CAPACITIES AND MORAL STATUS

Russell DiSilvestro

A Dissertation

Submitted to the Graduate College of Bowling Green State University in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2006

Committee:

R. G. Frey, Advisor

Younghhee Kim
Graduate Faculty Representative

Fred Miller

Loren Lomasky

Michael Tooley
ABSTRACT

R. G. Frey, Advisor

The thesis of this essay is that, because human organisms have the specific sorts of capacities that they do, they have the sort of moral status (which I call “serious” moral status) that includes, as one of its components, a strong moral presumption against being killed. The main argument of this essay has three steps: if an entity is human, it has a set of typical human capacities; if an entity has a set of typical human capacities, it has serious moral status; therefore, if an entity is human, it has serious moral status. Typical human capacities include both “active” capacities and “passive” capacities, and also include both “lower-order” capacities and “higher-order” capacities. Although both distinctions are somewhat rough, I have an active capacity to raise my arm on purpose, a passive capacity to feel pain when pricked, lower-order capacities to do both of these things right now, and higher-order capacities to obtain these lower-order capacities when asleep, anesthetized, or temporarily comatose. Whereas most writers who focus on the moral relevance of capacities tend to ignore an entity’s higher-order capacities, or its passive capacities, or both, I argue that an entity’s passive higher-order capacities are relevant to its moral status.

The five chapters of this essay are structured as follows. Chapter One explains the concepts of the main argument in more detail, and explains why personal pronouns such as “you” and personal names such as “Ronald Reagan” are applied to human organisms throughout the essay. Chapters Two and Three defend the first and second premises of the main argument, by focusing on human organisms who undergo temporary changes involving “incapacitation” of
one sort or another. Chapters Four and Five relate the main argument to two controversial arguments in contemporary applied ethics: the Argument From Potential, which focuses on normal human organisms at the beginning stages of life (such as human infants, fetuses, and embryos), and the Argument From Marginal Cases, which focuses on abnormal human organisms (such as human organisms that are disabled, diseased, or genetically deficient in some way).
To Heather
ACKNOWLEDGMENTS

I would like to thank several groups of people, without whom the writing of this dissertation would have been much more difficult, if not impossible. First, my committee, who generously gave both a great deal of time and a good deal of effort in various ways to help the dissertation come together: R. G. Frey, Fred Miller, Loren Lomasky, Michael Tooley, and Younghee Kim. Second, the members of Fred Miller’s dissertation reading group, who gave me helpful feedback on each part of the dissertation over a period of several months: Pete Cellelo, Kathy Erbeznik, Leanne Kent, Nico Maloberti, Chris Metivier, John Milliken, Jonathan Miller, Sangeeta Sangha, and Matt Stichter. Third, the different groups who provided me with financial support during the writing of the dissertation: The Bowling Green State University Philosophy Department, The Earhart Foundation, and The Charlotte and Walter Kohler Charitable Trust, who support the Kohler Fellows program at the Howard Center for Family, Religion, and Society. Finally, my immediate and extended family, who gave me various kinds of valuable support throughout the process: my wife Heather, my children Hannah, Grace, Bethany, and Joshua, my parents Ruth and Frank, my siblings Frank and Elizabeth, and my in-laws Phillip, JoAnne, Ian, Sean, and Nick McDaniel.
# TABLE OF CONTENTS

INTRODUCTION ................................................................................................................. 1

1. Topic, Thesis, and Main Argument of This Essay ......................................................... 1

2. Summaries of Each Chapter in This Essay .................................................................. 2

CHAPTER ONE: CAPACITIES, HUMAN ORGANISMS, AND PERSONS ....................... 6

1. Prologue: The Adventure of Reagan’s Brain .................................................................. 6

2. What Humans Are ......................................................................................................... 14

3. What Serious Moral Status Is ...................................................................................... 17

4. What Typical Human Capacities Are .......................................................................... 24

5. What Persons Are ......................................................................................................... 30

6. What We Are ............................................................................................................... 34

CHAPTER TWO: HUMANS AND THEIR CAPACITIES ................................................. 41

1. Humans Have Capacities .............................................................................................. 41

1.1. Ordinary Language Speaks about Capacities in Various Ways.............................. 43

1.2. Capacities May Be Active or Passive, Identity-preserving or Compositional, General or Specific ................................................................................................................. 45

