuers, then even if it is *prima facie* morally wrong to kill one of us, kind-based arguments would not apply generally, since, in most cases, the killing of a nonconscious embryo or fetus would not be the killing of one of us.

The Psychological Approach also seems to have clear implications for what I will call *Identity-based arguments*. Such arguments include at least one premise explicitly stating that I was once a fetus (or that each of us was once a fetus). For example, consider the following argument recently defended by Alexander Pruss:

1. I was once a fetus.
2. If I was once a fetus, then it would have been wrong to kill that fetus.
3. If it would have been wrong to kill me when I was a fetus, then it is wrong to kill anyone when he is a fetus.

Therefore,

4. It is wrong to kill anyone when he is a fetus.<sup>111</sup>

At first glance, it seems as if Pruss's argument is clearly refuted by the Psychological Approach. However, this is not necessarily the case. It is true that on *some versions* of the psychological approach, the first premise of Pruss's argument is false. For on some versions of the psychological approach, the kind of psychological continuity that is

---

<sup>111</sup> Alexander Pruss, "I was Once a Fetus: That is Why Abortion is Wrong." In *Persons, Moral Worth, and Embryos: A Critical Analysis of Pro-Choice Arguments*, edited by Stephen Napier (New York: Springer, 2006).

necessary and sufficient for one's identity over time is too robust to relate one to any fetus. For example, on some versions, the conditions of our identity over time require psychological connections involving memory or sustained character traits that are not plausibly possessed by any fetus or newborn infant.<sup>112</sup> But on other versions of the Psychological Approach, the psychological continuity that is relevant for our identity over time requires only the continuity of those areas of the brain that are involved in the spontaneous generation of consciousness.<sup>113</sup> On this view, it is true that I was once a fetus if the term 'fetus' refers only to a being that originated at some point between the 20<sup>th</sup> and 28<sup>th</sup> week of gestation. If this view were true, then defenders of the Psychological Approach could accept the premises and the conclusion of Pruss's argument, since the term 'fetus', as it appears in the argument, would refer only to beings that begin to exist after the 20<sup>th</sup> week of gestation. In this case, the argument may be not only valid but *sound*; and yet, it would apply to less than 1% of all actual cases of abortion. This is an interesting result. It shows that a defender of the Psychological Approach can accept Pruss's argument and yet also hold that abortion is morally permissible in the vast majority of actual cases. One can accept the argument while rejecting what one would naturally assume are its practical implications.

What I say here in response to Pruss's argument also applies to similar identity arguments against the killing of embryos, as well as identity arguments against the killing of adult human animals in a permanent vegetative state.

---

<sup>112</sup> See, e.g., Parfit, *Reasons and Persons*, Part III.

<sup>113</sup> See, e.g., Unger, *Identity, Consciousness, and Value*, and McMahan, *The Ethics of Killing*.



### 4.3.2 Universalizability Arguments

#### 4.3.2.1 Hare's Golden Rule Argument

In this section, I will show that the conclusion of the previous section—that the Psychological Approach limits the scope of kind-based and identity-based arguments even when it does not directly challenge the moral principles on which those arguments depend—extends to *universalizability arguments*—arguments that appeal to the hypothetical consequences of the universal performance of a certain type of act. I shall focus on the most famous pro-life universalizability argument—R.M. Hare's *Golden Rule Argument* against abortion (GRA).

Hare's argument proceeds as follows:

P1. We shouldn't do to others what we are glad they haven't done to us. (Golden Rule)

P2. I'm glad no one killed *me* when *I* was a fetus.

Therefore,

C. I shouldn't kill those who are fetuses.<sup>114</sup>

Before considering how, and to what extent, the truth of The Psychological Approach would restrict the application of the GRA's conclusion, it will be useful to consider the most common response to the argument, namely, rejecting Hare's formulation of the Golden Rule. One common criticism of Hare's golden rule is that it implies that the class of acts that I shouldn't perform is unacceptably broad.<sup>115</sup> For example, it has been argued

---

<sup>114</sup> R.M. Hare, "Abortion and the Golden Rule," *Philosophy & Public Affairs* 4 (1975): 201-222. Although this argument was originally made by Hare, variations exist; see e.g., Harry J. Gensler, "A Kantian Argument Against Abortion," *Philosophical Studies* 48 (1985): 57-72.

<sup>115</sup> For example, see David Boonin, "Against the Golden Rule Argument Against Abortion" *Journal of Applied Philosophy* 14 (1997) 187-197.

that Hare's golden rule implies that I shouldn't use contraception, insofar as this would prevent the existence of a person. In applying Hare's golden rule, I might reason as follows:

(1) We shouldn't do to others what we are glad they haven't done to us. (Golden Rule)

(2) I am glad that my parents refrained from using contraception, as this would have prevented me from existing.

Therefore,

(3) I shouldn't use contraception when this would prevent someone from existing.

A similar application of Hare's golden rule seems to imply that I shouldn't abstain from sexual intercourse, again insofar as this would prevent someone from existing. We can see how iteration of the rule would have me working overtime to guarantee the existence of as many new persons as possible.

However, on a more plausible understanding of Hare's golden rule, this absurd consequence is avoided. According to the more plausible understanding, I apply the rule via the following procedure: First, identify a range of possible scenarios in which the performance of an act of a certain type *X* would affect *me* in a certain way *Y*. Next, consider whether I am glad that others have refrained (or would refrain) from performing acts of type *X* when this would have affected me (or would affect me) in way *Y*. Finally, if I *am* glad, conclude that I shouldn't perform acts of type *X* when this would affect *others* in way *Y*. When I apply the rule in this way, the possible scenarios that I consider include only those in which *I* am affected by the performance of some act. This excludes any possible scenario in which my parents either used contraception or abstained from intercourse.

Since *I* would not have existed in such a scenario, it follows that any act that might have been performed in such a scenario could not have affected *me*. Thus, I do not conclude that I shouldn't use contraception or that I shouldn't abstain from intercourse. The application of Hare's golden rule *must* be restricted in this way, I believe, in order to avoid the morally absurd conclusion that I shouldn't perform an act if, had someone else performed an act of the same kind, I wouldn't have existed.

However if, when deciding whether I am glad that an act of type X has not affected me in way Y, the possible scenarios that I am allowed to consider exclude those in which I am not (or couldn't have been) affected in way Y, then if any version of The Psychological Approach is true, I cannot consider whether I am glad that no one killed me when I was a fetus *if the fetus whose life depends on my decision lacks the neural capacity for consciousness*. According to The Psychological Approach, I could not have existed without this capacity. Thus, I cannot imagine a possible scenario in which I am killed as a fetus that lacks this capacity. For example, if I couldn't possibly have existed as a 15-week-old fetus (one that lacks the neural capacity for consciousness) then in deciding whether or not to abort a 15-week-old fetus I cannot consider whether I am glad that no one aborted me when *I* was a 15-week-old fetus. This would be like deciding whether or not to break rocks by considering whether I am glad that no one broke me when *I* was a rock. In each case, my decision would rest on the false presupposition that I was (or could have been) a being of a certain kind. On any version of The Psychological Approach, the conditions of my existence will exclude my having been a fetus that lacks the neural capacity for consciousness, just as they clearly exclude my being (or having been) a rock. Since no actual fetus younger than 20 weeks has the neural capacity for consciousness, it

follows that if I am considering whether to have (or perform) an abortion, and the fetus that would be killed by the abortion is younger than 20 weeks, then when I apply Hare's golden rule I cannot consider how an abortion would affect this fetus.

Given the empirical facts about the onset of consciousness, if (any version of) The Psychological Approach is true, in no more than 1% of actual cases would the agent who correctly applied the golden rule conclude that an abortion shouldn't be performed *because of its effect on the fetus*. A possible consequence of The Psychological Approach is that, with respect to many actual decisions regarding abortion, one who correctly applied the golden rule would conclude that one *shouldn't refrain* from having (or performing) the abortion. For example, suppose that a certain pregnant woman's interests would be best served by having an abortion, and that there is more at stake for her than for anyone else who might be impacted by her decision whether or not to have an abortion, *excluding the fetus*. In that case, one can see how, if the pregnant woman and the doctor who is in a position to perform the abortion were to apply the golden rule, while disregarding how an abortion would affect the fetus, the pregnant woman would conclude that she shouldn't refrain from having the abortion and the doctor would conclude that she (the doctor) shouldn't refrain from performing it. If most actual cases were like this, then those who consistently (and correctly) followed Hare's golden rule would *support* abortion in most cases.

The Psychological Approach appears to negate, and maybe even reverse, the *prima facie* practical implications of *the Golden Rule Argument*, perhaps even without directly challenging this argument. For example, versions of The Psychological Approach on which we begin to exist with the onset of consciousness do not challenge any premise of

*the Golden Rule Argument*, but merely restrict the scope of its application. On these versions of The Psychological Approach, the application of the term ‘fetus’ as it appears in the second premise (P2) and the conclusion (C), is restricted to fetuses older than 20 weeks. Thus, one can accept Hare’s golden rule (P1), and one can be glad that one was not killed when one was a fetus (P2), and so conclude that one shouldn’t kill those who are fetuses (C); and yet still deny that in the vast majority of actual decisions regarding abortion, an abortion shouldn’t be performed.

What I say here in response to the golden rule argument against abortion also applies to golden rule arguments against the killing of embryos for the creation of new stem cell lines, as well as alternative methods of engineering mindless human organisms as a biological resource.

### **4.3.3 Future-Like-Ours Arguments**

One of the most well-known anti-abortion arguments is due to Don Marquis. It runs

P1. Killing an individual is *prima facie* seriously morally wrong if such killing deprives it of a future similar to that of an ordinary adult person.

P2. Abortion (a type of killing), in most cases, deprives an individual (a fetus) of a valuable future similar to that of an ordinary adult person.

Therefore,

C. Abortion is, in most cases, *prima facie* seriously morally wrong.<sup>116</sup>

This is commonly called *The Future-Like-Ours-Argument (FLOA)*.

---

<sup>116</sup> See Don Marquis, “Why Abortion is Immoral,” *The Journal of Philosophy* 86 (1989): 183-202.

The plausibility of P2 seems to depend on the assumption that each adult person was once a fetus that eventually *became* that person by becoming first an infant, then a child, then an adolescent, etc. It's hard to see how the FLOA could get off the ground otherwise. As Chris Heathwood says,

It would certainly seem that whether [the FLOA] can succeed depends upon issues of personal identity. For in order for the killing of a fetus to deprive it of a future like ours, it must be the case that, were the fetus not killed, there would later be some future person who is identical to the fetus – the person the fetus would grow up to become. *To put it another way*, killing a fetus can deprive it of a future like ours *only if each of us was once a fetus*.<sup>117</sup>

I suspect that many would share Heathwood's view that in order to deprive a fetus of a future like ours, it must be that each of us was once a fetus. However this isn't obvious, and, as I will now argue, it is incompatible with the commitments of some who defend the Psychological Approach.

#### **4.3.3.1 The Derivative Deprivation Objection: The human animal can be deprived of a FLO in virtue of its relation to the person.**

To see this, recall (Ch. 3) that those who deny the identity of the person and animal must say what the relationship between them is. Some views of the relation between the person and the animal lead to *the too many subjects problem*—if I (the person) am a subject of experience, and the animal associated with me is a subject of experience, and I am not

---

<sup>117</sup> See Chris Heathwood, "The Significance of Personal Identity to Abortion," *Bioethics* 25.4 (2011): 230. italics added. See also David DeGrazia, "The Harm of Death, Time-Relative Interests and Abortion," *The Philosophical Forum*, 38 (2007): 57–80. DeGrazia states: "In asserting that a fetus has a future like ours, the [FLOA] assumes that a fetus can grow into a paradigm person; conversely, it assumes that *each of us was once a fetus*." p. 69, italics added.

identical to this animal, then there are at least two subjects of experience where I am located—me (the person) and the animal.

Recall that the most common response to this problem for defenders of the Psychological Approach is to claim that the animal is conscious only *derivatively*—only in virtue of standing in a special relation of dependence to something that is conscious independently or “nonderivatively.” This is what I called *the Derivatively Conscious Animal View* (DCAV) (Ch. 3, Section 3.3.2). On some versions of DCAV, the animal is conscious in virtue of constituting me—a nonderivatively conscious person; on some versions, it is conscious in virtue of having me as a part. Recall McMahan’s analogies involving a tree that grows in virtue of having a part (a branch) that grows and a car that is noisy in virtue of having a part (a horn) that is noisy:

Suppose ... that over a certain period of time the only part of a tree that grows is a particular limb. ... A property of the part—growth—is in this instance necessarily a property of the whole. There are thus two things that are growing: the limb and the tree of which it is a part. Similarly, when I blow the horn in my car, the horn makes a noise but so does the car. There are two things that have the property of emitting a noise: the horn and the car of which it is a part. ... There is only one noise; and there is a clear sense in which there is only one noisemaker: the horn. But we attribute the making of the noise not just to the horn but also, in a derivative way, to the larger whole [of which it is part]. It is clear, however, that the car is not some additional occult presence that mysteriously joins the horn in producing the honking noise. Nor is the animal as a whole involved in the experience of consciousness except by [having as a part] that which is conscious. In the same sense in which the tree grows because its limb does, and in which the car honks because its horn does, the animal may be said to think, feel, and perceive, because I do.<sup>118</sup>

But a proponent of the FLOA may point out that if these analogies are accurate, then the animal associated with me *has my experiential properties*, if only derivatively. Moreover, she may point out that the animal would not have had my experiential properties if it had

---

<sup>118</sup> McMahan, *The Ethics of Killing*, 92 – 93.

died before I existed, and that, for this reason, dying before I existed would have deprived the animal of “all the experiences and enjoyments that otherwise would’ve been parts of [its] future.”<sup>119</sup> Thus, a defender of the FLO Argument can say that if the DCAV is correct, then, in most cases of abortion, killing a non-conscious fetus deprives it of a FLO, and that therefore, in most cases, killing a non-conscious fetus is *prima facie* seriously morally wrong. Let us call this objection to defenses of abortion predicated upon the non-identity of the person and animal *The Derivative Deprivation Objection*.

The Derivative Deprivation Objection seems to show that from

- (i) None of us was a nonconscious fetus.

one cannot directly infer

- (ii) No nonconscious fetus can be deprived of a FLO.

If the Psychological Approach is true, then in addition to (i) it seems we need some further premise in order to derive (ii). But what premise is needed?

#### **4.3.3.2 A Response to the Derivative Deprivation Objection: Denying that Animals Are Really Conscious**

One possible response is that derivative deprivation doesn’t matter—it just isn’t morally significant. Derivative deprivation isn’t *real deprivation*. Having a FLO matters only when this FLO is had nonderivatively. Having a FLO derivatively, in virtue of standing in some relation to a being that has it nonderivatively, is morally unimportant.

---

<sup>119</sup> Marquis, “Why Abortion is Immoral,” 183–202.



Why? Perhaps the best answer is that something's having a property *derivatively* means that it doesn't *really* have that property. This is compatible with the view that if object x has property F derivatively, attributions of F to x can be true (even though *strictly speaking* x doesn't have F). This is the view of Dean Zimmerman's truth-preserving Cartesian dualist, who

wants to say that ordinary uses of the "to have" of generic exemplification, as in "I have several characteristics in common with that statue", result in true statements, despite the fact that the thing referred to by "I" in some sense lacks all the empirical properties of the statue. The truth-preserving Cartesian must claim that the truth-conditions for many statements involving the "have" of property ascription allow for the borrowing of properties by things that, in some sense, don't really have them, but merely stand in certain relations to things that do.<sup>120</sup>

If the animal associated with me has my experiential properties only derivatively, and if to have a property derivatively (as opposed to having that same property nonderivatively) is to not really have it at all, it is clear how the Psychological Approach bears on the morality of killing nonconscious fetuses and embryos. For, given these assumptions, if the Psychological Approach is correct, then not only was I never a nonconscious fetus, but the nonconscious fetus that my mother carried in her womb would never *really* have experiences. If that is true, then it is hard to see how this fetus could ever have been deprived of a FLO. So perhaps the additional premise we are looking for is this: *If none of us was ever a nonconscious fetus, then no nonconscious fetus could ever really become phenomenally conscious.* The argument from the Psychological Approach to the conclusion that no fetus can be deprived of a FLO now seems straightforward: First, the Psychological Approach and the empirical facts regarding the timing of the onset of consciousness jointly imply that none of us was ever a fetus. Next, the crucial premise: If

---

<sup>120</sup> Zimmerman, "The Constitution of Persons by Bodies," 315.

none of us was ever a fetus, then no nonconscious fetus (no human animal, in fact) could ever really experience. Next, a plausible claim regarding the relation between experience and FLOs: If no fetus (no human animal) can ever really experience, then no nonconscious fetus (no human animal) can be deprived of a FLO. Hence, the conclusion: No nonconscious fetus (no human animal) can be deprived of a FLO. The FLOA fails.

However, some proponents of the Psychological Approach would want to deny that to have a property derivatively is to not really (strictly, literally speaking) have it at all. Baker, for example, would insist that when a human animal constitutes a person, it *really has* the mental properties necessary and sufficient for being a person; although the animal has these properties only derivatively, it *really* has them.

Similarly, Parfit rejects the view that the animal doesn't really (strictly, literally speaking) have the mental properties of the person. He finds this view needlessly paradoxical, and worries that it may lead to a kind of epiphenomenalism about human animals.<sup>121</sup> If what's *really* conscious isn't the animal but just some part of it, why not think that only the animal's stomach *really* digests, that only its lungs *really* breathe, etc.? Why think the animal *really* engages in any activity that can be wholly attributed to one of its proper parts?<sup>122</sup>

I think the defender of the embodied parts view has some wiggle room here. Perhaps the animal really has the properties of *some* of its parts, but does not really have the mental properties of its conscious controlling part, which, according to the embodied

---

<sup>121</sup> Parfit expressed this worry to me in conversation.

<sup>122</sup> Even if, at the end of this process, there were some properties that could only be attributed to my organism as a whole, we can expect this list of properties to be short. So we'd still end up with a view on which organisms were *mostly* epiphenomenal.

parts view, is what *I* am. Consider the following example. Suppose that I park my car on a steep hill and forget to engage the parking brake. As a result, my car begins rolling down the hill. In this case, I can truly shout ‘My car is rolling down the hill!’ even though, strictly and literally speaking, only the tires are rolling. This is similar to McMahan’s example of the noisy car and horn; in both examples, we can say that the car does something in virtue of having a part (or parts) that directly do it. But in my example it is clear that the car *does not literally do what its parts do*. Yet, the car could literally have other properties, such as speed, aesthetic beauty, size, shape, and mass, nonderivatively. There are some properties that the car really has and some that it doesn’t really have.

Alternatively, we might say that although my car literally rolls down the hill, the sense in which it rolls is different from the sense in which its tires roll. For example, the tires roll in the sense of moving along the ground by rotating around their respective central axes, whereas my car rolls in the sense of moving along the ground by resting on objects that are rolling in the first sense. One need not deny that both senses of ‘rolls’ are literal. Similarly, perhaps, I experience in the sense of being composed only of particles whose activities jointly constitute experience, whereas the animal experiences in the sense of having a part that experiences in the first sense. If the relation between me and the animal is construed in this way, it is clear that the animal is not an independent subject of experience.

Even if animals do not really have phenomenal properties, they might really have other mental properties, such as the property of being able to think (where thinking is unaccompanied by the phenomenal feel of thinking) and understand. If, as some philosophers of mind claim, thought and understanding can be adequately characterized

without reference to phenomenal feeling, then there may be nothing objectionable about the claim that the animal really thinks and understands (e.g. language) but lacks phenomenal feeling.<sup>123</sup>

Is it so hard to believe that animals don't really have phenomenal properties? One might argue that we should reject this view because it contradicts what biologists say. But this is an appeal to inappropriate authority. Cartesian dualism may contradict what most modern evolutionary biologists *say* about the mental capabilities of animals, but it doesn't contradict *modern evolutionary biology*. The empirical evidence and the scientific practices involved in the field of modern evolutionary biology are completely independent of the truth or falsity of Cartesian dualism. The same seems true of materialist versions of the embodied parts view on which the animal doesn't really have the phenomenal properties of its conscious controlling part.

Materialists might worry that the view is too similar to dualism. But this worry is unfounded. The view that I am the conscious controlling part of a human animal—and that this human animal is not *really* conscious is wholly unlike dualism with respect to what materialists find objectionable about that position—namely its rejection of materialism.