2. These Capacities Come in Hierarchies ......................................................................... 48

2.1. Several Philosophers Have Recognized That Capacities Come in Hierarchies ........ 48

2.2. Hierarchies of Capacities Can Be Clarified by the Concept of Developmental Distance ......................................................................................................................... 57

3. As Long As You Are Human, You Have a Certain Set of Higher-Order Capacities .... 64
<table>
<thead>
<tr>
<th>Chapter</th>
<th>Title</th>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Three</td>
<td>Why Capacities Matter</td>
<td>1. Another Temporary Change Argument</td>
<td>73</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.1. A First Strategy for Replying</td>
<td>78</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.2. A Second Strategy for Replying</td>
<td>81</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.3. A Third Strategy for Replying</td>
<td>84</td>
</tr>
<tr>
<td></td>
<td></td>
<td>1.4. Implications of the Temporary Change Argument</td>
<td>86</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. An Alternative Account: Actual, Continuing Subjects of Experience</td>
<td>90</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.1. The Account Itself</td>
<td>91</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.2. Complete Reprogramming and Other Challenges to the Account</td>
<td>93</td>
</tr>
<tr>
<td>Four</td>
<td>Capacities and the Argument from Potential</td>
<td>1. The Argument from Potential and the Argument of This Chapter</td>
<td>100</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2. Potential and Higher-Order Capacities</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.1. Normal Human Infants</td>
<td>105</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.2. Beyond Normal Human Infants</td>
<td>111</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.3. Objections and Replies</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.3.1. You Can Only Survive Certain Sorts of Upper Brain Damage: Tooley</td>
<td>113</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.3.2. You Could Not Have Been an Embryo: McMahan</td>
<td>124</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.3.3. You Could Not Have Been a Zygote: Singer and Kuhse</td>
<td>125</td>
</tr>
<tr>
<td></td>
<td></td>
<td>2.3.4. Even Given Identity, It Begs the Question to Assume Serious Moral Status: Boonin</td>
<td>130</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3. Higher-Order Capacities and the Argument from Potential</td>
<td>132</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.1 Human Germ Cells and Somatic Cells</td>
<td>132</td>
</tr>
</tbody>
</table>
3.2. Potential Presidents and Potential Persons ............................................. 138
   3.2.1. The Potential President Example ............................................... 139
   3.2.2. Re-electing the Potential President .......................................... 147

CHAPTER FIVE: CAPACITIES AND THE ARGUMENT FROM MARGINAL CASES 156
   1. The Argument from Marginal Cases and the Argument of This Chapter .......... 156
   2. Marginal Cases and Higher-Order Capacities ................................................. 161
      2.1 The Problem: Some Marginal Cases Appear Not To Have Any Higher-
      Order Capacity To Think ........................................................................ 161
      2.2. The Solution: Such Marginal Cases Have a Higher-Order Capacity To
      Think, But Need Technologically Sophisticated Therapy in Order To
      Realize It ................................................................................................. 165
         2.2.1. Marginal Cases Whose Marginality Is Due To Damage Have
         a Higher-Order Capacity To Think, But Need Brain Therapy
         of Various Kinds To Realize It ............................................................. 165
         2.2.2. Marginal Cases Whose Marginality Is Due To an Inherited
         Condition Have a Higher-Order Capacity To Think, But Need
         Genetic Therapy To Realize It ............................................................. 167
      2.3. An Objection: This Solution Is Too Generous in Ascribing a Capacity
      To an Individual ..................................................................................... 171
   3. Higher-Order Capacities and the Argument from Marginal Cases ............... 178
      3.1. The Main Dilemma ........................................................................... 178
      3.2. The First Horn of the Dilemma ......................................................... 179
         3.2.1. Tooley’s Cat, Boonin’s Spider, and McMahan’s Dog ............... 179
3.2.2. It Is Not Possible To Perform These Transformations on a Non-Human Animal

3.3. The Second Horn of the Dilemma

CONCLUSION

BIBLIOGRAPHY
INTRODUCTION

1. Topic, Thesis, and Main Argument of This Essay

The topic of this essay is the relationship between an entity’s capacities and its moral status. This relationship is important because many debates about the moral status of an entity eventually migrate towards a discussion of its capacities, such as the capacity to feel pain, or the capacity for self-consciousness, or the capacity to think rationally.

The thesis of this essay is that, because human organisms have the specific sorts of capacities that they do, they have the sort of moral status—what I shall call “serious moral status”—that includes, as one of its components, a strong moral presumption against being killed. This thesis is narrow because it focuses upon just one sort of entity and just one aspect of serious moral status, leaving to one side the many other sorts of entity that could have serious moral status and the many other interesting aspects of serious moral status itself.

The main argument of this essay has three steps:

1. If an entity is human, it has a set of typical human capacities.
2. If an entity has a set of typical human capacities, it has serious moral status.

Therefore,

3. If an entity is human, it has serious moral status.

This argument is sound because typical human capacities include both “active” capacities and “passive” capacities, and also include both “lower-order” capacities and “higher-order” capacities. Although both distinctions are somewhat rough, I have an active capacity to raise my arm on purpose, a passive capacity to feel pain when pricked, a lower-order capacity to do both of these things right now, and a higher-order capacity to do both of these things when asleep,
anesthetized, or temporarily comatose. Whereas most other writers who focus on the moral relevance of capacities tend to ignore an entity’s higher-order capacities, or its passive capacities, or both, I argue that an entity’s passive higher-order capacities are relevant to its moral status.

So, then, this essay brings together a debate in metaphysics, over the nature of capacities, and a debate in ethics, over the nature of moral status. I explain why the metaphysical distinctions between lower-order and higher-order capacities and between active capacities and passive capacities are helpful for answering questions in the moral arena concerning killing, and why the position I advance is better at solving certain problems than other positions.

2. Summaries of Each Chapter in This Essay

The five chapters of this essay are structured as follows. Chapter One explains the concepts of the main argument in more detail. Chapters Two and Three defend the first and second premises of the main argument. Chapters Four and Five relate the main argument to two controversial arguments in contemporary applied ethics.

Chapter One explains the sense of “human” that appears in the main argument, the sense of “moral status” that is operative in phrases like “serious moral status” and “the strong moral presumption against killing”, and the meaning of the phrase “typical human capacities.” This chapter also discusses how the concept of a person will be employed in this essay, and explains why personal pronouns such as “you” and personal names such as “Ronald Reagan” are applied to human organisms throughout the essay.

Chapter Two defends the first premise of the main argument: if an individual is human, it has a set of typical human capacities. Another way of putting this claim is this: there is some set, H, of capacities, such that for any individual X, if X is human, then X has H. This chapter
explains in more detail the distinctions between different kinds of capacities, such as the distinction between lower-order and higher-order capacities and the distinction between active and passive capacities. Finally, this chapter argues that an entity has the capacities in the set of typical human capacities as long as it has these capacities at some level or other: whether this level is higher-order or lower-order is not important.

While Chapter Two is about metaphysics, Chapter Three is about ethics. It defends the second premise of the main argument: if an individual has a set of typical human capacities, it has serious moral status. The main argument of this chapter is that there are cases of human organisms that are in such a state that the only possible basis for their serious moral status is their set of higher-order capacities. Since most of us—philosophers and non-philosophers alike—do believe that these human organisms have serious moral status, and since most of us do base this serious moral status on something, it must be the higher-order capacities of the organisms that we base it on. Importantly, this argument does not begin with the assumption that normal human infants have serious moral status. Therefore, even philosophers who do not start with that assumption can still accept this argument.

There are two main problem areas that threaten to undermine each of the two premises in the main argument. The first concerns normal human organisms at the beginning stages of life—such as human infants, fetuses, and embryos. The second concerns abnormal human organisms—so-called “marginal cases”—such as human organisms that are disabled, diseased, or genetically deficient in some way. Chapters Four and Five tackle these two problem areas, and relate them to two of the most hotly debated arguments in contemporary applied ethics: the Argument From Potential and the Argument From Marginal Cases.
Chapter Four argues that even human infants, fetuses, embryos, and zygotes have the typical human capacities that are sufficient to generate serious moral status. But this seems to make the main argument into a version of the Argument From Potential. The Argument From Potential has often been accused of leading to absurd conclusions and of relying on mistakes in moral reasoning. Therefore, Chapter Four also explains why the main argument of this essay does not lead to the absurd conclusion that human gametes or somatic cells have serious moral status, and that the main argument does not rely on any of the alleged mistakes in moral reasoning that are made by “potentiality” arguments in ethics.