I think the claim that *strictly and literally speaking* human animals (biological organisms of the species *Homo sapiens*) lack phenomenal properties is less troubling than it seems. Philosophical investigation often overturns claims that appear to be true. In his book *The Nature of the Physical World*, Arthur Eddington claims that because physics tells

---

<sup>123</sup> Ned Block, for example, holds this view. See Ned Block, "The Mind as the Software of the Brain." In *An Invitation to Cognitive Science*, ed. Daniel N. Osherson, Lila Gleitman, Stephen M. Kosslyn, S. Smith and Saadya Sternberg (Cambridge: MIT Press, 1995) 170-185.

us that matter is mostly empty space, it is, strictly speaking, false that his table is solid.<sup>124</sup> One might challenge this claim and insist that the sentence ‘Eddington’s table was solid’ is, strictly speaking, true. This person might say that physics has revealed only that we were wrong about what it takes for an object to be solid, or, perhaps, wrong about *the sense in which* certain objects are solid.<sup>125</sup> But one might also say, reasonably, that although objects such as tables aren’t really solid, sentences such as ‘Eddington’s table was solid’ are true; this view is similar to the “truth-preserving” version of Cartesian dualism considered earlier, according to which attributions of certain physical properties, such as the property of being six feet tall, to a wholly immaterial being, can be true even if this being doesn’t *really* have such properties. Those who deny that the table *really is* solid should at least say that it is *correct* to say of the table that it is solid. There is no good reason to stop saying things like this or to correct those who do. Similarly, there may be no good reason to ban attributions of phenomenal consciousness to human animals from our ordinary discourse, even if human animals aren’t really phenomenally conscious. It is only when doing philosophy that one has good reason to take seriously the possibility that the human animal sitting in one’s chair isn’t *really* phenomenally conscious.

A lot more could be said about this issue, but I will leave it here. As we will see in the next section, it may be possible to refute the Derivative Deprivation Objection without denying that human animals *really* are phenomenally conscious.

---

<sup>124</sup> Arthur Stanley Eddington, *The Nature of the Physical World* (New York: Macmillan, 1928).

<sup>125</sup> Sider discusses these different views in “The Evil of Death: What Can Metaphysics Contribute.” In *The Oxford Handbook of Philosophy and Death*, ed. Ben Bradley, Fred Feldman, and Jens Johansson (New York: Oxford University Press, 2013) 155-166. The example of Eddington’s attitude toward his table is borrowed from page 157.

#### 4.3.3.2 Can a Human Animal be Deprived of a Future Like Ours While it is Associated with a Psychological Continuer?

Perhaps the Derivative Deprivation Objection can be refuted by appealing to *moral intuitions* about cases involving the separation of the psychological continuer and the animal. For example, consider

##### **Body-Replacement**

I have a moderately debilitating condition that is expected to cause me to undergo severe mood swings, abdominal swelling, back pain, and other unpleasant effects over the next nine months. The condition also carries a very small but non-trivial risk of death. There are no expected compensating benefits—nothing that can be expected make up for the nine months of unpleasantness that I am expected to endure. I can avoid this unpleasantness only by undergoing body-replacement—a procedure in which one's (fully functional) brain is transplanted into the head of a different human animal—one that has been genetically engineered to develop without a brain. In my case, the empty-headed animal into which my brain would be transplanted has been genetically cloned from a sample of my DNA. Many years ago, I paid a research group to create this animal and maintain it, on the off chance that I would later need one or more of its body parts. Although body-replacements used to be costly and dangerous, they are now fast, safe, and cheap. If I undergo body-replacement, the human animal that is now sitting in my chair will be killed, but I will be healthy and pain-free. After the procedure, I will be intimately related to a new human animal that is exactly like the one now sitting in my chair. I will

be related to the former as I am currently related to the latter. Furthermore, my psychology will be preserved.

Next, consider two possible outcomes:

***Replacement:*** I undergo body-replacement, and avoid the nine months of unpleasantness and the non-trivial risk of death. The human animal that is now sitting in my chair is killed in the procedure.

***No Replacement:*** I do not undergo body-replacement, but instead endure the nine months of unpleasantness. The human animal that is now sitting in my chair is not killed. It remains associated with me as usual.

If *Replacement* obtains, the animal now sitting in my chair will be killed. Does that mean that it will also be deprived of a Future Like Ours (FLO)? If this animal is killed, then it will no longer have my (or any) mental properties. On the other hand, if *No Replacement* obtains, the animal will continue to exist and will have all of my mental properties. If the animal continues to live, it will have the remainder of my conscious life. If it dies, it will not have any more conscious life. If my body-replacement deprives the animal of a FLO, and if depriving any individual of a FLO is *prima facie* seriously morally wrong (e.g., if it is just like murdering someone) then my body-replacement is *prima facie* seriously morally wrong; even though I –the psychological continuer – survive. If *Replacement* obtains, the surgeons will save me from nine months of unpleasant pregnancy-like symptoms and a small albeit non-trivial risk of death, but this benefit for me will come at the cost of another individual, the animal, since this individual will be deprived of a future like ours. The surgeons will save one individual from a minor temporary disability only by murdering another individual.

But this conclusion is absurd. It depends on two problematic assumptions. The first is that if *Replacement* obtains, then the animal will be deprived of benefits that *I* would have if *No Replacement* were to obtain. The second problematic assumption is that it would be *prima facie* seriously morally wrong to deprive the animal of those benefits—that this deprivation would count as murder. But if both of these assumptions are correct, then for the purpose evaluating *Replacement* and *No Replacement*, the benefits that I would have if *No Replacement* were to obtain must be counted *twice*—once in order to account for their contribution to the fulfillment of *my* self-interest, and then once again in order to account for their contribution to the fulfillment of *the animal's* self-interest. The Derivative Deprivation Objection therefore requires an absurd *double-counting* of benefits.

Building on this observation, we can give the following argument in response to the Derivative Deprivation Objection:

1. If (i) killing an animal, and thereby preventing it from having (derivatively) the mental life of a person, deprives it of a FLO, and (ii) depriving an animal of a FLO is *prima facie* seriously morally wrong (as wrong as murder), then (iii) the killing of the animal in *Replacement* deprives it of a FLO, and is *prima facie* seriously morally wrong for this reason.
2. But the killing of the animal in *Replacement* *isn't* *prima facie* seriously morally wrong *for this reason*. For that would require an absurd double-counting of the benefits that I would have if *No Replacement* were to obtain.

Therefore,



3. At least one of (i) and (ii) is false: Either killing an animal cannot deprive it of a FLO when killing it prevents it from having (derivatively) the mental life of a person, or depriving an animal of a FLO isn't prima facie seriously morally wrong.

If this argument is sound, then the Derivative Deprivation Objection is decisively refuted. To preserve the objection, one needs to reject either 1 or 2. For reasons just discussed, I do not see how we can reject 2 without embracing an absurd double-counting of benefits. If proponents of the Derivative Deprivation Objection want to reject the argument, it seems they should reject 1.

One way to reject 1 is to appeal to the Hybrid Approach. Recall that on this approach, each human animal begins to exist as a nonconscious zygote or embryo, but once the animal's brain is able to support consciousness, it (the animal) goes where its brain goes. If the Hybrid Approach is true, then I am wrong to assume that the animal now sitting in my chair wouldn't survive my body-replacement; this animal would survive the operation as the individual that inherits my functional brain. Thus, on the Hybrid Approach, 1 (and, specifically, (iii)) involves a false presupposition, namely that my body-replacement kills the human animal associated with me. If my body-replacement doesn't kill this animal, it doesn't deprive this animal of a FLO. But on the Hybrid Approach, abortion *does* kill a human animal. Thus, if the Hybrid Approach is correct, the Derivative Deprivation Objection applies to cases of abortion but not to hypothetical cases of body-replacement, and hence, the argument that I presented in response to the Derivative Deprivation Objection does not succeed. However, as I argued in Chapter 3, the Hybrid Approach is false.

Since they cannot appeal to the Hybrid Approach, those who want to defend the Derivative Deprivation Objection should accept (i) and (ii), but deny that (i) and (ii) jointly entail (iii)—that the killing of the animal in *Replacement* deprives it of a FLO and is *prima facie* seriously morally wrong for this reason. They should argue that *Replacement* is a special case, and that the killing of the animal in this case is importantly different from the killing that occurs in most cases of abortion. They should argue that while preventing a nonconscious prenatal human animal from having (derivatively) the mental life of a person deprives it of a FLO, and is *prima facie* seriously morally wrong for this reason, in *Replacement* either (a) preventing the human animal from having (derivatively) the mental life of a person doesn't deprive it of a FLO, or (b) this *does* deprive the animal of a FLO, but in this case depriving the animal of a FLO isn't *prima facie* seriously morally wrong.

One possible argument for (a) appeals to an externalist view of animal mentality. In Chapter 2, I considered a view of this kind, proposed by Rory Madden. On Madden's view, a human animal can continue to have experiences even after it has ceased to exist—as long as its functional brain is preserved and certain other conditions are met. For reasons I gave in Chapter 2 (Section 2.2.5), we should reject Madden's view.

Another way to reject 1 is to establish a morally relevant difference between killing the animal in Body Replacement and aborting (and therefore, killing) a nonconscious prenatal human animal. Perhaps a crucial moral difference between killing the animal in *Replacement* and killing a nonconscious prenatal human animal is that the animal in *Replacement* is psychologically continuous with an individual that survives the body-replacement, whereas a nonconscious prenatal human animal that is killed in an abortion isn't psychologically continuous with any individual. As we saw in Chapters 1 and 2,

philosophers such as Parfit believe that psychological continuity—not identity—is what matters in survival. Couched in terms of individual self-interest, the view is that psychological continuity—not identity—is what provides the basis for an individual's having a self-interest in future (and past?) benefits and avoidance of future (and past?) harms. In Chapter 1, I assumed

*The Identity View of Self-Interest:*

- (1) S1 has a self-interest at t1 in S2 benefiting at t2(-t3), if and only if (a) S1 exists at t1, and (b) S1 and S2 are numerically identical.
- (2) The strength of the self-interest depends only on the size of the benefit(s).

I have assumed that the Identity View is true of *us*. This assumption implies that if we are psychological continuers, the Identity View is true of psychological continuers. But it does not follow that this view is also true (or false) of *human animals*. Those who accept the Identity View can claim that this view applies only to psychological continuers; but they can also accept a psychological continuity view of self-interest that applies only to human animals. For example, one who accepts the Identity View might accept

*The Psychological Continuity View of Animal Self-Interest:*

- (1) Animal A has a self-interest at t1 in individual S benefiting at t2(-t3), if (a) A exists at t1, and (b) A at t1 is appropriately psychologically continuous with S at t2(-t3).
- (2) The strength of the self-interest at t1 depends only on the size of the benefit(s) and the degree of psychological continuity between A at t1 and S at t2(-t3).

Importantly, the Psychological Continuity View of Animal Self-Interest provides only a *sufficient* condition for an animal's having a self-interest in some individual benefiting. One might accept the stronger view that psychological continuity is *both necessary and sufficient* for an animal's having this self-interest. However, if one accepts this stronger view, one will find it difficult to maintain that in most cases of abortion, killing a nonconscious prenatal human animal deprives it of a FLO and is *prima facie* seriously morally wrong for this reason. If the animal that is aborted and killed is not psychologically continuous with any individual, and if psychological continuity with some individual is necessary for an animal's having a self-interest in some individual benefiting, it follows that the prenatal animal does not have a self-interest in any individual benefiting. But in the moral framework that I assumed in Chapter 1, a framework that views individual self-interest as having fundamental moral importance, when an individual is deprived of a FLO, if this deprivation is to be considered morally significant, it must be understood in terms of the frustration of the self-interest that the individual has in some future individual having all the experiences and enjoyments that are jointly sufficient for a FLO. Thus, if a human animal has no self-interest in a future individual benefiting, then killing this animal cannot deprive it of a FLO. Because The Psychological Continuity View of Animal Self-Interest provides only a *sufficient* condition for an animal's having a self-interest in some future individual benefiting, this view is compatible with the possibility that a nonconscious prenatal human animal has a self-interest in some future individual benefiting, and that in most cases frustrating this self-interest is *prima facie* seriously morally wrong. For example, one who accepts the Psychological Continuity View of Animal Self-Interest can claim that *for animals* psychological continuity is sufficient for self-interest, but identity is

sufficient for self-interest *in the absence of psychological continuity*. (This is reminiscent of the Hybrid Approach discussed in Chapter 3, except that it is a view of self-interest, not identity.)

If the Psychological Continuity View of Animal Self-Interest is true, then for the animal that is killed in Body-Replacement, death may be no worse than ordinary survival. For, although the animal does not actually survive the body-replacement operation, its psychology is carried forward by me—the psychological continuer that survives the operation. One could argue that in this case, because the animal is psychologically continuous with me, and because its being psychologically continuous with me is sufficient (but not necessary) for its having a self-interest in me having certain benefits in the future, killing the animal doesn't actually deprive it of a Future Like Ours (FLO). On the other hand, one could further argue, because no nonconscious prenatal human animal is psychologically continuous with any individual, killing such an animal could deprive it of a FLO. If this is right, then there is an important moral difference between the killing of the animal in *Body-Replacement* and the killing of a nonconscious prenatal animal; only the latter killing can deprive an individual of a FLO. This conclusion refutes the first premise of my argument against the Derivative Deprivation Objection, which states that if killing an animal deprives it of a FLO and such deprivation is *prima facie* seriously morally wrong, then the killing of the animal in *Body-Replacement* deprives it of a FLO and is, for this reason, *prima facie* seriously morally wrong. Have we, at last, a successful defense of the Derivative Deprivation Objection?

No. The problem with this defense is that it succeeds only if human animals' self-interests have moral significance independent of that of psychological continuers'. But if

this is true, then once again, our moral assessments will require an absurd double-counting of benefits. Consider the animal in *Body-Replacement*. If, as the Psychological Continuity View of Animal Self-Interest implies, it has a self-interest in me benefiting after body-replacement, and this self-interest has independent moral significance, then my successful body-replacement is much better than it seems to be. This time, the problem is not that there are too many victims; rather, it is that there are too many *beneficiaries*. My undergoing body-replacement is clearly in *my* self-interest, as it allows me to avoid nine months of unpleasantness and a small risk of death. If my undergoing body-replacement is also in *the animal's* self-interest, and if, as we are assuming, I am not identical to the animal, it follows that *two self-interests* are fulfilled by my undergoing body-replacement. For the purpose of making a correct moral evaluation of the alternatives in this case, the benefits that I would enjoy after my body-replacement must be counted *twice*—once when we account for their contribution to the fulfillment of *my* self-interest, and once more when we account for their contribution to the fulfillment of *the animal's* self-interest. Double-counting benefits seems no less absurd here than in previous discussion.

The problem is worse than it may seem. If the Psychological Continuity View of Animal Self-Interest is true, then the animal now sitting in this chair has a self-interest in *me* benefiting next week—since it is psychologically continuous with me next week. And, clearly, *I* have a self-interest in me benefiting next week. It would seem to follow that there are two individuals now sitting in this chair, each of whom has a self-interest in me benefiting next week. This has absurd consequences. To see this, consider the following case.

*Choice between Lives.* I have never been related in the usual way to any human animal. I started out as a brain in a vat—the product of an expensive genetic engineering experiment. Soon after my creation, I was transplanted into an inorganic non-living body that is mechanically similar to most human animal bodies. Psychologically, I am like most other psychological continuers (those who are intimately related to human animals). Thus, I can think, feel, perceive, study, hold down a job, etc. Until very recently, I have lived a normal, happy life. However, I now have a medical condition that affects my brain and threatens my life.

You are related in the usual way to a human animal, and you have the exact same medical condition I have. Each of us will die very soon if his condition is untreated. Doctors can treat either one of us, but not both. If either of us is treated, his life will be extended by 20 years. The overall quality of remaining life would be the same for either of us. (Of course, if you are treated, then the human animal associated with you will continue to live as well.) Neither of our psychologies would be substantially altered by the treatment.

It seems obvious that the doctors have no more reason to treat you than they have to treat me. But if human animals and psychological continuers both have self-interests, and these self-interests have independent moral significance, then the Psychological Continuity View of Animal Self-Interest and the Identity View of Psychological Continuer Self-Interest jointly support the claim that the doctors have a stronger moral reason to treat you than they have to treat me. On the Identity View of Psychological Continuer Self-Interest, you and I each have a self-interest in being treated; and our self-interests are equally strong.

But on the Psychological Continuity View of Animal Self-Interest, the animal that is associated with you also has a self-interest in you being treated; and this self-interest is as strong as either of our self-interests. I am not associated with any human animal, and hence, no human animal or individual other than *me* has a self-interest in me being treated. If animal self-interest and psychological continuer self-interest have independent moral significance, then it seems the doctors ought to treat you. By treating you, they are treating *two patients*—you and the animal associated with you. If the self-interests of human animals and psychological continuers are equally morally significant (I see no reason to assume that they wouldn't be), treating you rather than me would be like saving two lives instead of one. But this is morally absurd.

Next, consider

*Choice between Pains.* You stand in the usual relation to a human animal. I am a brain in a vat—the product of an expensive genetic engineering experiment. I have never stood in the usual relation to any human animal, and no human animal is (or will be) psychologically continuous with me. Each of us will certainly endure a terrible pain. If neither of us is treated for his pain, then my pain and yours will be qualitatively identical, and so these pains will be of equal intensity and duration. Doctors can treat either you or me, but they cannot treat both of us. Whichever person is treated, the intensity of his pain will be reduced by half, but its duration will not be altered.

On the Identity View of Psychological Continuer Self-Interest, you and I each have a self-interest in being treated; and our self-interests are equally strong. But on the Psychological Continuity View of Animal Self-Interest, the animal that is associated with you also has a



self-interest in you being treated; and this self-interest is as strong as the self-interest that each of us has in being treated. Again, assuming that the self-interests of human animals and psychological continuers are equally morally significant (I see no reason to assume that they wouldn't be), treating you rather than me would be like reducing two separate pains, each of which is just like yours. It would then seem that the doctors have twice as much reason to treat you. But this is morally absurd. It seems obvious that the doctors have no more reason to treat you than to treat me.

Proponents of the Derivative Deprivation Objection can avoid these morally absurd consequences if they deny that a human animal's self-interest has inherent moral significance. But if the self-interest of a fully-developed adult human animal does not have inherent moral significance, then it is hard to see why the self-interest of a nonconscious prenatal human animal has inherent moral significance. Does a human animal's self-interest have inherent moral significance only if the animal is a nonconscious fetus? Surely not. If an ordinary healthy adult human animal's self-interest doesn't have inherent moral significance, then *no* human animal's self-interest has inherent moral significance. And, if no human animal's self-interest has inherent moral significance, then no nonconscious prenatal human animal's self-interest has inherent moral significance. But if this is the case, then why should we think that frustrating a nonconscious prenatal human animal's self-interest is *prima facie* seriously morally wrong?

The problem for the proponent of the Derivative Deprivation Objection is this: If human animals and psychological continuers have independent and inherently morally significant self-interests, and if the Identity View of Psychological Continuer Self-Interest and the Psychological Continuity View of Animal Self-Interest are both true, then our

moral evaluations will involve a kind of bizarre *double-counting*. When one of us is benefited or harmed, the amount of goodness or badness that is added to the world is *twice* what it appears to be, unless the individual that is benefited or harmed isn't associated in the usual way with any human animal—for example, if she is a brain in a vat. If you are a brain in a vat, then usually your pleasurable (or painful) experiences will immediately contribute to the world as much goodness (or badness) as they seem to. (Their immediate contribution may be different from their overall contribution, which may depend on their relation to certain other parts of the world.) But if, as most of us assume, you are not a brain in a vat, then usually your pleasurable (or painful) experiences immediately contribute to the world twice as much goodness or badness as they seem to, since each experience contributes to the fulfillment or frustration of two different self-interests. The proponent of the Derivative Deprivation Objection will want to avoid this morally absurd consequence. She can avoid it by claiming that the animal's self-interest isn't inherently morally significant, and that therefore the fulfillment or frustration of its self-interest contributes no value to the world. When we evaluate your pleasurable and painful experiences, we do not double-count. But in that case, the proponent of the Derivative Deprivation Objection cannot explain why frustrating the self-interest of *any human animal* is *prima facie* seriously morally wrong. The Derivative Deprivation Objection is therefore undermined. The objection is that abortion, in most cases, prevents a human animal from becoming (derivatively) conscious, and hence, deprives it of a future like ours—that is, *frustrates the self-interest it has* in some future individual incurring all of the benefits—all of the experiences and enjoyments—that constitute a typical future like ours. But if a human animal's self-interest isn't inherently morally significant, this is hard to believe.

To be clear, I am not denying that there is any morally relevant asymmetry between abortion and body-replacement that would make the former but not the latter *prima facie* seriously morally wrong. (I will return to this issue in Section 4.4.) My aim here is to refute the Derivative Deprivation Objection.

I have argued in this section that from (a) none of us was ever an early fetus (a non-conscious prenatal human animal) one cannot directly infer (b) no such fetus could be deprived of a future like ours (FLO). Some proponents of the Psychological Approach who claim that none of us (no psychological continuer) was ever an early fetus also claim that an early fetus can go on to constitute, or have as a part, a psychological continuer. In that case, they argue, the psychological continuer would have certain mental properties nonderivatively, while the human animal (which was once an early fetus) would have these same mental properties derivatively—this is what I called *the derivatively conscious animal view* (DCAV). A proponent of the future-like-ours argument (FLOA) can argue that if DCAV is correct, then an early fetus can be deprived of a FLO, since it can have (derivatively) all of the mental properties that are sufficient for having a FLO. This is the Derivative Deprivation Objection. This objection illustrates that the inference from (a) to (b) is invalid.