Chapter Five argues that even the most “marginal” human organisms still have the typical human capacities that are sufficient to generate serious moral status. This includes even human organisms that are in a persistent vegetative state and human organisms that have inherited terrible genetic deficiencies. Since it is possible that future technological advances in regenerative medicine and genetic therapy will enable these unfortunate human organisms to overcome their current conditions, it follows that these unfortunate human organisms have the relevant higher-order capacities right now. Such humans do not lack the relevant capacities; they merely lack the external assistance required to actualize these capacities. But this seems to lead directly to a version of the Argument From Marginal Cases, which would entail that, if these marginal human organisms really do have serious moral status, it follows that certain non-human animals have serious moral status as well. After all, it seems possible that future technological advances will enable non-human animals to develop the ability to do all the things that normal human organisms do. Therefore, Chapter Five also explains why the main argument of this essay does not entail that the non-human animals we are aware of have serious moral status, and
why the main argument does not entail a morally objectionable sort of “speciesism” or “anthropocentrism.”

If the argument of this essay is sound, this will have important implications for the proper way of framing a number of debates about killing human organisms, such as debates about killing in war, debates about killing in self-defense, debates about killing as state punishment, and debates about killing in the context of biomedical ethics. For if all human organisms really do have serious moral status, then there is a strong moral presumption against killing them—whether or not they are enemy combatants, whether or not they are attacking us with lethal force, whether or not they have committed capital offenses, and whether or not they are very undeveloped, diseased, or damaged.
CHAPTER ONE: CAPACITIES, HUMAN ORGANISMS, AND PERSONS

1. Prologue: The Adventure of Reagan’s Brain

To grasp the basic argument of this essay, it helps to start with a vivid example. When former president Ronald Reagan died from Alzheimer’s disease, it added new fuel to the ongoing moral and political debate about embryonic stem-cell research. Proponents of such research emphasized that it might lead to the discovery of cures for diseases like Alzheimer’s, and that it should therefore be aggressively funded by the federal government. Opponents of embryonic stem-cell research, on the other hand, emphasized that former president Reagan would have remained staunchly opposed to it, since this research typically involves the destruction of living human embryos, and he believed that every human being—here understood as a human organism—possessed a sacred and inviolable dignity from its conception until its natural death.

Assume for a moment, what may after all be true, that human embryonic stem-cell research will one day lead to a regenerative cure for Alzheimer’s disease. This sort of cure, let us assume, could work in the following restorative way: if the relevant stem-cell tissues were injected into a certain region of a patient’s brain after Alzheimer’s disease had done its characteristic damage to that region, then that region of the brain would become capable of being restored to the state it was in before the disease had done its characteristic damage.¹

In addition, assume for a moment, what was unfortunately not true, that such a cure for Alzheimer’s disease became available to Reagan’s doctors, but only after the disease had progressed quite far. Assume that the cure was put in their hands before Reagan’s biological

¹ If you have trouble getting very enthusiastic about using a cure obtained from stem-cell research that destroys human embryos, then change the example so that the cure was obtained from research that you consider morally unobjectionable: for example, from adult stem-cell research, which destroys no human embryos.
death, but after Alzheimer’s disease had done its characteristic damage on those regions of
Reagan’s brain that supported his distinctive personality. Assume that the disease had destroyed
those brain structures that supported Reagan’s distinctive memories, beliefs, and desires.
Assume further that the disease had also destroyed those brain structures that supported
perception, conscious self-awareness, and all other mental states. Assume that, although a living
human organism remained, there were no experiences at all being had by that organism.

Reagan’s doctors inject the relevant stem-cell tissues into the relevant region (or regions:
but for simplicity, assume it is just one region) of Reagan’s brain. That region of the brain
becomes capable of being restored to the state it was in before the disease had done its
characteristic damage. In a way similar to the way a starfish starts growing back a limb that has
been removed, Reagan’s brain starts growing back the structures that had been destroyed by
Alzheimer’s disease. If the original generation of brain structures in Reagan’s developmental
neurobiology had been captured on videotape, then watching this regeneration of brain tissues in
Reagan’s brain would be like watching a rerun of parts of the original videotape.