However, I argued, the objection can be met. I argued that if we are not human animals, then either no human animal can be deprived of a FLO, or such deprivation isn't wrong. To show this, I considered a case in which one of us, a psychological continuer, survives a hypothetical surgical procedure—body-replacement—which kills the human animal associated with him. I argued that killing this animal isn't *prima facie* seriously morally wrong, even though killing the animal prevents it from having (derivatively) the

experiences of the psychological continuer. On this basis, I concluded that killing the animal either doesn't deprive it of a FLO (that is, doesn't frustrate any self-interest that it has in a FLO) or deprives it of a FLO but isn't, for this reason, *prima facie* seriously morally wrong. In either case, the Derivative Deprivation Objection is undermined. The objection depends on a false assumption that is the conjunction of two claims: (a) preventing an animal from having (derivatively) the valuable experiences of a psychological continuer deprives it of a FLO and (b) depriving an animal of a FLO is *prima facie* seriously morally wrong.

Next, I considered some possible defenses of the Derivative Deprivation Objection. I argued that we should reject these defenses, since, if they succeeded, our moral assessment of certain outcomes would involve a morally absurd double-counting of benefits; we would count certain benefits *once* in order to determine their contribution to the fulfillment of the self-interest of the psychological continuer, and then again in order to determine their contribution to the fulfillment of the self-interest of the human animal. We should reject this morally absurd consequence.

I conclude that the Derivative Deprivation Argument fails, and that the Psychological Approach undermines the Future Like Ours Argument (FLOA), or at least restricts the scope of its application to late abortion.

My argument against the Derivative Deprivation Objection raises two important questions. First, should we say that killing a human animal doesn't deprive it of a FLO—that is, doesn't frustrate its self-interest in a FLO? Or should we say that although killing a human animal does deprive it of a FLO—does frustrate its self-interest in having a FLO—this is not *prima facie* seriously morally wrong? I think we should say the former. Killing

a human animal doesn't frustrate its self-interest in having a FLO. A possible explanation of this was considered in Section 4.3.3.2. It is that human animals—even ordinary, fully-developed, fully-responsive human animals—aren't *really* phenomenally conscious, even if sentences that attribute phenomenal consciousness to human animals can be *true*. This would provide the clearest explanation of why it is impossible to frustrate a human animal's self-interest. If no human animal can be phenomenally conscious, then, it seems, no human animal has a self-interest in anything. Human animals aren't *bearers* of self-interest.

An alternative explanation of why human animals aren't bearers of self-interest is that a fully-developed rational human animal would be fundamentally concerned only about the psychological continuer associated with it. Harold Noonan argues that a human animal cannot think or talk about itself in the first-person because all of its first-personal thoughts and utterances refer to the psychological continuer associated with it. If this is true, then perhaps no human animal is really capable of first-personal expressions of self-concern. One might argue that nothing that is incapable of first-personal expressions of self-concern is a bearer of self-interest. I am not committed to this claim, nor will I attempt to defend it; but I mention it here as a view that is worth exploring.

Finally, we might think it is just a primitive fact that human animals aren't bearers of self-interest. For example, when we recoil from the thought that my benefits and harms figure twice in moral evaluations, we might take this as conceptual evidence for the claim that there is only one kind of thing currently sitting in my chair that is a bearer of self-interest. Surely, if anything currently sitting in my chair is a bearer of self-interest, it is whatever *I* am. Since I am a psychological continuer, the only thing currently sitting in my

chair that is a bearer of self-interest is a psychological continuer. The human animal that is sitting here with me has no self-interest.

My argument against the Derivative Deprivation Objection raises another question. Earlier in this section, I considered a hypothetical case, *Choice between Lives*, in which doctors can either give a certain benefit to me—a psychological continuer that has never been related in the usual way to a human animal—or you—a psychological continuer that is related in the usual way to a human animal. Giving this benefit to you would preserve the life of the human animal associated with you, but giving it to me wouldn't preserve the life of any human animal. I argued that this wouldn't tip the balance of reasons in favor of giving the benefit to you. Extending the biological life of the human animal associated with you wouldn't even be a tie-breaking consideration. Does this imply that there is never a non-instrumental moral reason to extend the biological life of a human animal? Does biological human life have no intrinsic value whatsoever? I will address these questions in section 4.5, after presenting and defending (4.4) my argument for Premise 2 of the Personal Ontology Argument. But first (4.3.4.) I will consider the relevance of the Psychological Approach for another group of popular arguments against abortion and destructive embryonic stem-cell research—*Potential Person Arguments*.

#### **4.3.4 Potential Person Arguments**

In Section 4.3.2.1, I showed that the Psychological Approach greatly restricts the scope of certain pro-life arguments that explicitly claim that each of us is numerically identical to some past fetus or embryo (identity-based arguments and kind-based arguments) without directly challenging the premises of those arguments. In this section, I will argue that the

same is true of most pro-life arguments that appeal to the moral importance of the fetus's *potential to become* a person. I call these *Potential Person Arguments*. Such arguments seem reasonable only if the morally relevant sense in which the fetus can *become* a person assumes that it would *be* that person, i.e., that the fetus and the later person would be *one and the same thing*. Without this assumption, potential person arguments seem vulnerable to seemingly devastating objections. For example, one common objection to the claim that we shouldn't kill beings with the potential to become persons is that it implies that we shouldn't kill pairs of sperm and ova that have the potential to fuse, possibly leading to the existence of a new person. Although the sperm-ova pair wouldn't literally *be* a person (both sperm and ova would cease to exist before the arrival of the person), there is a sense of 'potential' and a sense of 'become' according to which the sperm-ova pair has the potential to become a person. For example, this pair of cells can *engender* a person through a continuous biological process initiated by their fusion. Similarly, unless the relevant notion of potential is qualified, there may be no plausible way of avoiding the absurd conclusion that it would be morally wrong to destroy objects that normally lack the potential to become persons (e.g., rocks) should they become magically endowed with such potential.<sup>126</sup>

If the relevant notion of potential presupposes that the being with potential to become a person would be numerically identical to the person that would exist once this potential was fully realized, then it seems possible to avoid these morally absurd

---

<sup>126</sup> Michael Tooley, "Abortion and Infanticide," *Philosophy and Public Affairs*, 2 (1972) 37-65. Michael Tooley's well-known criticism of arguments based on the potential of the fetus to become a person depends on examples similar to this. Tooley's famous case involves not an inanimate object but a kitten injected with a chemical that will cause it to develop the mental properties of an ordinary human person.

conclusions. For example, on any plausible view of our identity over time, the conditions of one's identity over time will preclude one's having been a sperm-ova pair or some inanimate object that might become magically endowed with the potential for personhood. However, if the relevant sense of potential is qualified in this way, then for the reasons that we have considered, on any version of The Psychological Approach, the moral conclusions of potential-person arguments will no longer apply to the vast majority of actual abortions or to *any* embryonic stem cell research.

#### **4.4 Establishing the Conclusion of the Personal Ontology Argument**

##### **4.4.1 The Symmetry Argument**

So far, I have argued that The Psychological Approach to personal identity restricts the scope of application of certain arguments against abortion and stem-cell research, or undermines certain key premises in these arguments. In this section, I argue directly for the conclusion of the Personal Ontology Argument. Since this argument depends on the falsity of Animalism and the Hybrid Approach, it is also an argument for Premise 2 of the Personal Ontology Argument, which states

If we are psychological continuers, then killing a human animal and/or using it as a biological resource is morally justified if this is necessary (or the only reasonable way) to achieve a certain good—the fulfillment of a weighty self-interest of some psychological continuer—and, apart from its impact on any human animal, it does not violate anyone's rights or produce any evil that nullifies or outweighs this good.

Recall my hypothetical *Body Replacement* in which replacing all but the consciousness-generating parts of my brain with inorganic substitutes is the only reasonable way to avoid



nine months of unpleasantness and a non-trivial risk of death. I have argued that my body-replacement would kill the human animal associated with me, and that this wouldn't be *prima facie* seriously morally wrong on the grounds that it deprives this animal of a future like ours. I will now present an argument that develops this line of reasoning further.

I claimed that killing the animal by performing a body-replacement isn't *prima facie* seriously morally wrong *for the reason that it deprives the animal of a future like ours*. But I also accept the stronger claim that killing the animal by performing a body-replacement isn't *prima facie* seriously morally wrong. Certainly, the killing of the animal wouldn't be an instance of murder. One might think that in *Body Replacement* what is at stake for me—avoiding nine months of moderate discomfort and a small risk of death—is too insubstantial to justify killing the animal. But surely, if the stakes are high enough, the killing would be justified. For example, if the moderately debilitating condition I described was expected to be permanent rather than temporary, and if killing the animal was the only reasonable way to prevent me from remaining in this condition, then killing the animal would be justified. If the benefits that body-replacement offers are substantial enough, if my self-interest in having a body-replacement is weighty enough, the doctors ought to give me one. There would be exactly *one* beneficiary of the operation, and there would be *no* victim.

It is epistemically possible, albeit unlikely, that Body Replacement actually obtains. Let us assume for the sake of argument that it does. Taking this assumption for granted, we should accept

***The Symmetry Argument:***

P1. My body-replacement kills a certain human animal when it is a healthy young adult; this animal was once a mindless fetus (and, before that, an embryo) and it could have existed in a permanent vegetative state.

P2. Killing the human animal when it is a healthy young adult is morally justified because this is (i) necessary (or the only reasonable way) to achieve a certain significant good—the fulfillment of a weighty self-interest of some psychological continuer—and (ii) apart from its effect on any human animal, it doesn't violate anyone's rights or cause any evil that nullifies or outweighs this good. (I assume (i) and (ii) are independently necessary and jointly sufficient to justify the act.)

P3. If killing the human animal when it is a healthy young adult is morally justified in virtue of (i) and (ii), then killing the same human animal in any of its other possible histories (for example, killing it as an early fetus or embryo, or as a patient in a permanent vegetative state, etc.) is morally justified, as long as this killing also has features (i) and (ii).

Therefore,

P4. Killing the same human animal in any of its other possible histories (for example, killing it as an early fetus or embryo, or as a patient in a permanent vegetative state, etc.) is morally justified, as long as this killing has features (i) and (ii).

P5. If killing the same human animal in any of its other possible histories is morally justified, as long as this killing has features (i) and (ii), then killing *any* human animal in any of its possible histories is morally justified, as long as *that* killing also has features (i) and (ii).

Therefore,

C. Killing any human animal in any of its possible histories is morally justified, as long as this killing has features (i) and (ii). In other words, killing a human animal is morally justified if this is necessary (or the only reasonable way) to achieve a certain significant good—the fulfillment of a weighty self-interest of some psychological continuer—and, apart from its effect on any animal, this killing does not violate anyone’s rights or produce any evil that nullifies or outweighs this good.

The conclusion of the Symmetry Argument is almost exactly the same as the conclusion of the Personal Ontology Argument. The only difference between the two conclusions is that the conclusion of the Personal Ontology Argument refers to the *killing and/or use* of human animals, whereas the conclusion of the Symmetry Argument refers only to the *killing* of human animals. I address the issue of using human animals in section 4.4.4.

The Symmetry Argument is extremely theoretically cost-efficient. Its moral implications are wide-reaching; it provides significant justificatory support for abortion prior to the 20<sup>th</sup> week of pregnancy; selective abortion; in vitro fertilization (involving the creation of multiple human embryos, some of which are caused to die, often in ways that involve killing as opposed to merely letting die); the killing of embryos for the purpose of creating new stem cell lines to further scientific research into finding new treatments and cures for diseases; the active euthanasia of, and harvesting of organs from, human animals that are brain dead or in a permanent vegetative state; the engineering and use of brainless human animals for the purpose of organ harvesting (known as “spare parts” cloning); the

deprioritization of health interventions targeted at preventing prenatal deaths;<sup>127</sup> and potentially life-saving whole body (or head) transplants akin to those discussed in Chapters 1 and 2, the first of which may happen as early as 2017! Yet, despite the broad range of issues that it touches, the Symmetry Argument has relatively few constraining commitments. It doesn't depend on a view of personhood, or the claim that what morally distinguishes non-conscious fetuses and embryos from normal adults is the latter's possession of conscious interests. It doesn't depend on a general theory of either the badness of death or the wrongness of killing. And it is compatible even with the claim that all living human animals have rights (although, for reasons that I consider later in the chapter, I think it is implausible to attribute rights to human animals if the Symmetry Argument is sound).

Some potential objections to the Symmetry Argument have already been considered and refuted in earlier sections. For example, I have already considered the view that killing a human animal when it is a non-conscious fetus deprives it of a future like ours, in virtue of the fact that if it were allowed to develop normally it would have (derivatively) the mental properties sufficient for a FLO.

---

<sup>127</sup> An important issue regarding prioritization for which personal identity seems especially relevant concerns how to incorporate deaths near the time of birth into estimates of the global burden of disease. The loss of future life years or disability-adjusted-life-years (DALYs) for a significant number of stillbirths are ignored, and so don't contribute to the recorded global burden of disease. Therefore, they aren't incorporated into cost-effectiveness estimates of health interventions targeted at preventing stillbirths. However, infant mortality currently counts as a loss of the infant's entire life-expectancy i.e. the total number of expected DALYs lost in its death. This practice of ignoring the burden of stillbirths while giving infant mortality its full weight seems oddly inconsistent, and can potentially lead to undesirable and seemingly unprincipled discontinuities in the prioritization of health interventions (see Dean Jamison, et al., "Incorporating Deaths Near the Time of Birth into Estimates of the Global Burden of Disease" in *Global Burden of Disease and Risk Factors*, ed. AD Lopez, CD Mathers, M. Ezzate, D. Jamison, and C. Murray (Washington, D.C.: World Bank, 2006), Chapter 6. The Psychological Approach would appear to help mitigate this problem, since it provides a justification for discounting the contribution to the burden of disease from prenatal death occurring prior to or even during synapse validation, even if the contribution of infant mortality is given its full weight.

We have also considered what one might call *the Unrealized Potential Objection*. In Body-Replacement the human animal is killed only after its potential has been fully realized. But when this same human animal was still a nonconscious fetus, its potential hadn't yet been fully-realized. Therefore, killing it *then* would've been wrong, even though killing it in *Body-Replacement* isn't wrong. This assumes that although killing a nonconscious human fetal animal or embryo is inherently wrong, once the animal becomes conscious, killing it is no longer inherently wrong. This is a very strange view. Granted, it is sometimes morally worse to destroy a thing *before* its potential has been realized rather than after. It would be wrong of me to destroy your birthday cake by intentionally overheating the oven and burning it to a crisp, even if it would not have been wrong for you and I to later eat, and therefore destroy, the same cake. But the claim that killing a human animal becomes *less* morally problematic as it develops seems ludicrous. If anything, the opposite is true. Simply put, the unrealized potential objection depends on an implausible view of the moral significance of potential.

However, there seems to be one important moral difference between the consequences of killing a human animal in early abortion, and killing the human animal in Body-Replacement—a difference that concerns the potential that an early fetus has to engender or bring into being a person (a psychological continuer). In Body-Replacement, when the human animal is killed, neither the total number of psychological continuers nor the total number of persons changes, whereas in early abortion, this number may change. For example, if the mother of the early fetus would have *one less child* as a result of having an early abortion, and her decision to have an abortion does not significantly alter anyone else's decision whether to procreate or have an abortion, then, as a result of the mother's

decision, the overall number of psychological continuers and the overall number of persons may be less than it would have been if the mother had instead decided to carry her fetus to term. For this reason, a normal, healthy, nonconscious fetal human animal has instrumental value that an ordinary, fully-developed, (derivatively) conscious human animal does not have—since the former but not the latter has the potential to alter the total number of psychological continuers, as well as the total number of persons. It is worth nothing that, at least in certain cases, the potential of an early embryo to engender a psychological continuer (and hence, a person) would not have much practical importance. For example, almost all embryos that are not destroyed for embryonic stem cell research are stored indefinitely, and few if any will be donated to fertile couples or single women who wish to have children. Thus, while a change in the number of psychological continuers and in the number of persons may mark a very important difference between ordinary cases of early abortion and Body-Replacement, currently it does not mark a very important difference between destructive embryonic stem cell research and Body-Replacement. I will address this objection near the end of the chapter.

#### **4.4.2 The Consent Objection**

One might raise the following objection to the Symmetry Argument's third premise: Killing the human animal in *Body-Replacement* is morally justified because the animal voluntarily consents to the body-replacement, and therefore *waives* its right not to be killed as a result of the body-replacement. But no embryo or fetus consents to abortion or to be used in scientific research. Thus, from the premise that killing the human animal in Body-Replacement is morally justified, one cannot infer the conclusion that killing a human

animal in any other circumstances, including those of early abortion or embryonic stem cell research, would be morally justified. Call this *the Consent Objection*.

This objection is intriguing, but also very odd. For example, for those on either side of the abortion issue, the rightness or wrongness of abortion does not depend significantly on lack of consent, nor is it plausible to suppose that it does. In general, the fact that one has consented to being killed isn't sufficient to justify killing one. One can't legitimately waive one's right not to be killed *merely* by consenting. Usually, there must be some additional reason to justify the killing, for example, that the person has a debilitating disease from which she can never recover, and her having this disease will mean that, for her, continued existence would be overall *worse* than immediate death. In Body-Replacement this is not true of the animal that is killed. It is not true that, if the animal lives, it experiences uncompensated suffering which, for it, makes its continued existence worse than its immediate death.

Second, it is not clear whether the animal can offer informed consent in this case, or even that it is capable of consenting at all. Suppose the animal (as well as the psychological continuer) is capable of referring to itself in the first person. If so, then given the details of the case, it may be mistaken about what will happen to it. If the animal thinks 'I will survive body-replacement', then it is mistaken, for body-replacement will in fact kill it. If deciding whether or not to undergo body-replacement is a life or death decision for the animal, and if, for it, the stakes are as high as they usually are when one faces a life and death decision, then we cannot plausibly claim that in this particular case the animal's *uninformed* consent is sufficient to justify killing it.

One might reject this rebuttal of the Consent Objection for the following reason: Even if the animal is mistaken about what will happen to it, if the animal were to correct its mistaken belief, this would have no moral significance, since whether or not the animal is right about what will happen to it cannot determine which decision it *ought* to make. Suppose the animal (and the psychological continuer) were to think ‘I, *the animal*, will not survive this procedure’. Then the animal would be not be mistaken about what will happen to it. Yet, the animal (and psychological continuer) could truly think, and should think ‘I, the psychological continuer, *will* survive the procedure’. Assuming the animal (and psychological continuer) are rational, this second thought should be sufficient to motivate it (and the psychological continuer) to undergo the procedure. For surely, the *right* decision is to undergo the procedure, whether or not the animal is mistaken about what will happen to it.

If this reply to my initial rebuttal of the Consent Objection is correct, it seems to provide a further reason to reject this objection. The crux of the reply is: in Body-Replacement, the right decision is to undergo body-replacement, and whether the animal mistakenly believes (*de se*) that it will survive the procedure does not affect the plausibility of this conclusion. But if this is true, then it is hard to see how the animal’s consenting to the procedure in the first place could by itself be significant enough to justify what would otherwise be a morally unjustified killing. Thus, it seems, if the killing of the animal in Body-Replacement is morally justified but the killing of a non-conscious fetus or embryo is unjustified, this cannot be fully explained by the fact that the animal consents in the first case but not the second.



Another worry about the Consent Objection is that it is implausible to suppose that the animal's consent has moral significance that is independent of the moral significance that my consent has. (Imagine the surgeon saying "Great! Now that we have both *your* consent *and* the consent of the human animal sitting in your chair, we can proceed with the surgery.") Intuitively, my consent is sufficient. We might say that I, the psychological continuer, am the fundamental or non-derivative consenter, while the animal associated with me consents only derivatively. As we saw, it is implausible to suppose that derivative harms and benefits have independent moral weight. The same idea goes for consent.

Some would say that the animal cannot even refer to itself in the first-person. For example, suppose that all of the animal's thoughts and utterances containing 'I' refer not to it but to me—the psychological continuer. This view has been proposed by Harold Noonan; he calls it *pronoun revisionism* (PR). If PR is correct, then when speaking in the first person the animal cannot consent to its having the body-replacement procedure. If the animal were to sincerely tell the hospital staff 'I consent to the procedure', it would be consenting to *my* having the procedure, since 'I' in the animal's mouth refers *only to me*. If this view of first-personal reference were true, it would further undermine the Consent Objection. For if the animal cannot even consent to the performance of an act X (where the performance or non-performance of X determines whether the animal dies) by sincerely uttering 'I consent to X', then, once again, it becomes very difficult to see how the animal's consent can be important enough to justify what would otherwise be an instance of murder.<sup>128</sup>

---

<sup>128</sup> Although Pronoun Revisionism is worth mentioning here in connection with the Consent Objection, I should also mention that for reasons I discussed in Chapter 1, a linguistic rule according to which utterances of 'I' refer only to psychological continuers wouldn't necessarily prevent animals from speaking

#### 4.4.3 The Afterlife Objection

In his illuminating discussion of the intrinsic value of human persons, Phillip Quinn considers the view that killing

violates human personhood because it destroys the biological life of the person's body, even if it does not destroy the person on account of the survival of bodily death by a soul which is reunited with a body at the resurrection in order to reconstitute the same person.<sup>129</sup>

There are some interesting similarities between Body-Replacement and the case that Quinn describes, in which a psychological continuer (who is an immaterial soul) is killed but continues to exist, awaiting the resurrection. In both of these hypothetical cases, the biological life of a human animal is destroyed while the psychological continuer associated with the animal survives. Furthermore, we may suppose that in each case the act that kills the human animal *benefits* the psychological continuer. For example, we may suppose that the case that Quinn describes involves the active euthanasia of a patient who would suffer greatly if he is not killed, and whose life would therefore be worse overall. According to what I will call *the afterlife objection*, body-replacement is immoral even when it benefits a psychological continuer and is necessary for (or the only reasonable means of) providing the desired benefit and does not have any morally problematic consequences other than its effect on the psychological continuer, the human animal, and those who perform the body-

---

and thinking about themselves in the first person. An animal that is associated with a psychological continuer is capable of first person *speaker* reference, even if it is not capable of first-person *semantic* reference. First person speaker reference would involve an inward-directed singular I-thought—an I-thought that points inward toward a subject of that thought.

<sup>129</sup> Phillip Quinn, "On the Intrinsic Value of Persons." In *Persons, Human and Divine*, edited by Peter van Inwagen and Dean Zimmerman (New York: Oxford University Press, 2007), 237 – 260. Here Quinn is reciting the view of Robert M. Adams in *Finite and Infinite Goods: A Framework for Ethics* (New York and Oxford: Oxford University Press, 1999), 111.

replacement. The performance of body-replacement on me is morally similar to the active voluntary euthanasia of a psychological continuer who goes to heaven when the human animal he is intimately associated with is killed. Such active voluntary euthanasia is immoral, the objection assumes, because it destroys something of great intrinsic value, namely, a biologically human life. This feature of active voluntary euthanasia is sufficient to make it immoral even when it is necessary to curtail the suffering of a psychological continuer. My hypothetical body-replacement has the same morally problematic feature—it destroys a biological human life of great intrinsic value. Therefore, it is, similarly, immoral, though necessary to benefit a psychological continuer.

Those who push the afterlife objection may, in addition, claim that destroying a human animal while it is intimately related to a psychological continuer wrongs the psychological continuer greatly. This is consistent with the plausible judgment that if you were to murder me *right now*, destroying the biological life of the human animal sitting in my chair, and causing my earlier-than-expected departure from the earthly realm, then not only would you destroy a biological life of great intrinsic value, you would also violate my right to life. Even if one is not a biologically living human animal, one can still possess a right to life. We may construe one's "right to life" as one's right to a certain continued arrangement between one and one's animal. For example, we may regard it as the right that one (a psychological continuer) has to continue standing in a certain relation (e.g., the constitutional relation) to a certain *biologically living* human animal. On this conception of our (i.e. psychological continuers') right to life, if you were to destroy the life of the human animal sitting in my chair, you would violate my right to life because, as a result of

your act, I would no longer stand to the animal in that relation (that is, in whatever relation I am entitled to stand in to it).

What should we think about these claims? Quinn plausibly denies that active euthanasia is always morally objectionable when it is done to avoid the (otherwise unavoidable) pain and suffering of the patient. However even if, implausibly, the active euthanasia of a psychological continuer is *always* morally objectionable, in Body-Replacement the active killing of the animal does not seem objectionable at all. Certainly, we ought to say, killing the animal in the conditions I described in my presentation of that case does not wrong me (the psychological continuer) or violate my rights, as one would if one were to kill the animal now as it sits in my chair (say by shooting it), thereby causing my earlier-than-expected departure from the earthly realm. Killing my animal to save my life, or to improve the quality of my life significantly, would not wrong me. Nor would it be immoral for the reason that it would destroy an intrinsically valuable biological human life.

I believe we should accept an even stronger claim. Suppose I will be severely physically incapacitated for some indefinite period of time unless I have a body-replacement and that surgeons who are fully capable of giving me a body-replacement are aware of this fact. These surgeons are employed at a certain medical institution—the only institution with the resources to perform body-replacements without putting the patient at significant risk of death or injury. Now suppose I go to the institution and request a body-replacement (my insurance company would cover the cost of the procedure). Can the surgeons rightfully refuse my request on the grounds that a body-replacement would destroy the biological life of the human animal associated with me? I am inclined to think

that they cannot. But even if they can rightfully refuse to treat me for this reason, such refusal would be grotesquely harsh. It would be like refusing a paraplegic's request that his biological but nonfunctional legs be amputated and replaced with fully functional prosthetic legs on the grounds that such amputation and replacement would severely mutilate a human animal, and would therefore unjustly damage this animal's intrinsically valuable biological life.

Moreover, whether or not such refusal would be grotesquely harsh, it seems clear that the surgeons would be justified in treating me. The afterlife objection to the second premise of the Symmetry Argument is unconvincing.

#### **4.4.4 Using Human Animals**

When I first introduced the Symmetry Argument, I mentioned that the conclusion of this argument applies only to the killing of human animals, and not to their *use* as a biological resource. In many cases in which a human animal is killed, it is also used for some further purpose. In this section, I argue that if it is morally justified to kill a human animal in the kinds of cases that I described, in which the act of killing is one for which there is a just cause, and this act is otherwise necessary (or reasonable) and proportionate, then it is also morally justified to use the human animal that is killed in these circumstances as a means of further good.

Using an individual as a means, in a way that also kills that individual, is usually thought to be morally worse than killing it without using it as a means. Kant's famous maxim that we should treat persons as ends in themselves, and never merely as means, is

undoubtedly the most common source of this idea, but it can be developed in many different ways.

However, for our purposes, we need not consider different accounts of why killing-and-using is morally worse than mere killing. It will suffice to consider a case that undoubtedly involves the killing and use of a human animal, and compare this with other cases in which the animal is merely killed, but not used. The following case is an extension of *Body-Replacement*:

### **Body-Replacement and Kidney Transplant**

I undergo body-replacement and avoid the nine months of unpleasantness and the non-trivial risk of death. The human animal associated with me is dismembered and killed in the procedure. Halfway through the procedure, while the animal is still alive and most of its body intact, the engineering team responsible for prepping my new synthetic body discovers that the filtration system intended to replace my original kidneys has malfunctioned. It is unanimously decided that I should be given a temporary pair of organic kidneys to filter toxins out of my blood until the new system has been fixed. The surgical team therefore removes the kidneys (my original kidneys) from the human animal and implants them in my (mostly synthetic) body, resulting in the death of the animal.

In this case, the human animal associated with me is not only killed, but also used as a means. If we believe that the killing of the human animal in *Body-Replacement* is morally justified, it would be very odd if we also believed that the killing of the human animal in *Body-Replacement and Kidney Transplant* is morally unjustified. In the former case, the animal is dismembered and killed; in the latter case, it is dismembered and killed, but some

of its parts—kidneys—are used for a further purpose. Removing the animal’s kidneys causes its death, but even if its kidneys had not been used, it would have died from the removal and scattering of its various parts—including its kidneys. It is strange to think that although *body-replacement* is morally justified, *body-replacement and kidney transplant* is morally unjustified on the grounds that in this case, the death of the human animal is the result of my original kidneys being reattached to me.

There are various complications we might consider. For example, we might consider how our moral evaluations would change if we imagined that the animal’s kidneys had been used to save someone else’s life rather than mine. (In that case, suppose that I wouldn’t have needed the kidneys.) However, when I think about different cases that include such added complications, I find it hard to believe that the moral significance of the distinction between killing-and-using and mere killing makes the kind of difference that would inhibit those who accept the Symmetry Argument from accepting Premise 2 of the Personal Ontology Argument. To me, this suggests that once we accept that killing a human animal is justified if it fulfills the weighty self-interest of a psychological continuer, we should also accept that killing *and using* the animal would be justified for the same reason.

#### **4.5 Weighing the Value of Self-Interest against the Value of Biological Life**

I have claimed that the fulfillment of a weighty self-interest of some psychological continuer constitutes a just cause for killing a human animal. But ‘weighty’ is vague. How weighty must the self-interest be? The answer depends, in part, on whether biological human life (the sort of life that defenders of the Biological Approach believe is necessary

for the persistence of human animals) has intrinsic value. If it doesn't, then the fulfillment of virtually any morally significant self-interest, such as the self-interest that I have in avoiding another mosquito bite, could justify killing a human animal.

However, I will assume that biological life is valuable and that it warrants respect. Perhaps there are convincing arguments against this view. If so, then we should reject it. But the view is intuitive, and reasonable as a default hypothesis. It also raises a problem for me. I claimed that in the hypothetical *Choice between Lives*, extending the life of the human animal associated with you does not constitute even a tie-breaking consideration in favor of giving a certain benefit to one psychological continuer rather than another. But how can we make sense of this claim if biological human life is intrinsically valuable and warrants respect? How is it that life's intrinsic value fails to tip the balance of reasons when all other considerations are equal?

The answer I suggest appeals to an idea that has been developed by Frances Kamm.<sup>130</sup> Kamm argues that it is sometimes inappropriate to make life and death decisions on the basis of utility. For example, if two patients are threatened with death, and doctors can save either patient, but not both, then if one patient is an engineer and the other a homeless person, this fact shouldn't influence the doctors' decision of whom to save. The fact that one patient is more likely to be beneficial to the rest of society is morally irrelevant in this context. Giving additional weight to this consideration would violate principles of distributive fairness. It is also irrelevant whether, by saving one patient rather than the other, the doctors would give very small benefits to any number of other people—where

---

<sup>130</sup> In her *Morality, Mortality: Death and Whom to Save from It, Vol. I* (New York: Oxford University Press, 1993), p. 147-151.



these benefits seem trivial compared to the benefit the doctors would give to either patient by saving her life. For example, suppose the doctors discover that one of the two patients is Willy Wonka, and that he will certainly give a Wonka bar to each living person on Earth if, but only if, he is the patient who survives. An enormous amount of extra utility would be gained from each of 7.3 billion people enjoying a Wonka bar, but this is irrelevant to the doctors' decision because the extra good for each person—the enjoyment of eating a small quantity of chocolate—is trivial compared to the good that either patient would gain by avoiding a premature death.

According to Kamm, this “principle of irrelevant utilities” is relevant to life and death decisions in at least two different types of case. In one type of case, the extra utilities are irrelevant because they are of the wrong kind. The extra good of having one more engineer in society is significant, but goods of this kind are irrelevant in the context of the doctors' decision. The good of saving a person's life and the social good of having one more engineer in society are goods belonging to different “spheres”, and the only goods that are relevant to the life and death decisions of medical doctors are goods belonging to the sphere of healthcare.

In another type of case, the extra goods are irrelevant because they are too small relative to what is at stake for those who are immediately affected by a life and death decision. Even if these extra goods belong to the right “sphere”, their size makes them irrelevant. The sphere of healthcare is concerned with avoiding both premature death and hangnails, but intuitively, we should prevent a single premature death rather than prevent *any* number of hangnails, assuming we cannot prevent both.

I am not sure whether the principle of irrelevant utilities has the particular application Kamm thinks it has—for example, whether it applies to the particular cases that we have just considered—but I want to suggest that this principle, or something like it, applies to cases in which a life and death decision must be made, and where the goods to be gained or lost by making one decision or another include both biological human life and the fulfillment of the self-interest most of us have in continuing to exist. For example, if we can save only one of two psychological continuers from death (here ‘death’ just means the ceasing to exist of something that is one of us) and we can decide whom to save, and all other relevant considerations apart from those concerning the preservation of biological human life are equal, then in this context biological human life is an irrelevant “utility”, in other words, an irrelevant good.

Biological human life is also an irrelevant good when our decision would affect only one psychological continuer. For example, if body-replacement would kill the human animal that is sitting here but it is the only reasonable way to extend my life (i.e. my personal history) by fifty years, then the fact that it would destroy the biological life of the animal is an irrelevant consideration.

These claims are consistent with an argument made famous by Michael Sandel. At a meeting of the President’s Council on Bioethics, Sandel argued that the view that human embryos have the same moral status as ordinary adult human persons (in virtue of their being biologically human) has implications that very few of us would accept. To emphasize this point, Sandel asked each person in his audience whether, if they were trying to escape a burning building, and could easily rescue a crate of ten frozen embryos or a five-year-old girl, but not both, they would save the embryos. Sandel believed that most

people would have the intuition that they ought to save the five-year-old girl, and that this suggests they don't really believe that human embryos have the same status as ordinary adult human persons.

I suggest that something similar to Kamm's principle of irrelevant utilities applies in this case. The stakes for the five-year-old (who is one of us) are high enough that biological life is an irrelevant consideration. Thus, not only should we save the five-year-old rather than ten human embryos, we should save the five-year-old rather than *any number* of human embryos. This assumes, of course, that all other morally relevant factors are equal. For example, it assumes that the embryos would remain frozen once rescued, that they wouldn't go on to engender new psychological continuers with happy lives, etc. Holding these other factors fixed is necessary in order to focus on the two kinds of values that we are now considering, biological human life and the self-interest most of us have in continuing to exist.

In Sandel's thought experiment, we are not asked to imagine that we must *kill* the human embryos in order to save the five-year-old. But now suppose that, just as we are about to abandon the embryos in order to save the five-year-old, the crate topples onto the floor and blocks the only available path between us and the five-year-old. The only way to reach the five-year-old is to step directly on top of the crate of embryos, thereby crushing all of them (this would involve killing rather than letting die). If, in order to rescue the five-year-old, we were prepared to abandon the embryos to the fire, would we really abandon the five-year-old to the fire in order to avoid stepping on the embryos? I find it extremely difficult to believe that we shouldn't step on the crate in order to reach the five-year-old. Furthermore, when I consider this case carefully, my intuitions are not

significantly influenced by the number of embryos that I imagine are in the crate. I think that we should save the five-year-old no matter how many embryos we must crush.

What about when the stakes for the psychological continuer are not as high? Consider the following case.

*Stephalopagus.* Steph is a psychological continuer and the only known adult survivor of cephalopagus—a rare instance of conjoined twinning in which two human animals share a single head and brain. In this case, Steph's single head is part of each animal and there is no physical feature that we can point to that would make us think that the head belongs more to one animal or the other. Steph is neurologically connected to each human animal and can control and coordinate the movements of their arms and legs. Aside from its limited mobility, each human animal is perfectly healthy. Moreover, Steph is happy, and over the years she has grown quite attached to both animals, which she regards as living parts of a single oddly-shaped body. Still she believes (correctly) that she would be happier if she had to drag around only one human animal instead of two. So she decides to undergo surgery in which one of the animals is separated from the other just below the neck, before being dissected, studied, and, ultimately, tossed into a container for biohazardous waste. The head remains with the surviving animal, which becomes the only human animal associated with Steph after the surgery.

In this case the self-interest in question is not as weighty as in life and death cases. However, it is hard to deny that killing the human animal in this case is morally justified. This gives us some idea of the strength of the self-interest that can outweigh the intrinsic value of biological life. There may be a threshold of self-interest above which no amount

of additional biological human life can justify failing to fulfill it, but below which tradeoffs between the value of human life and self-interest fulfillment can be made. But it may be difficult to determine where this threshold is.

#### **4.6 Do Human Animals Have Rights?**

Some believe that natural objects, including biological organisms, have rights.<sup>131</sup> Although The Symmetry Argument is logically compatible with this claim, I think we should reject it. Certainly, I would deny that human animals have *serious* rights, such as the right not to be destroyed. One *could* argue that although human animals have serious rights, these rights are almost always trumped by the competing rights of psychological continuers. But this assumes an implausible view of rights. If an individual has a right to something, then it should almost always be impermissible to deny her it. In my view, any so-called right that is very easily trumped or outweighed is too flimsy to plausibly be considered a right.

#### **4.7 Conclusion and Summary of Chapter 4**

In this chapter, I have argued that the Psychological Approach, when combined with plausible empirical claims about the timing of the onset of consciousness, greatly limits the number of actual cases in which the conclusions of most familiar pro-life arguments apply, and that it also supports the conclusion of the Personal Ontology Argument, which states

Killing a human animal and/or using it as a biological resource is morally justified if this is necessary (or the only reasonable way) to achieve a certain good, the

---

<sup>131</sup> Christopher Stone, "Should Trees Have Standing--Toward Legal Rights for Natural Objects" *Southern California Law Review* 45 (1972): 450–87.

fulfillment of a weighty self-interest of some psychological continuer, and apart from its impact on any human animal, it does not violate anyone's rights or produce any evil that nullifies or outweighs this good.

I began by examining the moral implications of the Psychological Approach for certain pro-life arguments against abortion and embryonic stem cell research. I argued that the Psychological Approach restricts the number of actual cases in which the conclusions of most familiar pro-life arguments apply. In some cases, the Psychological Approach restricts the scope of the argument without directly challenging any of the argument's major moral assumptions. In other cases, it restricts the scope of the argument without challenging *any* of the argument's premises.

I then gave a positive argument for the conclusion of the Personal Ontology Argument, which depends on the Psychological Approach, and I defended this argument against objections. Finally, I discussed how the self-interest of psychological continuers might be weighed against the value of the biological life of human animals. I argued that when the fulfillment of a weighty self-interest of a psychological continuer is at stake, biological human life is a morally irrelevant consideration. I appealed to something like Kamm's principle of irrelevant utilities, which implies that in some contexts, certain kinds of goods are relevant to what we ought to do, but others are not. Finally, I argued that killing a human organism can be morally justified even when the self-interest that is fulfilled is significantly less weighty than the self-interest most of us usually have in continuing to exist.

## 5. Do Reductionist Views Threaten the Personal Ontology

### Argument?

#### 5.1 Introduction to Chapter 5

In Chapters 1 through 4, I defended the two premises of the Personal Ontology Argument. In this chapter, I will consider a potential challenge to this argument, as well as a more general problem that has broader implications for normative theory, and particularly for aggregation. In what follows, I will be talking about the aggregation of *wellbeing*. However, readers should keep in mind my assumption about the fundamental moral importance of *individual self-interest*. Ultimately, those who care about self-interest will want to say that the overall goodness or badness of a state of affairs is largely (but perhaps not wholly) determined by aggregating the fulfillment of different individuals' self-interests. On this view, individual units of wellbeing, which I have been calling *benefits* (this is a somewhat non-standard use of the term 'benefit') matter morally because of their contribution to the fulfillment of individual self-interest. The fulfillment of individual self-interest, not wellbeing (or benefit), is what matters most fundamentally for the purpose of determining overall goodness. As we shall see in this chapter, this assumption faces serious challenges. In the previous chapter, we saw how wellbeing and self-interest fulfillment might be thought to come apart. In this chapter, we will consider another reason for thinking that they come apart. However, the arguments in this chapter put serious pressure on the view that self-interest is of fundamental moral importance. These arguments provide some support for the view that wellbeing is of fundamental importance.

Although the main aim of this chapter is to defend the Personal Ontology

Argument, the bulk of the chapter is devoted to the more general puzzle about aggregation, which is highly relevant to the main aim. In setting up the general puzzle, I begin with a discussion of a problem that arises for utilitarianism and other standard additive aggregationist views. For the purpose of this chapter, I will often use the terms ‘person’ and ‘psychological continuer’ interchangeably, and will often use the former in place of the latter.

## 5.2 The Question of Value-Bearers

**Total Utilitarianism** implies the following principle and criterion:

***Utilitarian Principle (U):*** Action  $a_1$  is morally better, and produces a better outcome, than action  $a_2$ , iff  $a_1$  produces more wellbeing than  $a_2$ .

***Sum Total Criterion (ST):*** Action  $a_1$  produces more wellbeing than action  $a_2$  iff the cardinal sum of wellbeing produced by  $a_1$  is greater than that produced by  $a_2$ .

ST faces the following problem. Suppose that time stretches infinitely into the future. Then under certain conditions, each available action produces an infinite, and therefore, undefined cardinal sum of wellbeing. Thus, ST implies that no such action produces more wellbeing than another. But this seems false. For example, if two actions both produce an infinite amount of wellbeing, but one of these actions produces more wellbeing *at every time*, then intuitively it produces *more* wellbeing.<sup>132</sup> To capture this

---

<sup>132</sup> See, e.g., Peter Vallentyne, “Utilitarianism and Infinite Utility,” *Australasian Journal of Philosophy* 71 (1993): 212-217.



intuition, several writers proposed the following criterion:

***Time-Centered Criterion (T):*** Action a1 produces more wellbeing than action a2 iff there is some time *t* such that the cardinal sum of wellbeing produced up to any time *later than t* by a1 is greater than the cardinal sum of wellbeing produced up to that same later time by a2.<sup>133</sup>

For actions that produce a finite amount of wellbeing, it seems, T always agrees with ST. Whenever the cardinal sum of wellbeing produced by action a1 is greater than that produced by action a2, there will be some time *t*, such that the cardinal sum of wellbeing produced up to any time later than *t* by a1 is greater than that produced up to that same later time by a2. However, unlike ST, T can discriminate between actions that produce infinite amounts of wellbeing. For example, it implies that if a1 and a2 both produce an infinite amount of wellbeing, but the amount of wellbeing produced by a1 *at each time* is greater than that produced by a2, then a1 produces more wellbeing than a2. For this reason, the conjunction of T and U, which I will call ***Time-Centered Utilitarianism (TCU)***, seems more plausible than the conjunction of ST and U.

However, as James Cain has shown, TCU has the following undesirable feature: it is insensitive to how much wellbeing an action produces *for each person*. To see this, suppose that the following outcomes, A and B, contain exactly the same persons and times.