Even though the injection makes a region of Reagan’s brain capable of being restored to
its pre-disease state, there is no guarantee that this region will be restored to exactly the same
structure it had before the disease took its toll. For even if we assume that Reagan’s brain
develops the characteristic structures necessary for supporting perception, conscious self-
awareness, and other mental states, the exact details of those characteristic structures might be
very different than the structures that supported the distinctive memories, beliefs, and desires that
most people would associate with Reagan. For all we know, the restoration might result in the
brain structures characteristic of Bill Clinton. This might be viewed as a medical miracle gone
badly wrong indeed (especially if you are a Republican!).
No doubt Reagan’s doctors, friends and family would do all that was in their power to make sure that his brain would not eventually gain the brain structures characteristic of Bill Clinton—or, more precisely (and more seriously), they would do all they could to make sure that his brain would eventually regain the exact same structures that it had before the disease. They would try to speak and act in such a way that resultant brain structures supporting beliefs, desires, and so on would be as similar as possible to the original brain structures that supported beliefs, desires, and so on before the disease began its destructive course. For example, the doctors, friends and family would show their recovering patient lots of old photographs and newspaper articles, play him countless hours of old speeches, debates, and documentaries, and even recount lots of old stories that only they and Reagan could possibly know about. They would have their recovering patient read both biographies and autobiographies of Reagan, and would serve their recovering patient all of Reagan’s favorite foods, play all of Reagan’s favorite music and movies, and visit all of Reagan’s favorite places. All this would be done in an attempt to give their recovering patient as close an approximation as possible of the pre-disease personality.

Nevertheless, as I have described this case so far, there would be very real limitations to what even the most sincere and determined efforts of Reagan’s friends and family could achieve by way of recreating Reagan’s distinctive beliefs, desires and so on. But these limitations can be overcome by introducing another bit of imaginary technology into the case. Imagine that the doctors have at their disposal a ‘brain scanner’ similar to the sort used in the recent Arnold Schwarzenegger movie *The Sixth Day.* Such a brain scanner can do two things. First, it can

---

2 You may object at this point that introducing movie-stars-turned-governors-of-California is too extravagant for even a philosophical thought experiment. But this objection cuts no ice for the family of Reagan, since Reagan was a movie-star-turned-governor-of-California himself. The truth (of history) is often stranger than (science-) fiction, and sometimes even repeats itself.
record a “snapshot” of the exact structure of a given brain at a given time—call this ‘scanning out’. Second, it can reproduce that exact structure into the same brain at a different time, and/or a different brain at the same time, and/or a different brain at a different time—call this ‘scanning in’. The only limitation is that any brain being ‘scanned into’ must already have a certain degree of structure to begin with. Although the brain scanner can restructure your brain so that it has the structure of Bill Clinton’s brain, it cannot do the same thing to your intestines.

Assume Reagan’s doctors use the brain scanner as follows. First, they ‘scan-out’ Reagan’s brain structure on a certain agreed-upon day before the Alzheimer’s disease has had much of a chance to run its destructive course. Second, they reproduce that exact structure into Reagan’s regenerated brain, after the brain has been given the stem-cell injection and has been able to develop the minimum structure needed for the ‘scan-in’. The result is that the brain structures supporting beliefs, desires, and so on, on the day of the ‘scan-in’, would be exactly the same as the brain structures that supported beliefs, desires, and so on, on the day of the original ‘scan-out’.

Call this example The Adventure of Reagan’s Brain. I would like to pose the following basic question about it: what happens to Reagan in this adventure?

It might seem odd that I think it makes sense to pose this question. After all, have not the last several paragraphs been discussing precisely what happens to Reagan? Would not someone who has read these paragraphs already know, in a more or less schematic way, what happens to Reagan?

Although it may be true that this question has a ring of redundancy for many people who have already read the last several paragraphs, the fact is that there are sharply different ways of
answering it. This can be seen most clearly by presenting my own answer to the question and by explaining some typical ways of disagreeing with my answer.

My answer to this question has two parts. First, Reagan survives each stage of this adventure. Second, Reagan retains his moral status during each stage of this adventure. This answer may seem to be somewhat uncontroversial. But since some will dispute each part of this answer, and since I eventually rely on this answer to argue for some very controversial claims below, I should say a bit more about what I mean.

First, Reagan survives each stage of this adventure. He was there at the beginning, before the disease began its work. He was still there in the middle, after the disease had taken its terrible toll. And he was still there at the end, after the injection and scan-in. Reagan persisted throughout the adventure. It was, in fact, his adventure, and the different parts of the adventure, such as the psychological deterioration and recovery, were his psychological deterioration and his recovery.

Some will object that Reagan literally ceased to exist at some point in this adventure. What really survives each stage of the adventure, this objection will maintain, is a living human organism. A living human organism was there at the beginning, before the disease began its work. That same living human organism was there in the middle, after the disease had taken its terrible toll. And that same living human organism was there at the end, after the injection and scan-in. A living human organism persists throughout the adventure. But Reagan does not. For Reagan is not a living human organism. According to this objection, Reagan is something else.