**A.** Starting at the beginning of Year 1, and continuing forever, exactly one person is born each year, and each person lives exactly 3 years. Furthermore, the *n*th person

---

<sup>133</sup> Ibid., 216.

born has  $-2^n$  units (a *negative* amount) of wellbeing for the first year of her life, and  $(2^n \times 2)$  units (a *positive* amount) of wellbeing for the remaining 2 years of her life.

**B.** The same as A, except that the  $n$ th person born has  $2^n$  units (a *positive* amount) of wellbeing for the first year of her life, and  $(-2^n \times 2)$  units (a *negative* amount) of wellbeing for the remaining 2 years of her life.<sup>134</sup>

A and B are compared in Diagram 1 (page 202). The total wellbeing for each person is obtained by adding the numerical scores within the rectangles in each row of the diagram, whereas the total wellbeing in each year is obtained by adding the numerical scores in each column of the diagram.

---

<sup>134</sup> This case is adapted from James Cain, "Infinite Utility," *Australasian Journal of Philosophy* 73 (1995): 401-404. The design of Diagram 1 is adapted from Peter Vallentyne, "Infinite Utility,: Anonymity and Person-Centeredness," *Australasian Journal of Philosophy* 73 (1995): 417.

Diagram 1:

A							
--Time--							
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
--Persons--							
Person 1	-2	2	2	--	--	--	--
Person 2	--	-4	4	4	--	--	--
Person 3	--	--	-8	8	8	--	--
Person 4	--	--	--	-16	16	16	--
Person 5	--	--	--	--	-32	32	32
.....							
.....							
Total (each year):	-2	-2	-2	-4	-8	.....	.....

B							
--Time--							
	Year 1	Year 2	Year 3	Year 4	Year 5	Year 6	Year 7
--Persons--							
Person 1	2	-2	-2	--	--	--	--
Person 2	--	4	-4	-4	--	--	--
Person 3	--	--	8	-8	-8	--	--
Person 4	--	--	--	16	-16	-16	--
Person 5	--	--	--	--	32	-32	-32
.....							
.....							
Total (each year)	2	2	2	4	8		

In B, the total wellbeing at each time (i.e., in each year) is *positive*, whereas in A it is *negative*. Thus, the total wellbeing at each time is greater in B than in A. For this reason, TCU implies that an action resulting in B produces more wellbeing and is morally better than an action resulting in A. But this is implausible. Not only is the total wellbeing for each *person* greater in A than in B, but the total for each person in A is positive, whereas

in B it is negative. Thus, according to traditional utilitarian standards, in A, each person has a life that is worth living, whereas in B, each person has a life that is worth *not* living.

Peter Vallentyne has argued that the crucial insight provided by such examples is that utilitarians should aggregate wellbeing counting by *person* rather than by *time*. Thus, he proposes replacing TCU's *Time-Centered Criterion* with the following:

***Person-Centered Criterion (P):*** If the persons that exist or that will exist

are exactly the same if action a1 is performed as if action a2 is performed, then if there is a finite subset S of these persons such that for any finite set S' of persons that contains S, the total wellbeing produced for S' by a1 is greater than that produced for S' by a2, then a1 produces more wellbeing than a2.<sup>135</sup>

Call the conjunction of U and P ***Person-Centered Utilitarianism (PCU)***. PCU is similar to TCU, except that the relevant units for aggregating wellbeing are *sets of persons* rather than times. Because PCU aggregates wellbeing across *persons*, it can accommodate the intuition that one action is morally better than another if it produces more wellbeing for each person. For example, PCU implies that in Cain's example, an action resulting in A produces more wellbeing, and is morally better than an action resulting in B.

The conflict between TCU and PCU raises the question: What are the relevant value-bearing entities for the purpose of aggregating wellbeing, and for the purpose of aggregating value more generally? I call this *the question of value-bearers*. Value-bearers are the entities whose assigned "local" values, when combined in a certain way, jointly determine overall, or total, or "global" value of an outcome (as well as the contribution

---

<sup>135</sup> Ibid., 418. Vallentyne's formulation differs only slightly from mine.

that different acts make to this value). Cain's example shows that treating persons rather than times as the relevant value-bearers makes a difference to aggregation. In particular, his example shows that when persons are specified as the relevant value-bearers, the aggregation of goods seems to reflect our special concern *for persons*, as opposed to times, or experiences, or something else.

The *special concern for persons* that I have in mind can be understood by considering the phenomenon of sympathetic identification. When deciding how to act, we often consider whether the consequences of our actions would be desirable from the different forward-looking perspectives of those whom we might affect. We thereby gain a better understanding of what these persons have *prudential reason* to want—i.e., of what is in their *self-interest*. (Here, I am assuming that the morally relevant class of persons can include *merely possible* persons whose lives would be good or bad for them if they existed. I do not assume that our special concern for persons must be restricted to those who presently exist or to those who will certainly exist, or to any other subclass of persons.) Person-centered views of aggregation such as PCU seem plausible because the actions that they favor are those that seem best or most desirable for persons in this sense—these views seem to favor actions that maximize the fulfillment of individual self-interest (other things being equal).

I will argue that whether there is a plausible answer to the question of value-bearers depends on the nature of personal identity. According to *reductionist views*, facts about persons and their identity over time just consist in more particular facts about brains and

bodies, and the relations between certain physical and mental events.<sup>136</sup> I will argue that if a reductionist view is true, there is no intuitively plausible answer to the question of value-bearers; specifically, if a reductionist view is true, it is likely that any plausible combination of views about the individuation of persons and about individual self-interest will fix the value-bearers in such a way that when we aggregate their respective local values, the result that we get will be totally implausible. On one standard view of the individuation of persons (which I discuss below), any answer to the question of value-bearers will fail to reflect the special concern for persons that I have characterized, and hence, will fail to reflect the importance of individual self-interest. Alternatively, if we adopt a non-standard view of the individuation of persons, then although our answer to the question of value-bearers might reflect our special concern for persons and the importance of their self-interest, it also leads to the kind of bizarre double-counting of benefits that we saw in Chapter 4; the immediate contribution that a single benefit makes to overall goodness can be many times greater than it appears to be, since this benefit can be “shared” by many different people—in other words, it can contribute directly to the fulfillment of many different self-interests. I argue that an account of aggregation that involves this kind of double-counting cannot adequately account for the moral significance of *temporal wellbeing*—wellbeing the sum of which is derived counting *by time*.

One important upshot of the discussion in this chapter is that there are certain

---

<sup>136</sup> For discussions of reductionist views of persons and personal identity see Derek Parfit, *Reasons and Persons*, Part III, and David Shoemaker, “The Irrelevance/Incoherence of Non-Reductionism About Personal Identity,” *Philo* 5 (2002): 143-160.

views about the nature of persons (reductionist views provide one example) on which it is hard to treat persons—or any person-like entities—as the relevant value-bearers for aggregation without embracing a bizarre double-counting of benefits. This conclusion threatens to undermine Premise 2 of the Personal Ontology Argument. As we saw in Chapter 4’s discussion of the Symmetry Argument, my support for Premise 2 depends on my rejection of the view that human animals have morally significant self-interests that are independent of the self-interests of psychological continuers. I supported my rejection of this view by emphasizing the bizarre double-counting of benefits that it implies. But if reductionist views lead to the same kind of double-counting when these views are spelled out in such a way that they adequately reflect our special concern for persons, then reductionists may have reason to *accept* the double-counting of benefits. For example, if one is firmly committed to a reductionist view but also to our special concern for persons, then one might have good reason to accept double-counting (as I will show, it is hard to avoid this commitment). In that case, one might have no reason to accept the arguments that I gave in Chapter 4—arguments on which Premise 2 of the Personal Ontology Argument crucially depends. Another important upshot of the discussion in this chapter is that if a reductionist view is true, then it seems our account of aggregation must fail to capture either the moral importance of persons and their self-interests or the moral importance of temporal wellbeing; it is very hard to see how reductionists can capture both.

Here, I provide a more detailed outline of the chapter. In section 5.3, I consider different possible ways of individuating persons on the assumption that a reductionist view is true. For this purpose, I consider cases of personal fission and fusion similar to those discussed

by Derek Parfit.<sup>137</sup> In section 5.4, I argue that a person-centered view of aggregation, such as PCU, when combined with the most common view of the individuation of persons, fails to reflect our special concern for persons. In section 5.5, I argue that if a person-centered view is combined with a non-standard view of the individuation of persons, then although this person-centered view might reflect our special concern for persons, it will fail to account for the moral significance of temporal wellbeing. In section 5.6, I consider two possible solutions to the problem of how to account for the moral significance of both persons *and* temporal wellbeing. I argue that neither solution succeeds, and that the problem remains. In section 5.7, I consider whether accepting the double-counting of welfare that comes from combining reductionist views with our special concern for persons undermines the Personal Ontology Argument. I argue that while this consideration may weaken the Personal Ontology Argument, we should still accept this argument. Section 5.8 summarizes the chapter.

### 5.3 Fission and Fusion

Here is an example of fission:

***My Fission.*** My right and left cerebral hemispheres are redundant realizers of my psychological profile (memories, beliefs, desires, and other mental capacities). The corpus callosum which connects my two hemispheres is severed, and each is transplanted into a different body similar to mine. Afterwards, these bodies belong to what appear to be

---

<sup>137</sup> See Derek Parfit, "Personal Identity," *The Philosophical Review* 80 (1971): 3-27.



two separate persons, each of whom has one of my cerebral hemispheres, and a psychological profile qualitatively identical to mine.

In fusion, the process is reversed:

***Our Fusion.*** You and I have qualitatively identical psychological profiles, redundantly realized by the right and left hemispheres of our cerebrums. After each of us loses one cerebral hemisphere, our remaining hemispheres are transplanted into a single body, where their severed corpora callosa are stitched together. Afterwards, this body and the cerebrum composed of our cerebral hemispheres belong to what appears to be a single person who is psychologically indistinguishable from us.

What happens to persons who divide and fuse? The answer depends on the nature of personal identity. One possibility is that facts about persons and personal identity involve certain “further facts” over and above the familiar facts concerning brains, bodies, and interrelated physical and mental events. For example, persons might be indivisible souls. In that case, when one’s brain is divided, one goes wherever one’s soul goes. In this case, there are different possibilities regarding what happens to one, each of which is consistent with the more familiar facts about brains and bodies being what they are. Thus, it is consistent with one’s brain and body being split up a certain way that one follows one’s *right* cerebral hemisphere, or that one follows one’s *left* cerebral hemisphere, or that one is obliterated, or that one floats off somewhere unseen, etc., etc. However, if one accepts a reductionist view of persons and personal identity, then one must claim that the facts about persons and personal identity just consist in the more particular facts about brains and bodies. There are no further facts, e.g., concerning indivisible souls. Thus, the question

of what happens when one undergoes fission or fusion depends on other considerations regarding how we think persons should be individuated.

I will assume that a reductionist view is true. I will ignore non-reductionist views, in particular those that entail that persons are indivisible, primarily because the puzzles about aggregation that I want to discuss do not obviously arise on such views. It will be useful to adopt a view of how persons persist through time that is neutral with respect to the most popular ways of individuating persons among those who defend a reductionist view. For this purpose, I will assume that a person is composed of different *person-stages*, each of which is a *temporal part* of that person.  $X$  is a temporal part of  $Y =_{df.}$   $X$  is a part of  $Y$  and every part of  $Y$  that does not overlap  $X$  exists only at times at which  $X$  does not exist.<sup>138</sup> Any temporal interval through which a person persists, and in which she exemplifies certain person-like characteristics, is occupied by one of her person-stages. I will also assume that the criterion of personal identity over time is non-branching psychological connectedness and/or continuity with an appropriate cause. Following Parfit, I call this *Relation R*.<sup>139</sup> All of one's person-stages are *R-interrelated* (each is *R*-related to every other).

The standard view of the individuation of persons in cases of personal fission and fusion is what I will call

***No Survivors:*** A person who undergoes fission ceases to exist, and each of her fission offshoots is a new person who comes into existence at the moment of

---

<sup>138</sup> This working definition is adopted by Eric Olson. See Olson, *What Are We?* 100. For alternative definitions and a more extensive discussion of temporal parthood see Ted Sider, *Four-Dimensionalism* (New York: Oxford University Press), Chapter 3.

<sup>139</sup> Derek Parfit, *Reasons and Persons*, 215.

fission. Similarly, two or more persons who undergo fusion cease to exist, and their fusion offshoot is a new person who comes into existence at the moment of fusion.

Parfit defends a view similar to *No Survivors*.<sup>140</sup> On his view, after my cerebral hemispheres are divided and transplanted, I cannot plausibly be identified with one of my offshoots but not the other. Because I am related in the same way to both, such identification would be arbitrary. Nor can I plausibly be identified with both of my offshoots. This would involve a “great distortion in our concept of a person” since my offshoots are clearly separate (non-identical) people.<sup>141</sup> On Parfit’s view, the best description of *My Fission* is that I am not identical with either of my offshoots. Thus, when my cerebral hemispheres are divided and transplanted, I *cease to exist*. However, according to Parfit, I would have survived if only one of my cerebral hemispheres had been transplanted. Thus, for Parfit, whether or not I am identical to some future person depends not only on the intrinsic properties of and relations between me and this person, but on what *other* persons exist.<sup>142</sup>

Although I don’t survive my fission, Parfit thinks that I should regard it as I would ordinary survival. For example, I shouldn’t think that I would benefit by bribing a nurse

---

<sup>140</sup> Ibid., 256-258.

<sup>141</sup> Ibid., 256.

<sup>142</sup> This is a problematic feature of *No Survivors*, as well as all “closest continuer” theories of personal identity, such as that defended by Robert Nozick, *Philosophical Explanations* (Oxford: Clarendon Press, 1981). These theories violate what David Wiggins calls the ‘only a and b’ rule, according to which, whether a person, a, at some time, is identical to a person, b, at some other time, depends only on the intrinsic properties of and relations between a and b, and therefore cannot depend on the existence or non-existence of any other person. See David Wiggins, *Sameness and Substance Renewed* (Cambridge University Press, 2001), 96.

to destroy my left cerebral hemisphere, thereby preventing my fission.<sup>143</sup> Assuming I would survive if only one of my cerebral hemispheres was successfully transplanted, Parfit reasons, what matters in survival for me would not be lost merely because my other cerebral hemisphere was also successfully transplanted. A “double success” in this case cannot plausibly be considered a failure. According to Parfit, other things being equal, whether I regard some scenario as being as good (or about as good) as ordinary survival should not depend on the mere existence or nonexistence of another person. But given the implication of *No Survivors* that my continued existence *does* depend on the mere existence or non-existence of another person, it seems I can regard my fission as being as good (or about as good) as ordinary survival, even though I do not literally survive it. Thus, Parfit argues, my identity cannot be the basis of what matters to me in day-to-day survival, i.e., it cannot be what grounds my prudential concern about the future.

Parfit’s argument for this conclusion can be strengthened by appealing to examples of personal fusion. According to Parfit, fusion differs from fission in that some but not all of what matters in survival would be lost if the persons involved had different characteristics, desires, and intentions.<sup>144</sup> However, I am assuming that the persons who undergo fusion have qualitatively identical psychological profiles and that they are nearly perfect physical duplicates. Thus, we can plausibly assume that what matters in survival for these persons would be preserved, although they would cease to exist in the process.

One might argue that if what appeared to be two different persons were exactly psychologically similar prior to their fusion, we should say that, in fact, there is only one

---

<sup>143</sup> This way of putting the point is originally due to Robert Nozick in *Philosophical Explanations*.

<sup>144</sup> See Derek Parfit, “Personal Identity,” 17-19.

person, though he is wholly present in each of two separate locations, or, perhaps, that although there are two separate human bodies prior to the fusion, each of these bodies is part of a single divided person. I find these claims implausible, but their plausibility is irrelevant for our purposes. For we can assume that the psychological profiles of those who undergo fusion differ *enough* that we cannot count these persons as one. Even if their psychological profiles are somewhat different, we can plausibly assume that *at least some* of what matters in ordinary survival is preserved when they undergo fusion. In this case, what matters would hold *to some degree*.

In support of this claim, consider the fact that certain people have undergone hemispherectomy—a procedure in which one of the two cerebral hemispheres is surgically removed. This sometimes has the effect of improving the patient's quality of life but at the cost of altering her psychologically. It is hard to deny that if the psychological alteration were relatively minor, and the improvement in physical wellbeing significant enough, it would be in the person's self-interest to undergo the surgery. For similar reasons, it is hard to deny that it might be rational to undergo fusion, even if the psychological profiles of the persons involved are somewhat different, and hence, even if their fusion product has a psychological profile that is somewhat different from either of theirs.

Parfit's argument for the claim that identity is not what matters seems plausible, assuming that facts about personal identity over time consist in more particular facts about brains and bodies, and the relations between certain physical and mental events. However, those who accept a reductionist view can, following David Lewis, try to rescue the view that identity is what matters by rejecting *No Survivors* in favor of what I will call

***Cohabitation***: When a person undergoes fission, prior to the fission she is one of two or more distinct co-located persons. When fission occurs, these co-located persons come apart. They are just like different roads that overlap at a certain location and branch off in different directions at some other location. After the fission, each of the previously co-located persons is physically realized by a separate portion of matter, just as each of many overlapping roads is realized by a separate portion of tarmac where it diverges from the others. Similarly, after two or more persons undergo fusion, they are distinct but co-located persons physically realized in the same body. These persons are like roads that are separated at one location but merged at another location.<sup>145</sup>

Lewis and others who defend ***Cohabitation*** assume that a person who undergoes fission *survives* and that, before the fission, she is one of at least two distinct persons who share a *single* person-stage. Similarly, they assume, persons who undergo fusion survive by sharing a single post-fusion person-stage. According to Lewis and most other defenders of ***Cohabitation***, when two or more persons share a single person-stage, each of them is the subject of that person-stage's experiences, thoughts, and other mental events. Thus, in cases of fission and fusion, a single stream of consciousness is shared by multiple persons.

***Cohabitation*** and ***No Survivors*** are the best known and, I believe, most plausible views about how persons are individuated in cases of personal fission and

---

<sup>145</sup> David Lewis, "Survival and Identity," in *The Identities of Persons*, ed. Amelie O. Rorty (Berkeley: University of California Press, 1976).

fusion. There are other less plausible views but I will not discuss them here. Nor will I discuss the relative merits of *No Survivors* and *Cohabitation*. Instead, I will focus on what these views imply regarding the aggregation of goods across persons.

#### **5.4 A Conflict between Person-Centered Views of Aggregation, No Survivors, and Our Special Concern for Persons**

Consider the following case:

##### **Fusion Rehabilitation**

Derek and David have identical psychological profiles which, as in most philosophical examples of fission and fusion, are redundantly realized by each of their cerebral hemispheres. They also have nearly qualitatively identical histories, so that each is like a nearly perfect replica of the other. At time  $t_0$ , they each lose one cerebral hemisphere as the result of a disease that permanently infects their bodies, and diminishes their quality of life. Whereas each enjoys 2 units of wellbeing per year at full health prior to  $t_0$ , while suffering from the disease, each enjoys only 1.2 units of wellbeing per year.

Five years later (i.e., five years after  $t_0$ ) at time  $t_5$ , a new medical procedure becomes available involving fusion. If Derek and David undergo this procedure, their remaining cerebral hemispheres will be transplanted into a disease-free body where their corpora collosa will be joined. The person who emerges from the fusion will be fully psychologically continuous with Derek as

well as with David, and he will be at full health. However, Derek and David must wait for a suitable body to become available for transplant.

Now, consider two possible outcomes:

***Early Fusion.*** A body is procured immediately. Derek and David undergo fusion at time  $t_5$  (five years after  $t_0$ ). The person who emerges from their fusion lives for ten years, and enjoys 20 units of wellbeing (2 units per year) before dying at time  $t_{15}$ .

***Late Fusion.*** Five years pass before a body is procured. During this time, Derek and David continue to enjoy only 1.2 units of wellbeing per year. (Thus, during this time, each enjoys a total of only 6 units of wellbeing. Once a body is finally procured, they undergo fusion at time  $t_{10}$  (ten years after  $t_0$ ). The person who emerges from their fusion lives five years, and enjoys 10 units of wellbeing (2 units per year) before dying at  $t_{15}$ .

Assuming *No Survivors*, in ***Early Fusion*** and ***Late Fusion*** there are three persons involved: Derek, David, and their fusion offshoot, who is a separate person. For simplicity, let us suppose that in both ***Early Fusion*** and ***Late Fusion*** the very same person emerges from the fusion of Derek and David. Call this person *Dervid*. Also let us suppose that no one other than Derek, David, and Dervid is affected for better or worse in either outcome.

***Early Fusion*** and ***Late Fusion*** are compared in Diagram 2 (page 216). As in



Diagram 1, the total wellbeing for each person is obtained by counting the numerical values distributed horizontally within each bar, and the total wellbeing at each time is obtained by counting the numerical values distributed vertically in the columns representing the three temporal intervals (i.e., between  $t_0$  and  $t_5$ , between  $t_5$  and  $t_{10}$ , and between  $t_{10}$  and  $t_{15}$ ).