Different versions of this objection may differ on precisely what sort of entity Reagan really is. Some will say that he is a living human organism with a particular sort of brain structure, but that once that particular sort of brain structure is gone, Reagan is gone as well,
even though the living human organism remains. Others will say that he is not a living human organism of any kind, but rather is a kind of psychological entity that happens to be associated with a living human organism. Those who take this psychological route may differ among themselves about precisely what sort of psychological entity Reagan is: on one formulation, Reagan is an actual, continuing subject of experiences; on another formulation, he is a different sort of psychological entity. Nevertheless, despite their differences, all versions of this objection are united in their denial that Reagan is the living human organism that persists through this adventure. They all gloss their relevant concept of the sort of entity Reagan is—“actual, continuing subject of experiences”, for example—in such a way that this sort of entity is not a living human organism. Reagan may be associated with a living human organism. Reagan may be dependent in important ways upon a living human organism. But according to this objection, Reagan is not a living human organism.

Consequently, this objection maintains, the mere fact that a living human organism persists through the adventure does not guarantee that Reagan persists through the adventure. In particular, by some point in the progression of the Alzheimer’s disease, Reagan ceases to exist. Again, different versions of this objection will differ on where this point is: on some versions, Reagan ceases to exist when the disease destroys the brain structures that support his particular personality traits; on other versions, Reagan ceases to exist when the disease destroys the brain structures that support his particular memories. (Different versions of this objection will also differ on whether there really is one single point before which Reagan exists and after which he does not. But I shall henceforth ignore this wrinkle.) But according to all versions of the objection under discussion, the living human organism in the middle of the adventure is not
Reagan. Indeed, that living human organism is not anybody at all. It is just an anonymous, impersonal nobody.

I believe this objection is mistaken. If a given theory of personal identity commits one to making this objection, then this is a mark against that theory. If a given theory of personal identity cannot accommodate the claim that it was Reagan himself that survived this adventure, then this is a good reason for rejecting the theory, and not a good reason for rejecting the claim. Cases like The Adventure of Reagan’s Brain make especially clear the counterintuitive implications of, and the high ontological price of, certain theories of personal identity. We can all imagine perfectly well—or so I say—what would be involved in Reagan himself getting his brain repaired in the way described in this adventure.

So much (for now) for the first part of my answer. The second part of my answer was that Reagan retains his moral status during each stage of this adventure. This can be illustrated with an example. Imagine one of Reagan’s enemies sneaks into his hospital room, the night before the stem-cell injection, and ends his life (I am assuming that it really is his life) by giving him a painless injection of some drug that stops his heart from beating. I claim that Reagan’s enemy has violated Reagan’s moral status. This is a wrong-making feature of his action. Of course, his action could have possessed additional wrong-making features, if, for example, he had killed Reagan painfully, before the Alzheimer’s disease had begun, in the presence of Reagan’s family and friends. But this will not distract us from the fact that, by ending Reagan’s life the night before his stem-cell injection, Reagan’s enemy violated Reagan’s moral status.

Some will deny this second part of my answer. Most who deny it will do so because they deny the first part. “After all,” they might say, “although Reagan’s enemy killed a living human organism, he did not kill Reagan. Reagan was not there to be killed.” But it is possible to
believe the first part of my answer and still deny the second part. Perhaps Reagan’s moral status is a fiction: it doesn’t exist to begin with, so he can’t lose it even when Alzheimer’s takes its toll. Or perhaps Reagan’s moral status is real, but is the sort of thing that fluctuates over the course of his biological life depending on what else is true of him. When his personality is intact, before the disease runs its course, Reagan has moral status, but when the disease obliterates his personality, Reagan’s moral status is obliterated along with it. (It is possible to believe the second part of my answer without believing my first part. But only if you believe that Reagan can have moral status at times when he does not exist. I shall not explore this option further.)

I believe that it is a mistake to deny this second part of my answer. If a given theory of moral status commits one to denying this part of my answer, then that is a mark against that theory. If a given theory of moral status cannot accommodate the claim that Reagan himself retains his moral status, then this is a good reason for rejecting the theory, and not a good reason for rejecting the claim. Cases like The Adventure of Reagan’s Brain make especially clear the counterintuitive implications of, and the high moral price of, certain theories of moral status. We can all see quite clearly—so I say—that Reagan retains his moral status in this adventure.