Diagram 2:

		--Time--			
		t0	t5	t10	t15
Early Fusion	<u>Person</u>				<u>Total for each person</u>
	Derek	6			6
	Dervid		10	10	20
	David	6			6
	Total at each time	12	10	10	
Late Fusion	<u>Person</u>				<u>Total for each person</u>
	Derek	6	6		12
	Dervid			10	10

David	6	6	12
Total at each time	12	12	10

---

Person-centered views of aggregation favor actions resulting in *Late Fusion* to those resulting in *Early Fusion*. For example, recall *Person-Centered Utilitarianism* (PCU), which discriminates between different infinite-wellbeing-producing actions while treating persons rather than times as the relevant value-bearing entities. According to PCU, action  $a_1$  produces more wellbeing than action  $a_2$  iff there is a finite set  $S$  of persons such that for any finite set  $S'$  that contains  $S$ , the total wellbeing produced for  $S'$  by  $a_1$  is greater than that produced for  $S'$  by  $a_2$ . Now with respect to those actions that produce only finite wellbeing, PCU seems to agree with standard total utilitarianism; whenever the cardinal sum of wellbeing produced by  $a_1$  is greater than that produced by  $a_2$ , it seems there will be some finite set of persons  $S$  such that  $a_1$  produces more wellbeing than  $a_2$  for  $S$  and for every larger finite set of persons containing  $S$ . In **Fusion Rehabilitation**, the relevant set  $S$  is the set whose members are Derek, David, and Dervid. In *Late Fusion* this set has 34 units of wellbeing, whereas in *Early Fusion* it has only 32 units of wellbeing. And since, we are assuming, no one other than Derek, David, or Dervid is affected by an action resulting in either *Early Fusion* or *Late Fusion*, PCU implies that an action resulting in *Late Fusion* produces more wellbeing and is morally better than an action resulting in *Early Fusion*.

However, for reasons we considered in the previous section, *Early Fusion* seems

more desirable than *Late Fusion for persons*. In other words, it seems better with respect to the fulfillment of persons' self-interest. It seems that, at  $t_0$ , from Derek's forward-looking perspective, fusion is as good (or nearly as good) as ordinary survival. Derek should not think that he would benefit by bribing a nurse to destroy David's one remaining cerebral hemisphere, thereby preventing their fusion. The same remarks also seem to apply to David who, at  $t_0$ , is in the same situation as Derek. But if fusion is as good (or nearly as good) as ordinary survival for both Derek and David, then, at  $t_0$ , from each of their forward-looking perspectives, it is better to fuse *earlier* rather than later. Practically speaking, for each of them, fusing earlier is not like dying earlier, but is instead like having the quality of his life improve earlier. This seems plausible, especially given that Derek and David are nearly perfect replicas of each other prior to their fusion. But even if we assume that there are incompatibilities in Derek's and David's psychological profiles and that, therefore, for each of them, fusion is substantially less desirable than ordinary survival, the difference in wellbeing that is brought about by their fusion could compensate for this; thus, it could still be true that, at  $t_0$ , for each of them, *Early Fusion* is more desirable than *Late Fusion*. This might be like a case of hemispherectomy in which the patient's life is dramatically improved, but her psychological profile significantly altered.

We can plausibly claim that *Early Fusion* is more desirable than *Late Fusion* for both Derek and David or, at least, that the details of the case could be filled out in a way that would make this true. Now consider Dervid, who is the only other person in **Fusion Rehabilitation** for whom it matters whether *Early Fusion* or *Late Fusion* obtains. In *Early Fusion* Dervid has 20 units of wellbeing, whereas in *Late Fusion* he has only 10,

and so, for Dervid, *Early Fusion* is clearly more desirable than *Late Fusion*. So it seems *Early Fusion* is more desirable than *Late Fusion* for each of the three persons involved. It therefore seems that, overall, *Early Fusion* is more desirable than *Late Fusion* for persons. PCU seems to ignore this important fact. It evaluates different actions only by how much wellbeing they produce for persons. For this reason, it implies that one action is morally better than another even when the consequences of the former are less desirable from the perspective of each person who is affected. From the perspective of any given person, an outcome in which that person ceases to exist sooner but is R-related to some future person might be just as good as an outcome in which he ceases to exist later, even if his ceasing to exist sooner implies that his life thereby contains *less* wellbeing. For this reason, our special concern for persons will sometimes favor actions that produce less wellbeing for persons (as well as less wellbeing when counting by time). But according to PCU, the contribution that any person's life makes to the total wellbeing of an outcome is determined solely by how much wellbeing that life contains. For this reason, PCU *cannot* favor one action over another if the former produces *less* wellbeing for persons. Thus, the conjunction of *No Survivors* and PCU does not adequately reflect our special concern for persons as I have characterized it. It does not favor the performance of actions that result in what each person, or what persons in general, have most prudential reason to want.

### 5.5 Cohabitation and the Moral Significance of Temporal Wellbeing

We can avoid this result by individuating persons (and thus, value-bearers) in such a way that whenever our special concern for persons favors one action over another, person-

centered views such as PCU imply that the former produces more wellbeing and is morally better than the latter. For this purpose, we can appeal to *Cohabitation*, according to which fission and fusion involve multiple persons who are co-located. According to *Cohabitation*, no new person is created or destroyed when fission or fusion occurs. Thus, in **Fusion Rehabilitation**, when Derek and David undergo fusion, each continues to exist afterwards. They are the only persons affected for better or worse in this case.

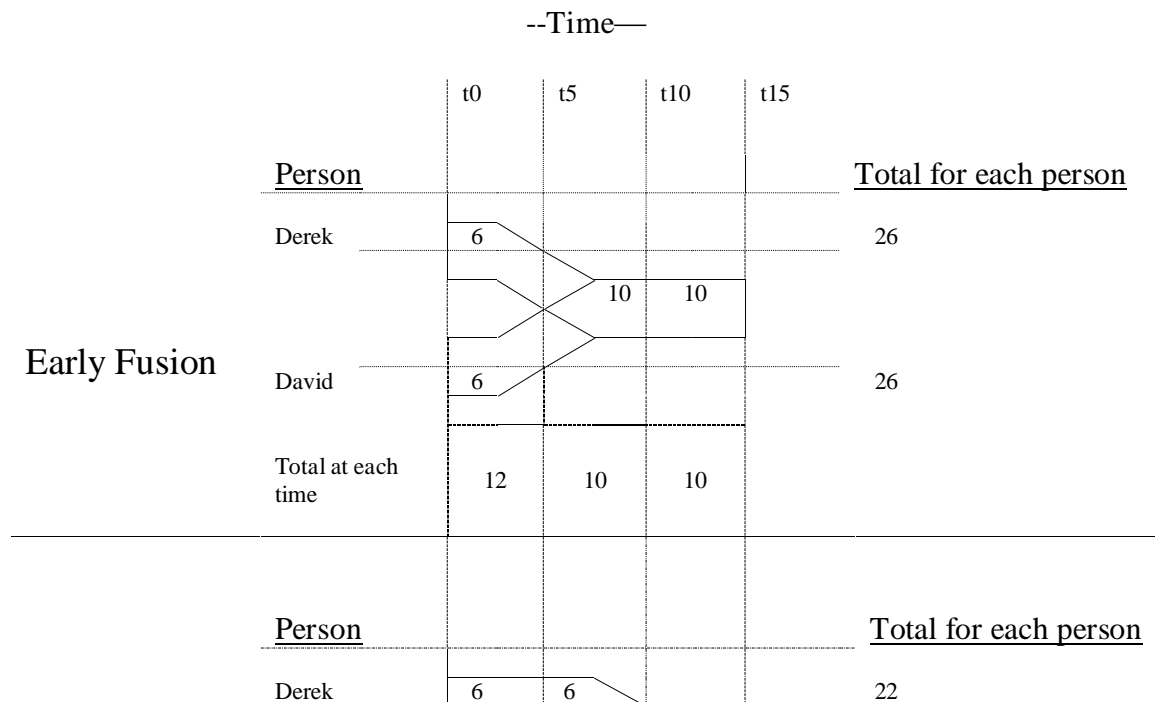
Assuming *Cohabitation*, it seems that, relative to  $t_0$ , for both Derek and David, it is better to fuse earlier rather than later. Therefore, it seems, the practical implications for Derek and David are the same assuming *Cohabitation* as they are assuming *No Survivors*. The difference is how this is reflected in the aggregation of wellbeing. If we assume *Cohabitation* then, counting by person, *Early Fusion* contains 52 units of wellbeing (26 for Derek and 26 for David) whereas *Late Fusion* contains only 44 units (22 for Derek and 22 for David). Therefore, the conjunction of *Cohabitation* and PCU implies that an action resulting in *Early Fusion* produces more wellbeing and is morally better than an action resulting in *Late Fusion*. So the conjunction of *Cohabitation* and PCU seems to reflect our special concern for persons, at least in this particular case.

On the other hand, *Time-Centered Utilitarianism* (TCU) does not seem to reflect this special concern. Counting by time (rather than by person) *Early Fusion* contains only 32 units of wellbeing, whereas *Late Fusion* contains 34 units. Therefore, TCU implies that an action resulting in *Late Fusion* produces more wellbeing and is morally better than an action resulting in *Early Fusion*. One surprising and important implication of *Cohabitation* is that PCU and TCU disagree not only in certain cases involving infinite wellbeing, but also in certain cases involving *finite* wellbeing. Because we are assuming

*Cohabitation*, PCU and TCU disagree in their respective evaluations of the two alternative outcomes, just as they disagree in their respective evaluations of the different outcomes in Cain's infinite case. In this case, the conflict is possible not because the two outcomes contain infinite amounts of wellbeing, but because the persons in these outcomes share some of their person-stages. The question of which view is more plausible can therefore be raised without appealing to cases involving infinite wellbeing.

*Early Fusion* and *Late Fusion* are compared in Diagram 3, assuming *Cohabitation*. The physical processes represented in Diagram 3 are exactly the same as those represented in Diagram 2. But the numbers are different because we are now presupposing a different view of the individuation of persons. (Instead of three non-overlapping rectangles there are two overlapping rectangles, each representing one of the two persons in the fusion.)

Diagram 3:



### Late Fusion

				10
David	6	6		22
Total at each time	12	12	10	

One might think that, as in Cain's example of infinite wellbeing, in *Fusion Rehabilitation*, PCU better reflects our special concern for persons and is more plausible than TCU. When one puts one's self in the shoes of each of the two persons in this example, and imagines what his life would be like in each of the two possible outcomes, it is in each person's self-interest to prefer *Early Fusion* to *Late Fusion*. This conclusion is more obvious, now that we are assuming a view of the individuation of persons according to which each person's life *literally* contains more wellbeing in one of the two possible outcomes. Thus, one might think that *Fusion Rehabilitation* reinforces Vallentyne's claim that utilitarians should adopt a person-centered view of the aggregation of wellbeing.

However, the conjunction of *Cohabitation* and PCU is highly problematic for the following reason. The conclusion that an action resulting in *Early Fusion* produces more wellbeing than an action resulting in *Late Fusion* is reached by *double-counting* the amount of temporal wellbeing at those times in *Early Fusion* and *Late Fusion* at which Derek and David share a person-stage. For example, in *Early Fusion* the person-stage beginning at t5 and ending at t10, and the person-stage beginning at t10 and ending at t15, each have only 10 units of wellbeing; but because each is a temporal part of *two persons*, each contributes *20 units* of wellbeing to the total amount of wellbeing that *Early*

*Fusion* contains—10 units *for Derek* and 10 units *for David*. Such “double-counting” is the result of combining a person-centered view of aggregation such as PCU with *Cohabitation*. Thus, aggregating across *persons* rather than *times*, person-centered views of aggregation count the wellbeing in any particular temporal region  $n$  times in order to derive the total wellbeing of the outcome, where  $n$  is the number of persons overlapping at that particular temporal region.

To see why this is problematic, consider the following case.

### **Spaceflight versus Tele-fusion**

Recently, 1000 people with very similar psychological profiles have all contracted a disease which causes pain so intense that for each of them, every moment of waking life is worse than temporary unconsciousness. Thus, a year in the life of any one of these 1000 diseased persons contains a negative amount of wellbeing. Fortunately, there is an effective treatment for the disease. However, because the disease is so rare, this treatment is administered only once every two years at a remote medical facility on Mars. And, as luck would have it, the window for receiving treatment has just ended! Each of the 1000 must therefore endure two more years of suffering before being treated.

There is a question as to how the 1000 will reach Mars in time to receive the next available treatment. Two methods of interplanetary travel are available. The first method, which is cheaper but much slower, involves taking a space shuttle. If the 1000 take the space shuttle to Mars today, they will reach the medical facility in exactly two years, arriving just in time to receive the desired treatment.



The second method of travel, which is more expensive but much faster, is teletransportation whereby one enters a booth in which one is disassembled at the molecular level and then “beamed” to one’s destination, where one is reassembled from different molecules. (Because the person-stages prior to disassembly are R-related to the person-stages after reassembly, teletransportation preserves personal identity.) There are different ways teletransportation can occur. In most cases, a single person is disassembled and then reassembled shortly afterward. However, it is also possible for persons to undergo tele-fission and tele-fusion. In tele-fusion, two or more persons are simultaneously disassembled and “beamed” to a single booth where they are reassembled. However, the reassembly is unusual; rather than two separate bodies, only one body is involved in the reassembly process. In accordance with *Cohabitation*, this body corresponds to a single person-stage that is a temporal part of each person who undergoes tele-fusion.

Now consider two possible outcomes:

***Spaceflight.*** The 1000 diseased people all take the space shuttle to Mars. They suffer for two agonizingly long years before finally arriving and receiving treatment.

***Tele-fusion.*** The 1000 undergo tele-fusion. What appears to be a single person steps out of the teletransportation booth on Mars, and continues to suffer for two agonizingly long years before the medical team finally arrives to administer the treatment.

According to *Cohabitation*, *Spaceflight* and *Tele-fusion* contain exactly the same persons. Those who take the space shuttle to Mars in *Spaceflight* are teletransported to Mars in *Tele-fusion*. Thus, in *Tele-fusion* no new person is created or destroyed; what appears to be a single person suffering on Mars is actually a temporal part of 1000 different persons. For this reason, *Cohabitation* and PCU jointly imply that the total wellbeing for each person is the same in *Spaceflight* as in *Tele-fusion*. From each person's perspective, it makes no difference (with respect to her wellbeing) which outcome obtains. However, it seems obvious that *Spaceflight* contains much more suffering along the temporal dimension, and therefore, has a much greater negative sum of temporal wellbeing than *Tele-fusion*. In *Spaceflight*, because each of the 1000 persons is realized in a separate body, there is a large temporal region that is filled with 1000 people who are suffering, whereas in *Tele-fusion* this very same temporal region is completely devoid of suffering. PCU ignores this difference; fundamentally, it is insensitive to how much temporal wellbeing an outcome contains. But intuitively, temporal wellbeing matters; it can contribute to the goodness or badness of different outcomes, and can make a difference regarding the moral evaluation of different actions. But *Cohabitation* and PCU together seem to imply that temporal wellbeing doesn't matter in this way. I call this *the problem of insensitivity* to temporal wellbeing. Views that are insensitive to temporal wellbeing imply, for example, that it doesn't matter morally how much of the world is filled with people who are suffering, or, for that matter, how much of the world is filled with people who are well off.

The conjunction of *Cohabitation* and PCU is problematic for another reason. Consider the following variation of *Tele-fusion*.

**Tele-fusion 2.** The 1000 remain on Earth and suffer for two agonizingly long years before undergoing tele-fusion. What appears to be a single person steps out of the teletransportation booth on Mars, and continues to suffer for several minutes before the medical team arrives and administers the treatment.

Consider the several minutes of suffering between the time the 1000 people undergo tele-fusion and the time the medical team arrives on Mars and administers the treatment. This suffering is of course bad, and so its corresponding cardinal sum of wellbeing is negative. But **Cohabitation** and PCU jointly imply that the contribution that this short period of suffering makes an *enormous* difference to the badness of the outcome. Because the person-stage in which this suffering occurs is a temporal part of 1000 different persons, PCU multiplies the corresponding negative cardinal sum of wellbeing by 1000. Thus, what is intuitively a very small negative contribution to the value of an outcome is 1000 times greater than it seems. The problem can be magnified endlessly. If 1,000,000 persons tele-fuse, PCU counts the suffering of their shared person-stage 1,000,000 times to determine overall value. If 1,000,000,000 persons tele-fuse, PCU counts this suffering 1,000,000,000 times, etc. Assuming both **Cohabitation** and PCU, the contribution that several minutes of suffering make to overall value might be equivalent to that of several years, or several *million* years, of (equally intense) suffering incurred by a single person, or by many different non-overlapping persons. But intuitively, several minutes of suffering couldn't make such a huge difference in value. The conjunction of **Cohabitation** and PCU not only ignores the apparent moral significance of huge

differences in temporal wellbeing, it magnifies the apparent moral significance of very small differences in temporal wellbeing. I call this *the problem of hypersensitivity to differences in temporal wellbeing*.

The conjunction of **Cohabitation** and PCU faces yet another problem, which I call *the problem of future-dependence and past-dependence*. Suppose that two years from now you will undergo fission, and that your fission will produce 1000 separate (i.e., non-overlapping) person- stages. Then **Cohabitation** implies that each of the person-stages prior to your fission is a temporal part of 1000 separate persons, and so PCU and **Cohabitation** jointly imply that the contribution to total wellbeing that is made by each of your experiences prior to your fission is 1000 times greater than it seems. For example, if you stub your toe sometime between now and your fission, then the negative contribution of this toe-stubbing to total wellbeing is equivalent to the contribution to total wellbeing that would be made (under normal circumstances) by 1000 separate toe-stubbings. This is very hard to believe. The contribution to total wellbeing of what happens at a given time should not be compounded by an event of fission occurring at some later time (future-dependence). Similarly, it should not be compounded by an event of *fusion* occurring at some *earlier* time (past- dependence).<sup>146</sup>

One can see a dilemma emerging. Initially, we encountered the following problem: With **No Survivors** as our view of the individuation of persons and PCU as our principle of aggregation, it seemed we couldn't evaluate actions according to their contribution to the fulfillment of persons' self-interest; thus, our view of aggregation failed to reflect our

---

<sup>146</sup> Note that the conjunction of **Cohabitation** and PCU implies that the contribution that any particular part of the world makes to total utility depends on its relation to other parts of the world. Thus, the relevant value-bearing entities are not *separable* in the sense that is required by standard versions of total utilitarianism.

special concern for persons. Initially, it seemed we could avoid this problem by replacing *No Survivors* with *Cohabitation*. *Cohabitation* and PCU together seem more in line with the aim of maximizing the fulfillment of persons' self-interest, and so perhaps this combination of views better reflects our special concern for persons. However, it reflects this concern by *double-counting* along the temporal dimension. And, as we saw, such double-counting renders our view of aggregation insensitive, as well as hypersensitive, to differences in temporal wellbeing. It also leads to the problem of future-dependence and past-dependence; the contribution that a certain amount of temporal wellbeing makes to the total value of an outcome depends, implausibly, on future and past events of fission and fusion. Thus, on a reductionist view, it is difficult to see how we can individuate persons to make our theory of aggregation reflect both the moral significance of persons (and self-interest) and the moral significance of temporal wellbeing.

This problem cannot be avoided by treating person-like entities other than *whole persons* as the relevant value-bearers. Some writers argue that the objects of moral concern, as well as the referents of our personal pronouns, are entities other than whole persons, e.g., strongly psychologically connected person-stages, or momentary subjects of experience within the life of a whole person.<sup>147</sup> But if *persons* can divide and fuse, then presumably so can other person-like entities that have temporal extension. Therefore, we can ask whether, in cases of fission and fusion, there are *three* such person-like entities, or *two* such entities that overlap, etc. And, it seems, however we answer this question,

---

<sup>147</sup> See, e.g., Derek Parfit, "Later Selves and Moral Principles," in *Philosophy and Personal Relations*, ed. A. Montefiore (London: Routledge and Kegan Paul, 1973), 139-142, and Galen Strawson, *Selves: An Essay in Revisionary Metaphysics* (New York: Oxford University Press, 2009).

we will end up with the same problem: Either our account of aggregation will fail to reflect our special concern for the relevant person-like entities, or it will encounter the problems of insensitivity, hypersensitivity, and future-dependence/past-dependence. Thus, it seems, the most general statement of our problem is this: If a reductionist view of personal identity is true, then no matter what entities are specified as the relevant value-bearers, and no matter how these entities are individuated, our account of aggregation will either fail to adequately reflect the apparent moral significance of persons *or person-like entities*, or fail to adequately reflect the moral significance of temporal wellbeing. Let us call this *The Reductionist Dilemma*.

## **5.6. Two Possible Solutions to the Reductionist Dilemma**

In this section, I consider two possible responses to The Reductionist Dilemma. In order to be successful, these responses should resolve the problems of insensitivity, hypersensitivity, and future- and past-dependence.