If you disagree with either part of my answer, I invite you to consider them as elements in what I take to be a “better picture” of both personal identity and moral status than the pictures often presented by contemporary philosophers who write on these topics. These parts of my answer are my starting points, and this essay is an attempt to both defend these starting points from rival accounts of personal identity and moral status, and to develop from these starting points an alternative approach to personal identity and moral status. When this is done, I believe that we will be led to a view that (a) identifies persons with human organisms, yet (b) allows for the possibility that persons could exist in a disembodied state. Along the way, I will argue that
the view I defend is better than its rivals at grappling with at other sorts of cases, including cases
where the upper brain (or even whole brain) of a human organism has been completely
reprogrammed, removed, or destroyed.

The main argument of this essay, it will be recalled, has three steps:

1. If an entity is human, it has a set of typical human capacities.

2. If an entity has a set of typical human capacities, it has serious moral status.

Therefore,

3. If an entity is human, it has serious moral status.

To understand this argument, it is important to understand what I mean by “human”, “serious
moral status,” and “a set of typical human capacities.”

2. What Humans Are

The term “human” is sometimes used ostensively. When asked what beings in the world
are human, most of us would begin our answer simply by pointing and by saying, “why, those
things, over there, are human…this thing, right here, is human…I myself, for example, am
human…I’m a human being.” But ostension only takes us so far, because eventually different
people with different understandings about what property “human” refers to will point to the
same thing and disagree about whether it is human.

When we attempt to go beyond mere ostension, we quickly encounter a number of
different ways, or families of ways, for using the term “human”. The term “human” sometimes
refers to being an individual member of the species Homo sapiens, or being an individual
member of the species Homo sapiens with a certain sort of mental life, or being a rational
animal, or being made in God’s image. This term sometimes refers to the property of possessing
basic “human” rights, such as the right to life, liberty, and the pursuit of happiness. But it also sometimes refers to a property referred to in the writings of Immanuel Kant, of being an “end in itself” that deserves the respect of all rational agents. Furthermore, the term “human” sometimes refers to the properties we use as evidence for classifying an individual as a member of the species Homo sapiens. Thus “human” sometimes refers to a broadly phenotypic property, such as the property of being a featherless biped, or the property of having phenotypes “like those things, over there” (using ostension to designate the comparison class). It sometimes refers to a broadly genotypic property, such as the property of possessing a certain number of chromosomes, or the property of having a certain sort of DNA, or the property of having genotypes “like those things over there” (again, using ostension to designate the comparison class). It sometimes refers to a broadly genealogical property (provided that one of the other senses of “human” is already in place), such as the property of being an ancestor of human beings, or the property of being a descendant of human beings, or the property of being a relative of human beings, or the property of being able to produce viable offspring with human beings.

According to the main argument of this essay, an entity is “human” if it has possessed, at some time in its history, the same basic genotype as you and I possess right now. More precisely: an entity is human if it has possessed, at some time in its history, a structure made up out of one or more cells that have the same basic genotype as your cells and my cells have right now. Cells have the same basic genotype as yours and mine just in case they have DNA of the sort that you and I have even though chimpanzees do not. It is difficult to give a more precise characterization than this. But a creature with ZNA instead of DNA in its cells—where ZNA is a structure radically different than DNA, not having a double helix structure, not even having base pairs, and so on—would not have the same basic genotypes as us. Consider an organism whose
cells have ZNA instead of DNA in them, yet whose phenotypes are indistinguishable from the phenotypes of Hillary Clinton: the organism looks and behaves just like her, and even has experiences just like hers (this organism knows the answer to the question “What is it like to be Hillary Clinton?”). Hillary Clinton and this organism have different basic genotypes; she is human, this organism is not. But Hillary Clinton and Shaquille O’Neal have the same basic genotype; they are both human.

So, then, an entity is “human” if it has possessed, at some time in its history, a structure made up out of one or more cells that have the same basic genotype as your cells and my cells have right now. The phrase “at some time in its history” is designed to handle at least two sorts of thought-experiments. First, assume Hillary Clinton happens to be an immaterial soul that can exist disembodied after her biological death as a human organism. My definition of “human” would entail that she is still human during this disembodied phase of her existence. For even when she is disembodied, it is still true that she had possessed, at some time in her history, the relevant sort of basic genotype. This, I believe, is a desirable result, since it allows for Hillary Clinton to continue being human even after her biological death as a human organism. Second, assume Hillary Clinton happens to be an immaterial soul that can exist disembodied before her biological life as a human organism. My definition of “human” would entail that she is not human during this disembodied phase of her existence. For when she is disembodied, it is not true that she had possessed, at some time in her history, the relevant sort of cell structure. This too, I believe, is a desirable result, since it does not make Hillary Clinton human before her biological life as a human organism. Thus, although it is generally harmless to replace phrases like “human” and “human being” with “human organism” throughout this essay, this strategy
will run intro trouble in a case where a human (like Hillary Clinton) continues to exist disembodied after its biological death as a human organism.