### **5.6.1 Does the Reductionist Dilemma Rest on a Mistake in Quantifying Experience?**

One possibility is that the problems of insensitivity, hypersensitivity, and future- and past-dependence seem to arise due to a mistaken assumption about the amount of experience, or number of individual experiences, in cases of personal fission and fusion. So far I have assumed that if *Cohabitation* is true then multiple persons can be the subjects of the same token experiences. But one might think that wherever there are 1000 overlapping experiencing persons there are in fact 1000 distinct streams of experience—one for each

person. Call this *Multi-Stream Cohabitation*.<sup>148</sup> If we accept this view, then we should not say that the contribution to total wellbeing that is made by each momentary experience occurring in a given person-stage depends on the number of persons of which this person-stage is a temporal part; instead, we should say that wherever there appears to be a single experience occurring in this person-stage, the actual number of experiences occurring depends on the number of persons of which that person-stage is a temporal part. However, *Multi-Stream Cohabitation* seems just as implausible (if not *more* implausible) than the problem of future-dependence and past-dependence. It has the absurd implication that the amount of experience (or number of experiences) at a given time can depend (non-causally) on events of personal fission or fusion occurring at earlier or later times. Such a bizarre proliferation of experience seems utterly inexplicable.

It is worth mentioning here that the absurdity of *Multi-Stream Cohabitation* does not merely derive from *Cohabitation*'s implication that the number of persons now sitting in your chair depends on whether you undergo fission in the future. One might think that if the number of *persons* at a given time depends on whether fission or fusion occurs at other times, then there is no additional absurdity in the claim that the number of *experiences* at that time depends on whether fission or fusion occurs at other times. However, this is mistaken. To see why, recall the temporal parts theorist's analogy: A person is extended in time as a road (at a time) is extended in space. Assuming that two or more roads can overlap, it is not absurd to suppose that, for example, whether a hitchhiker is standing on one or two (or three or four) roads at some particular spatial location

---

<sup>148</sup> For a discussion of this view see Eugene Mills, "Dividing without Reducing," *Mind* 102 (1993): 37-51.

depends on whether the road forks at some *other* spatial location. However, it *is* absurd to suppose that the amount of tarmac or the number of potholes in some particular segment of road depends (non-causally) on whether the road forks at some other spatial location. We shouldn't think that where there appears to be a single pothole lying at the intersection of two roads there are actually two spatially coincident potholes—each belonging uniquely to one of the two intersecting roads. Similarly, *if* persons are extended in time as roads are in space, then it is not absurd (or at least seems much less absurd) that the number of persons sitting in your chair depends on whether there is a fission in your future; but it is still *very absurd* that the number of experiences now occurring in your brain depends on whether there is a fission in your future. The absurdity of this claim is *independent* of any absurdity inherent in *Cohabitation*.

### **5.6.2 What if Both the Personal Dimension and the Temporal Dimension Have Moral Significance?**

Another possible response to The Reductionist Dilemma appeals to a *pluralist* view of what the relevant value-bearers are. The view that I have in mind is one according to which *both* persons (or person-like entities) *and* times are value-bearers. Such a view can mitigate the problem of insensitivity to temporal quantities of wellbeing. For example, a view that aggregates wellbeing counting by time *as well as* by person might imply that if two possible outcomes, A and B, contain the same persons, and each person has the same wellbeing in A as in B, but A contains a greater temporal quantity of wellbeing than B, then A's total wellbeing is greater than B's. However, because such a view treats persons as value-bearers in addition to times, it is hard to see how it can avoid the problem of



hypersensitivity. If a very small temporal quantity of wellbeing can contribute to the total wellbeing of each of many different persons, then it seems the pluralist view will be hypersensitive to such quantities. For the same reason, it is hard to see how this view can avoid the problem of future-dependence and past-dependence.

I suspect that if we were to treat both persons and times as value-bearers, our account of aggregation would also fail to adequately reflect our special concern for persons. To see why, recall Cain's counterexample to TCU, in which two outcomes, A and B, contain the same persons and times, and *each person* has more wellbeing in A, even though the total wellbeing at *each time* is greater in B. In this case, our special concern for persons clearly favors A, and A seems to be *the better outcome* despite the fact that there is less wellbeing at each time in A than in B. But because an action resulting in A and an action resulting in B would both produce an infinite cardinal sum of wellbeing, if we aggregate wellbeing counting both *by person* and *by time*, it is difficult to see how we will get the desired result that an action resulting in A produces more wellbeing than an action resulting in B. (For the same reason, it is difficult to see how we could get the *implausible* result that an action resulting in B produces more wellbeing than an action resulting in A.) Thus, pluralism about value-bearers may simply embrace *both* horns of the dilemma.

## **5.7 Should Reductionists Accept the Personal Ontology Argument?**

### **5.7.1 Does Double-Counting In Cases of Fission and Fusion Undermine the Personal Ontology Argument?**

If we are reductionists, and we treat persons as the relevant value-bearers for aggregation, and if we want our account of aggregation to reflect our special concern for persons, then it is hard to see how we can avoid accepting some kind of double-counting of benefits. If we accept double-counting, then our view of aggregation fails to reflect the apparent moral significance of temporal wellbeing. But the rejection of double-counting provides the basis for Chapter 4's defense of the Personal Ontology Argument. My defense of Premise 2 of that argument relies on the claim that for the purpose of moral evaluation, particular benefits and harms should not be counted twice on the grounds that these benefits and harms are incurred non-derivatively by psychological continuers and derivatively by human animals. Is there a principled reason to embrace double-counting in cases of fission and fusion but reject double-counting in cases of intimately related psychological continuers and human animals? Why double-count in one case but not in the other?

### **5.7.2 Double-Counting and the Distinction between Derivatively Conscious and Non-Derivatively Conscious Entities**

Perhaps one difference is that in cases of fission and fusion, no one is conscious *only derivatively*. Any person involved in fission or fusion is a non-derivative subject of consciousness. But some who accept a temporal parts ontology want to say that a person (a psychological continuer) is conscious only in virtue of having temporal parts—person-stages—that are conscious non-derivatively. For example, this seems to have been David Lewis's view. If it is inappropriate to double-count particular benefits and harms incurred by those who are conscious non-derivatively, then philosophers who follow Lewis should

reject double-counting not only in cases of intimately associated persons and animals but also in cases of fission and fusion; they should reject double-counting in cases of both kinds.

### 5.7.3 Double-Counting and the Number of Conscious Perspectives

Another important observation is that double-counting is intuitively more implausible in cases of intimately associated persons and animals than in cases of persons who undergo fission and fusion. For example, as I saw in the discussion of my example *Choice between Pains* in Chapter 4, it is very difficult to believe that if one experiences a horrible pain, then one's pain contributes only half as much to the total badness of the outcome if one is a brain in a vat rather than an ordinary person that is related in the usual way to a human animal. If the human animal would experience this very same pain in a derivative sense, this would have no fundamental moral importance. The outcome would not be *twice* as bad. On the other hand, it doesn't seem as absurd to suppose that if a certain person-stage will contain some amount of suffering, we could diminish the immediate contribution of this suffering to total badness by ensuring that the person-stage is a temporal part of *one* person rather than *two*. (For example, we might do this by preventing an anticipated instance of fission or fusion.) By ensuring this, we might ensure that there is *only one* diachronically unified conscious perspective from which it would be appropriate, as a matter of self-interest, to dread this suffering.

The foregoing suggests that the number of discernable conscious perspectives is morally important. When a person undergoes fission, and when multiple persons undergo

fusion, there are multiple unified perspectives that are clearly discernable. For example, in My Fission there are two such perspectives after my fission, and in Our Fusion there are two such perspective prior to our fusion. On the other hand, wherever there is exactly one person related in the usual way to a human animal, it seems there is only one diachronically unified conscious perspective from which prudential concerns about the future or past may be expressed. This makes double-counting in such cases seem especially implausible—even more implausible than double-counting in cases of personal fission and fusion. For instance, in my example *Fusion Rehabilitation*, David and Derek occupy different conscious perspectives; and from each of these perspectives, it makes good prudential sense to want *Early Fusion* rather than *Late Fusion*. This provides an intuitive rationale for double-counting. In cases involving only one unified perspective, the rationale for double-counting is much less intuitive.

#### **5.7.4 Double-Counting and the Claim that Human Animals are not, Strictly**

##### **Speaking, Conscious**

Next, there is the question of whether human animals are, strictly speaking, conscious. If they are not, then it seems there is no good reason to double-count in cases of intimately related persons and animals, even if there is good reason to double-count in cases of personal fission and fusion. I think a reductionist would say that the following two statements do not express genuinely competing hypotheses about reality, but that they merely express different descriptions of reality:

- (A) A normal adult human animal is not, strictly speaking, conscious, but it stands

in a certain intimate relation to a psychological continuer that is conscious.

- (B) A normal adult human animal is, strictly speaking, conscious, but only derivatively—only in virtue of standing in a certain intimate relation to a psychological continuer that is conscious independently or non-derivatively.

Reductionists who claim that (A) and (B) express what are merely different descriptions of reality might have reason to think that one of these descriptions is better than the other. But they might also deny that anything of moral importance turns on this. They might claim that what is of moral importance can be specified in a way that is completely neutral with respect to the question “Which of (A) and (B) is the better description?” It is hard to see what reason these reductionists would have for double-counting particular benefits and harms that occur within the life of a person that is related in the usual way to a human animal.

#### **5.7.5 Summary: Reductionists Who Accept Double-Counting in Cases of Fission and Fusion Can Accept the Personal Ontology Argument**

If there is no good reason to double-count *only* in cases of fission and fusion—and not in cases of intimately related persons and animals, then reductionists who accept double-counting in the former cases would have sufficient reason to accept double-counting in the latter cases. I believe these reductionists would also have sufficient reason to reject the Personal Ontology Argument. However, as we saw, these reductionists may indeed have good reason for double-counting *only* in cases fission and fusion. To justify this restriction, they can point to several differences between the two kinds of cases. They can

point to the fact that cases of fission and fusion involve multiple entities of the same kind, all of which are conscious non-derivatively, whereas cases of intimately related persons and animals involve two entities of different kinds—one that is conscious non-derivatively and another that is conscious only derivatively; they can argue that cases of personal fission and fusion involve different discernible conscious perspectives from which prudential concerns about future (and past) benefits and harms can be expressed, while in cases involving intimately related persons and animals, there is only one such conscious perspective; finally, they can say that human animals are not, strictly speaking, conscious, or that those who say this and those who say that human animals are, strictly speaking, conscious, but only in a derivative sense, are merely providing different descriptions of reality, and nothing of real moral significance depends on which of these descriptions is best.

If reductionists accept double-counting in certain cases, this wouldn't automatically give them reason to accept double-counting in the kinds of cases that are relevant to the success of the Personal Ontology Argument. These reductionists would still be able to accept this argument.

#### **5.7.6 Would a Radically Impersonal View of Aggregation Affect The Personal Ontology Argument?**

I have considered what we ought to say if we are reductionists who accept some form of double-counting. But if we are reductionists, then ultimately we may find double-counting too much to bear. In that case, we might reject the view that persons, or any person-like

entities, are the relevant value-bearers for aggregation. We might reject any account on which total goodness is determined by aggregating the fulfillment of different persons' self-interests, and we might accept an account on which total goodness is determined solely by aggregating temporal well-being (and ill-being) across arbitrarily fixed, non-overlapping time slices. This account of aggregation is radically impersonal, insofar as it rejects not only the importance of personal identity but that of persons and person-like entities as well. This radically impersonal account of aggregation poses no obvious challenge to the Personal Ontology Argument. However, this account might make it harder to justify abortion in cases in which having or performing an abortion would result in less temporal well-being. As I mentioned in Chapter 4, I am open to the possibility that abortion can be morally unjustified when it results in less total well-being.

## 5.8 Conclusion and Summary of Chapter 5

In setting up the problem of specifying the relevant value-bearers for aggregation, I initially considered a debate between Peter Vallentyne and James Cain regarding the quantification of infinite wellbeing. By showing how one action can produce a positive amount of wellbeing for every person even when it produces a negative amount of wellbeing at every time, Cain showed that it matters whether we take the relevant value-bearers to be persons or times. Vallentyne responded by claiming that, at least for utilitarians, the relevant value-bearers are *persons*, not times. I argued that whether there is a plausible account of what the value-bearers are for aggregation depends on the nature of personal identity. Specifically, if a *reductionist view* is true, then no matter how persons are individuated, our view of aggregation will fail to capture something that seems

important—either the moral importance of persons (and their self-interest) or that of temporal wellbeing. This problem arises even if we reject the idea that there is a morally relevant difference between promoting the good by creating people and promoting the good by improving the lives of those who currently exist (or who certainly will exist). Indeed, as I have shown, if persons are individuated in a certain way, the problem can arise even in cases involving finite wellbeing in which no one is created or destroyed.

Our account of aggregation fails to capture the moral importance of persons if it ranks outcomes with respect to their overall goodness (or badness) in a way that is insensitive (or only accidentally sensitive) to the fulfillment of the self-interests of the persons who exist in those outcomes. Our account fails to capture the moral importance of temporal wellbeing if it ranks outcomes in a way that is insensitive (or only accidentally sensitive) to differences in temporal wellbeing, as well as hypersensitive to very tiny differences in temporal wellbeing, as a result of double-counting quantities of temporal wellbeing (benefits and harms). It is therefore possible that if we are reductionists, then we must either reject the moral importance of persons or embrace double-counting.

If we embrace double-counting, then at first glance, it seems we have no good reason to accept the Personal Ontology Argument, since the defense of that argument requires that we reject double-counting. However, I argued that we may have reason to accept double-counting when two persons undergo fission or fusion, but not when a person and a human animal are associated in the usual way. In each kind of case, certain benefits and harms are in some sense “shared” by two entities, but there are crucial differences between these cases, including a difference in the number of conscious perspectives from which the “shared” benefits and harms may be regarded as prudentially good or bad.



Therefore, even if reductionists accept double-counting in certain cases, they would still have reason to reject double-counting in the cases that are relevant to my defense of the Personal Ontology Argument.

On the other hand, if we are reductionists, we might reject double-counting and accept the unimportance of persons (and person-like entities). In that case, our account of aggregation would be radically impersonal, but would present no obvious challenge to the Personal Ontology Argument.

## 6: Conclusion and Summary

In this final chapter, I will summarize the course of my investigation and discuss some of the remaining problems. A lot of material has been covered, and the summary that I present here differs somewhat from Chapter 1's introduction.

### 6.1 Summary of the Personal Ontology Argument and My Defense of It

In this dissertation, my central aim was to defend

*The Personal Ontology Argument:*

Premise 1: We are psychological continuers.

Premise 2: If we are psychological continuers, then killing a human animal and/or using it as a biological resource is morally justified if this is necessary (or the only reasonable way) to achieve a certain good—the fulfillment of a weighty self-interest of some psychological continuer—and apart from its impact on any human animal, it does not violate anyone's rights or produce any evil that nullifies or outweighs this good.

Therefore,

Killing a human animal and/or using it as a biological resource is morally justified if this is necessary (or the only reasonable way) to achieve a certain good—the fulfillment of a weighty self-interest of a psychological continuer—and apart from its impact on any human animal, it does not violate anyone's rights or produce any evil that nullifies or outweighs this good.

Chapters 2 and 3 jointly constitute my defense Premise 1. In Chapter 2, I argued that the case for the Psychological Approach is extremely powerful, and that the intuitions that motivate this approach cannot be easily explained away or accommodated by other approaches to personal ontology. In Chapter 3, I argued that Biological and Bodily Approaches do not seem to have significant theoretical advantages over the Psychological Approach, and that this, when taken together with the conclusion of Chapter 2, implies that Biological and Bodily Approaches are not as good (i.e. are not as theoretically plausible) as the Psychological Approach. I also argued that the Hybrid Approach is less plausible than the Psychological Approach. Even if this approach could accommodate the intuitions that motivate the Psychological Approach, it does do so at too great a cost. However, I argued, the Hybrid Approach may be unable to accommodate these intuitions. All things considered, I argued, we should reject this approach.

In Chapter 4, I established Premise 2 and the conclusion of the Personal Ontology Argument. For this purpose, I gave a positive argument, the Symmetry Argument, for the conclusion of the Personal Ontology Argument. (The plausibility of the Symmetry Argument depends on the Psychological Approach, which I defended in Chapters 1 and 2.) I defended the Symmetry Argument against objections. Finally, I discussed how the self-interest of psychological continuers might be weighed against the value of the biological life of human animals. I argued that when the fulfillment of a weighty self-interest of a psychological continuer is at stake, biological human life is a morally irrelevant consideration. I appealed to something like Kamm's principle of irrelevant utilities, which implies that in some contexts, certain kinds of goods are relevant to what we ought to do, but others are not. Finally, I argued that killing a human organism can be morally justified

even when the self-interest that is fulfilled is significantly less weighty than the self-interest most of us usually have in continuing to exist.

## **6.2 The Challenge of Reductionism**

In Chapter 5, I considered a potential challenge to Premise 2 of the Personal Ontology Argument. This potential challenge is posed by a reductionist views, according to which facts about persons and their identity over time just consist in certain more particular facts about bodily and psychological continuity. Reductionist views are not the only views that present this challenge, but for the sake of simplicity, I limited my focus to these views. I argued that if reductionist views are true, then no matter how persons are individuated, our view of aggregation will fail to capture something that seems important—either the moral importance of persons (and their self-interest) or that of temporal wellbeing. We fail to capture the moral importance of temporal wellbeing by double-counting benefits and harms in cases involving personal fission and fusion.

If we embrace this double-counting of benefits and harms, then it may seem that we should accept the Personal Ontology Argument, since the defense of that argument requires that we reject double-counting of benefits and harms. However, I argued that we may have reason to accept double-counting when two persons undergo fission or fusion, but not in the cases that matter for the Personal Ontology Argument—cases in which a person and a human animal are associated in the usual way. In each kind of case, certain benefits and harms are in some sense “shared” by two entities, but there are seemingly important differences between these cases, including a difference in the number of conscious perspectives from which the “shared” benefits and harms may be regarded as

prudentially good or bad. Therefore, even if reductionists accept double-counting in certain cases, they would still have reason to reject double-counting in the cases that are relevant to my defense of the Personal Ontology Argument.

On the other hand, if we are reductionists, we might reject double-counting and accept the unimportance of persons (and person-like entities). In that case, our account of aggregation would be radically impersonal, but would present no obvious challenge to the Personal Ontology Argument.

### 6.3 Concluding Remarks

I have defended *The Personal Ontology Argument*, but I do not know that this argument is sound. There is a great deal of *metaphysical uncertainty* surrounding Premise 1, and a great deal of *moral uncertainty* surrounding Premise 2. Furthermore, even if the argument is sound, there is a great deal of *empirical uncertainty* about which particular practices are morally justified given the truth of its conclusion. Resolving all of this uncertainty would be extremely valuable, but I am unable to resolve it. What should be done in light of the seemingly ineliminable uncertainty surrounding my argument, and how does this bear on the practical importance of that argument? These are very good questions. But at present I am unable to answer them. For that, I would need a theory of what it is rational to do (and believe) under conditions of metaphysical and moral uncertainty. But I do not have even a partial theory. For now, I can only point readers

toward some of the relevant literature on normative uncertainty.<sup>149</sup> Even if readers accept *The Personal Ontology Argument* they may think that the appropriate course of action, given our current moral uncertainty, is to prohibit the practices that this argument supports. I am inclined to think that, given our current moral uncertainty, at least some of these practices, such as embryonic stem cell research, should be allowed to continue. However, I am not as confident about all of the practices that I have considered. More careful research and thought is needed.

---

<sup>149</sup> Discussion of this topic can be found in Andrew Sepielli, "What To Do When You Don't Know What To Do," in *Oxford Studies in Metaethics*, ed. Russell Shaffer Landau, and Jacob Ross, "Rejecting Ethical Deflationism." *Ethics* (2006): 742-768.

## Bibliography

- Adams, Robert Merrihew. (1989) 1997. "Should Ethics Be More Impersonal?" In *Reading Parfit*, edited by Jonathan Dancy, 251–289. Oxford: Blackwell. Pagination refers to the Oxford edition.
- . 1999. *Finite and Infinite Goods: A Framework for Ethics*. New York: Oxford University Press.
- Agich, George J., and Royce P. Jones. 1986. "Personal Identity and Brain Death: A Critical Response." *Philosophy and Public Affairs* 15: 267–74.
- Anscombe, Elizabeth, 1981. *Ethics, Religion and Politics. The Collected Philosophical Papers of G.E.M. Anscombe*, vol. III, Oxford: Blackwell.
- Associated Press. 2011. *Brazil: Two Headed Baby Born*, retrieved from [http://www.huffingtonpost.com/2011/12/21/brazil-two-headed-baby\\_n\\_1163283.html](http://www.huffingtonpost.com/2011/12/21/brazil-two-headed-baby_n_1163283.html).
- Ayers, M. 1991. *Locke*, Vol, 2. London: Routledge.
- Baker, Lynn Rudder. 2000. *Persons and Bodies: A Constitution View*. Cambridge, MA: Cambridge University Press.
- . 2005. "When does a Person Begin?" *Social Philosophy and Policy* 22: 25-48.
- . 2007. *The Metaphysics of Everyday Life*. Cambridge, MA: Cambridge University Press.
- Bayne, Tim, and Barry Dainton. 2005. "Consciousness as a Guide to Personal Identity." *Australasian Journal of Philosophy* 83: 549–571.
- Becker, Gerhold K., ed. 2000. *The Moral Status of Persons: Perspectives on Bioethics*. Amsterdam-Atlanta, GA: Rodopi.
- Beckwith, Francis. 2007. *Defending Life: A Moral and Legal Case Against Abortion Choice*. New York: Cambridge University Press.
- Blatti, Stephan. 2007. "Animalism, Dicephalus, and Borderline Cases." *Philosophical Psychology* 20: 595-608.
- Block, Ned. 1995. "The Mind as the Software of the Brain." In *An Invitation to Cognitive Science*, edited by Daniel N. Osherson, Lila Gleitman, Stephen M. Kosslyn, S. Smith & Saadya Sternberg, 170-185. Cambridge: MIT Press.

- Block, Sergey, George Newman, Jennifer Behr, and Lance J. Rips. 2001. "Inferences About Personal Identity." In *Proceedings of the twenty-third annual conference of the cognitive science society*, edited by Johanna D. Moore and Keith Steinning, 80–85.
- Boonin, David. 1997. "Against the Golden Rule Argument Against Abortion" *Journal of Applied Philosophy* 14: 187–197.
- . 2001. *A Defense of Abortion*. New York: Cambridge University Press.
- Brill, H. Skott. 2003. "The Future-Like-Ours Argument, Personal Identity, and the Twinning Dilemma." *Social Theory and Practice* 29: 419–30.
- Brink, David O. 1990. "Rational Egoism, Self, and Others." In *Identity, Character, and Morality*, edited by O. Flanagan and A. Rorty, 339–378. Cambridge, MA: MIT Press.
- . 1997b. "Rational Egoism and the Separateness of Persons." In *Reading Parfit*, edited by Jonathan Dancy, 96–134. Oxford: Blackwell.
- Broome, John. 1991. "Utilitarian Metaphysics?" In *Interpersonal Comparisons of Well-Being*, edited by Jon Elster and John E. Roemer, 70–97. Cambridge: Cambridge University Press.
- . 2004. *Weighing Lives*. Oxford: Oxford University Press.
- Butler, Joseph. (1736) 1975. "Of Personal Identity." In John Perry, *The Analogy of Religion*, 99–105. Berkeley: University of California Press.
- Cain, James. 1995. "Infinite Utility." *Australasian Journal of Philosophy* 73: 401–404.
- Campbell, Tim, & Jeff McMahan. 2010. "Animalism and the Varieties of Conjoined Twinning." *Theoretical Medicine and Bioethics* 31: 285–301.
- Carter, W.R. 1982. "Do Zygotes Become People?" *Mind* 91: 77–95.
- . 1999. "Will I Be a Dead Person?" *Philosophy and Phenomenological Research* 59:167 – 171.
- Chalmers, David J. 1998. "The Problems of Consciousness." In *Consciousness: At the Frontiers of Neuroscience*, edited by H. Jasper, L. Descarriers, V. Castellucci and S. Rossignol, 29 – 59. New York, Lippincott-Raven.
- Conee, Earl. 1999. "Metaphysics and the Morality of Abortion." *Mind* 108: 619–46.
- Dainton, Barry. 1992. "Time and Division." *Ratio* 5: 102–128.
- Dainton, Barry, and Tim Bayne. 2005. "Consciousness as a Guide to Personal Persistence." *Australasian Journal of Philosophy* 83 no.4: 549–571.
- Dancy, Jonathan, ed. 1997. *Reading Parfit*. Oxford: Blackwell.
- Darwall, Stephen. 1982. "Scheffler on Morality and Ideals of the Person." *Canadian Journal of Philosophy* XII: 247–255.
- DeGrazia, David. 1999a. "Persons, Organisms, and the Definition of Death: A Philosophical Critique of the Higher-Brain Approach." *Southern Journal of Philosophy* 37: 419–40.
- . 1999b. "Advance Directives, Dementia, and 'the Someone Else Problem'." *Bioethics* 13: 373–391.
- . 2003. "Identity, Killing, and the Boundaries of Our Existence." *Philosophy and Public Affairs* 31: 413–442.
- . 2005. *Human Identity and Bioethics*. Cambridge: Cambridge University Press.



- . 2007. "The Harm of Death, Time-Relative Interests and Abortion." *The Philosophical Forum* 38: 57–80.
- Dennett, Daniel. 1976. "Conditions of Personhood." In *The Identities of Persons*, edited by Amelie Oksenberg Rorty, 175–196. Berkeley: University of California Press.
- Eddington, Arthur Stanley. 1928. *The Nature of the Physical World*. New York: Macmillan.
- Feldman, Fred. 1992. *Confrontations with the Reaper*. New York: Oxford University Press.
- . 2000. "Basic Intrinsic Value." *Philosophical Studies* 99 (3): 319–346.
- . 2000. "The Termination Thesis." *Midwest Studies in Philosophy* 24 (1): 98–115.
- Fischer, John Martin, and Daniel Speak. 2000. "Death and the Psychological Conception of Personal Identity." *Midwest Studies in Philosophy* XXIV: 84–93.
- Gendler, Tamar Szabo. 2002. "Personal Identity and Thought-Experiments." *The Philosophical Quarterly* 52: 34–54.
- Gensler, Harry J. 1985. "A Kantian Argument Against Abortion." *Philosophical Studies* 48:57–72.
- George, Robert and Patrick Lee. 2008. *Body-Self Dualism in Contemporary Ethics and Politics*. New York: Cambridge University Press.
- George, Robert and Christopher Tollefsen. 2008. *Embryo: A Defense of Human Life*. New York: Knopf Doubleday Publishing Group.
- Glover, Jonathan. 1977. *Causing Death and Saving Lives*. Harmondsworth: Penguin.
- Green, Michael, and Daniel Wikler. 1980. "Brain Death and Personal Identity." *Philosophy and Public Affairs* 9: 105–33.
- Grice, H.P. 1941. "Personal Identity." *Mind* 50 (200): 330–350.
- Hacking, Ian. 1991. "Two Souls in One Body." *Critical Inquiry* 17: 838–67.
- . 1995. *Rewriting the Soul: Multiple Personality and the Sciences of Memory*. Princeton, NJ: Princeton University Press.
- Hare, R.M. 1975. "Abortion and the Golden Rule." *Philosophy & Public Affairs* 4: 201–222.
- Hasker, William. 2001. *The Emergent Self*. Ithaca, NY: Cornell University Press.
- . 2005. "On Behalf of Emergent Dualism." In *In Search of the Soul: Four Views of the Mind-Body Problem*, edited by Joel B. Green and Stuart L. Palmer, 78. Downers Grove, IL: InterVarsity Press.
- Heathwood, Chris. 2011. "The Significance of Personal Identity to Abortion." *Bioethics* 25: 230–232.
- Hershenov, David. 2005. "Do Dead Bodies Pose a Problem for Biological Approaches to Personal Identity?" *Mind* 114: 31–59.
- . 2005. "Persons as Proper Parts of Organisms." *Theoria* 71: 29–37.
- Holland, S., K. Lebacqz, and L. Zoloth, eds. 2001. *The Human Embryonic Stem Cell Debate*. Cambridge, MA: MIT Press.
- Holtug, Nils. 2010. *Persons, Interests, and Justice*. New York: Oxford University Press.
- Hovorakova, M., Likovsky, Z., and Peterka, M. 2008. "A Case of Conjoined Twins: Cephalothoracopagus Janiceps Disymmetros." *Reproductive Toxicology* 26: 178–82.

- Hudson, Hud. 2001. *A Materialist Metaphysics of the Human Person*. Ithaca, N.Y.: Cornell University Press.
- Jamison, Dean, Shahid-Salles, Sonbol, Jamison, Julian, Lawn, Joy, and Zupan, Jelka. 2006. "Incorporating Deaths Near the Time of Birth into Estimates of the Global Burden of Disease." In *Global Burden of Disease and Risk Factors*, edited by AD Lopez, CD Mathers, M. Ezzate, D. Jamison, and C. Murray, Chapter 6. Washington, D.C.: World Bank.
- Johnston, Mark. 1987. "Human Beings." *Journal of Philosophy* 84: 59–83.
- Kamm, F.M. 1993. *Morality, Mortality: Death and Whom to Save from It*. New York: Oxford University Press.
- Kokcu, A., Cetinkaya, M.B., Aydin, O., and Tosun, M. (2007). "Conjoined Twins: Historical Perspective and Report of a Case." *The Journal of Maternal-Fetal and Neonatal Medicine* 20: 349-56.
- Korein, Julius. 1997. "Ontogenesis of the Brain in the Human animal: Definitions of Life and Death of the Human Being and Person." *Advances in Bioethics* 2: 1-74.
- Kripke, Saul. 1980. *Naming and Necessity*, 2nd edition. Cambridge, MA: Harvard University Press.
- Kuhse, Helga, and Peter Singer. 1990. "Individuals, Humans, and Persons: The Issue of Moral Status." In *Embryo Experimentation*, edited by Peter Singer, Helga Kuhse, Stephen Buckle, Karen Dawson and Pascal Kasimba, 65–75. Cambridge: Cambridge University Press.
- Lane, Robert. 2003. "Why I Was Never a Zygote." *The Southern Journal of Philosophy* 41: 63–83.
- Lee, Patrick and Germain Grisez. 2012. "Total Brain Death: A Reply to Alan Shewmon." *Bioethics* 26: 275-284.
- Lewis, David. 1971. "Counterparts of Persons and Their Bodies." *Journal of Philosophy* 68: 203–211.
- . 1976. "Survival and Identity." In *The Identities of Persons*, edited by Amelie Oksenberg Rorty, 17–40. Berkeley: University of California Press.
- Liao, Matthew. 2006. "The Animal View Defended." *The Monist* 89: 334–350.
- Lizza, John P. 1993. "Multiple Personalities and Personal Identity Revisited." *The British Journal for the Philosophy of Science* 44: 263–274.
- Locke, John. (1694) 1975. "Of Identity and Diversity." In *Essay Concerning Human Understanding* reprinted in John Perry, *Personal Identity*, 33–52. Berkeley: University of California Press.
- Mackie, David. 1999. "Animalism versus Lockeism: No Contest." *Philosophical Quarterly* 50 (196):369-376.
- MacIntyre, Alasdair. 1984. *After Virtue*. Notre Dame: University of Notre Dame Press.
- Madden, Rory. 2011. "Externalism and Brain Transplants" In *Oxford Studies in Metaphysics* 6, edited by Karen Bennett and Dean Zimmerman, 3. Oxford University Press: Oxford.
- Marquis, Don. 1989. "Why Abortion is Immoral." *Journal of Philosophy* 86: 183–202.
- . 1998. "A Future Like Ours and the Concept of Person: A Reply to McInerney and Paske." In *The Abortion Controversy*, edited by Louis P. Pojman and Francis J. Beckwith, 372–386. Belmont, CA: Wadsworth.

- Martin, Raymond, and John Barresi, eds. 2003. *Personal Identity*. Oxford: Blackwell Publishing.
- McDowell, John. 1997. "Reductionism and the First Person." In *Reading Parfit*, edited by Jonathan Dancy, Oxford: Blackwell.
- McInerney, Peter K. 1990. "Does a Fetus Already Have a Future-Like-Ours?" *Journal of Philosophy* 87: 264–268.
- McMahan, Jeff. 1995. "The Metaphysics of Brain Death." *Bioethics* 9: 91–126.
- . 1998. "Brain Death, Cortical Death, and Persistent Vegetative State." In *A Companion to Bioethics*, edited by H. Kuhse and P. Singer, 254–255. Oxford: Blackwell.
- . 2002. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press.
- . 2009. "Death, Brain Death, and Persistent Vegetative State." In *A Companion to Bioethics*, 2<sup>nd</sup> edition, edited by H. Kuhse, and P. Singer. Malden, MA: Wiley-Blackwell.
- Merricks, Trenton. 2001. *Objects and Persons*. New York: Oxford University Press.
- Miller, Kenneth. 1996. "Together Forever." *Life* (April): 44–56.
- Mills, Eugene. 1993. "Dividing without Reducing." *Mind* 102: 37–51.
- Nagel, Thomas. 1974. "What is it Like to be a Bat?" *Philosophical Review* 83: 435–50.
- Nichols, S. and M. Bruno. 2010. "Intuitions about personal identity: An empirical study" *Philosophical Psychology* 23(3): 293–312.
- Noonan, Harold. 2003. *Personal Identity*. New York: Routledge.
- Nozick, Robert. 1981. *Philosophical Explanations*. Oxford: Clarendon Press.
- Oderberg, David S. 1997. "Modal Properties, Moral Status, and Identity." *Philosophy and Public Affairs* 26: 259–298.
- Olson, Eric T. 1997a "Was I Ever a Fetus?" *Philosophy and Phenomenological Research* 58: 95–109.
- . 1997b. *The Human Animal: Personal Identity Without Psychology*. New York: Oxford University Press.
- . 2001. "Material Coincidence and the Indiscernibility Problem." *Philosophical Quarterly* 51 no.204:337–355.
- . 2003. "An Argument for Animalism." In *Personal Identity*, edited by Raymond Martin and John Barresi. 318–334. Oxford: Blackwell Publishing.
- . 2004. "Animalism and the Corpse Problem." *Australasian Journal of Philosophy* 82 no.2: 265–274.
- . 2007. *What Are We?* New York: Oxford University Press.
- . 2008. "Replies." *Abstracta* 4 (Special Issue I): 32–42.
- . 2010. "Ethics and the Generous Ontology." *Theoretical Medicine and Bioethics* 31: 259–270.
- Ord, Toby. 2008. "The Scourge: Moral Implications of Natural Embryo Loss." *The American Journal of Bioethics* 8.7: 12 – 19.
- Parfit, Derek. 1971. "Personal Identity." *The Philosophical Review* 80: 3–27.
- . 1972. "Later Selves and Moral Principles." In *Philosophy and Personal Relations*, edited by Alan Montefiore, 137–169. London: Routledge and Kegan Paul.
- . 1976. "Lewis, Perry, and What Matters." In *The Identities of Persons*, edited by

- Amelie Oksenberg Rorty, 91–107. Berkeley: University of California Press.
- . 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- . 1986. “Comments.” *Ethics* 96: 832–872.
- . 2001. “The Unimportance of Identity.” In *Identity*, edited by Henry Harris, 13–45. Oxford: Oxford University Press.
- . 2008. “Persons, Bodies, and Human Beings.” In *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne and Dean W. Zimmerman. Oxford: Blackwell Publishing.
- . 2012. “We Are Not Human Beings.” *Philosophy* 87 no.1, 5–28.
- Perry, John. 1975. *Personal Identity*. Berkeley: University of California Press.
- . 1978. *A Dialogue on Personal Identity and Immortality*. Indianapolis, IN: Hackett Publishing Company.
- Persson, Ingmar. 1999. “Our Identity and the Separability of Persons and Organisms.” *Dialogue* 38 (03): 519.
- Pruss, Alexander. 2006. “I Was Once a Fetus: That is Why Abortion is Wrong.” In *Persons, Moral Worth, and Embryos: A Critical Analysis of Pro-Choice Arguments*, edited by Stephen Napier. New York: Springer.
- Puccetti, Roland. 1988. “Does Anyone Survive Neocortical Death?” In *Death: Beyond Whole-Brain Criteria*, edited by Richard Zaner, 75–90. Dordrecht, The Netherlands: Kluwer.
- Quinn, Phillip. 2007. “On the Intrinsic Value of Persons.” In *Persons, Human and Divine*, edited by Peter van Inwagen and Dean Zimmerman, 237 – 260. New York: Oxford University Press.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: The Belknap Press of Harvard University Press.
- . (1975) 1999. “The Independence of Moral Theory.” *Proceedings and Addresses of the American Philosophical Association* XLVIII: 5–22. Reprinted in John Rawls, *Collected Papers*, edited by Samuel Freeman, 286–302. Cambridge, MA: Harvard University Press.
- . 1992. *Political Liberalism*. New York: Columbia University Press.
- Reid, Thomas. (1785) 1975. “Of Memory.” In *Essays on the Intellectual Powers of Man*. Reprinted in John Perry, *Personal Identity*, 107–112. Berkeley: University of California Press. Page references are to the reprinted version.
- Rips, L. J., Blok, S., and Newman, G. 2006. “Tracing the Identity of Objects.” *Psychological Review* 113: 1–30.
- Ross, Jacob. 2006. “Rejecting Ethical Deflationism.” *Ethics*: 742–768
- Rovane, Carol. 1993. “Self-Reference: The Radicalization of Locke.” *Journal of Philosophy* 90: 73–97.
- Schechtman, Marya. 1996. *The Constitution of Selves*. Ithaca: Cornell University Press.
- Sepielli, Andrew. 2009. “What To Do When You Don’t Know What To Do.” In *Oxford Studies in Metaethics* 4: 5–28.
- Shewmon, D. Alan. 1998. “Brainstem death, Brain Death, and Death: A Critical Reevaluation of the Purported Equivalence.” *Issues in Law and Medicine* 14: 125–45.

- Shoemaker, David W. 1999. "Selves and Moral Units." *Pacific Philosophical Quarterly* 80: 391–419.
- . 2007. "Personal Identity and Practical Concerns." *Mind* 116: 316–357.
- . 2009. *Personal Identity and Ethics: A Brief Introduction*. Ontario, Canada: Broadview Press.
- Shoemaker, Sydney. 1970. "Persons and Their Pasts." *American Philosophical Quarterly* 7: 269–285.
- . (1985) 1997. "Critical Notice of Derek Parfit's *Reasons and Persons*." *Mind* 94: 443–453. Reprinted as "Parfit on Identity." In *Reading Parfit*, edited by Jonathan Dancy, 135–148. Oxford: Blackwell.
- . 1996. *The First-Person Perspective and Other Essays*. Cambridge: Cambridge University Press.
- . 1999. "Self, Body, and Coincidence." *Proceedings of the Aristotelian Society* 73: 287–306.
- . 2008. "Persons, Animals and Identity." *Synthese* 162 (3): 313 – 324.
- Sider, Theodore. 2001a. *Four-Dimensionalism*. Oxford: Oxford University Press.
- . 2001b. "Criteria of Personal Identity and the Limits of Conceptual Analysis." *Philosophical Perspectives* 15: 189–209.
- . 2013. "The Evil of Death: What Can Metaphysics Contribute." In *The Oxford Handbook of Philosophy and Death*, edited by Ben Bradley, Fred Feldman, and Jens Johansson, 155 – 166. New York: Oxford University Press.
- Smith, Barry and Berit Brogaard. 2003. "Sixteen Days." *Journal of Medicine and Philosophy* 28: 45–78.
- Snowdon, P.F. 1990. "Persons, Animals, and Ourselves." In *The Person and the Human Mind*, edited by Christopher Gill, 83–107. Oxford: Clarendon Press.
- Sosa, Ernest. 1990. "Surviving Matters." *Nous* 24: 297–322.
- . 2000. "Contextualism and Skepticism." In *Philosophical Issues*, Supplement to *Nous* 34, edited by J. Tomberlin, 1–18.
- Stone, Christopher. 1972. "Should Trees Have Standing? – Toward Legal Rights for Natural Objects." *Southern California Law Review* (1972): 45: 450–87.
- Strauss, L.T., Gamble, S.B., Parker, W.Y., Cook, D.A., Zane, S.B., and Hamdan, S. 2006. "Abortion Surveillance – United States, 2003." *Morbidity and Mortality Weekly Report*, 55: 1–32.
- Strawson, Galen. "What is the Relation between an Experience, the Subject of the Experience, and the Content of the Experience?" unpublished manuscript.
- Thomson, J. J. 2008. "People and Their Bodies". In *Contemporary Debates in Metaphysics*, edited by Theodore Sider, John Hawthorne and Dean W. Zimmerman, 155–176. Malden, MA: Blackwell Publishing.
- Tooley, Michael. 1972. "Abortion and Infanticide." *Philosophy and Public Affairs*, 2: 37–65.
- Unger, Peter. 1990. *Identity, Consciousness, and Value*. New York: Oxford University Press.
- Vallentyne, Peter. 1993. "Utilitarianism and Infinite Utility." *Australian Journal of Philosophy* 71: 212–217.
- . 1995. "Infinite Utility: Anonymity and Person-Centeredness." *Australasian Journal*

- of Philosophy* 73: 413-420.
- van Inwagen, Peter. 1990. *Material Beings*. Ithaca, NY: Cornell University Press.
- Vyasa, Krishna, and Ganguli, Kisari. 1886. *Mahabharata*. Current English translation Retrieved from <http://books.google.com/books?id=A9RZAAAAMAAJ&pg=PA59&lpg=PA59&dq>
- White, Robert, Wolin, L.R., Massopust, L.C. Jr., Taslitz, N., and Verdura, J. 1971. "Cephalic Exchange Transplantation in the Monkey" *Surgery* Jul: 70 no. 1: 135 – 39.
- Wiggins, David. 1980. *Sameness and Substance*. Cambridge, MA: Harvard University Press.
- . 2001. *Sameness and Substance Renewed*, 96. New York: Cambridge University Press.
- Williams, Bernard. 1970. "The Self and the Future." *The Philosophical Review* 79: 161–180.
- . 1973. *Problems of the Self*. Cambridge: Cambridge University Press.
- Zimmerman, Dean. 2001. "The Constitution of Persons by Bodies: A Critique of Lynne Rudder Baker's Theory of Material Constitution." *Philosophical Topics* 30: 295-338.