3. What Serious Moral Status Is

The phrase “moral status” has at least five distinct, but closely related, meanings. First, talk about the moral status of something is often simply a shorthand way of referring to its morally salient features, or those of its features that are prominent from a moral point of view. For example, when we debate whether some action is morally required, or morally permissible, or morally impermissible, we are debating the moral status (in this first sense) of that action. Examples of this first sense of moral status are found in article titles like “The Moral Status of Abortion.”

Second, the phrase “moral status” is sometimes used to focus on the morally salient features of particular individuals in the world. For example, when we debate whether some entity has rights, or dignity, or intrinsic value, we are debating the moral status (in this second sense) of that entity. An example of this second sense of moral status is found in the following advertisement describing a recent philosophy conference:

The notion of an entity’s moral status is among the most fundamental and pervasive of our moral concepts. Current moral debates over stem cell research or therapeutic cloning, for example, revolve largely around questions concerning the moral status of embryos—the sort of intrinsic importance they may have (or lack),

---

making them worthy (or unworthy) of certain kinds of moral consideration—just as debates over abortion or animal rights have long dealt with questions about the moral status of fetuses or of non-human animals. And the idea of the moral status of persons is central to many accounts of our own fundamental rights and obligations to one another, placing it at the very heart of ethical thought.4

This second concept of moral status is narrower than the first one, since it can only apply to concrete individuals in the world, and cannot apply to things like actions or policies. However, the claim that some item has moral status, in either the first sense or the second sense, is not very informative because it does not tell us what sort of moral status that item has. For example, the claim that some action has moral status does not tell us whether that action is morally required, or morally permissible, or morally impermissible, and the claim that human embryos have moral status does not tell us exactly what “the sort of intrinsic importance” is that they have “making them worthy…of certain kinds of moral consideration.” The claim that something has moral status, in either this first or second sense, is rather like the claim that something has height: even if we know that something has height, we still do not know just how tall it is.

Third, there is Mary Anne Warren’s concept of moral status:

To have moral status is to be morally considerable, or to have moral standing. It is to be an entity towards which moral agents have, or can have, moral obligations. If an entity has moral status, then we may not treat it in just any way we please; we are morally obliged to give weight in our deliberations to its needs, interests, or well-being. Furthermore, we are morally obliged to do this not

merely because protecting it may benefit ourselves or other persons, but because
its needs have moral importance in their own right.\(^5\)

On Warren’s gloss of the concept of moral status, the claim that some entity has moral status is
very informative. For her, a thing has moral status just in case it has a cluster of other properties,
such as being an entity towards which moral agents have moral obligations and being an entity
whose needs have moral importance in their own right.

Fourth, Elizabeth Harman’s concept of moral status builds upon the idea of a harm
“mattering morally” as follows:

A harm to a being “matters morally” just in case there is a reason not to perform
any action that would cause the harm and the reason exists simply in virtue of its
being a harm to that thing, and simply in virtue of the badness of the harm for that
thing. A thing has moral status just in case harms to it matter morally…Some
examples will help to illustrate this terminology. There are reasons not to harm
both Alice and her car; but only Alice has moral status. Harms to Alice provide
reasons against action simply in virtue of being harms to her. But harms to
Alice’s car provide reasons against action only in virtue of being harms to Alice;
so these harms do not matter morally because the reason against action does not
exist simply in virtue of the harm’s being to that thing.\(^6\)

\(^5\) Mary Anne Warren, *Moral Status: Obligations to Persons and Other Living Things* (Oxford: Clarendon

Elizabeth Harman, “Moral Status” (Ph.D. diss., MIT, 2003), available online at
http://hdl.handle.net/1721.1/17645. As she puts it in the dissertation, “If something is ever harmed, then it has moral status just in case we have reasons
not to cause harms to it simply in virtue of the badness of the harms for it” (15). Harman recognizes that her notion of moral status “is somewhat close to the intuitive meaning of “moral status” even though it retains a somewhat
stipulative character that reflects her own particular methodological purposes. For example, she explains her choice
to include the phrase “if something is ever harmed” in her definition like this: “The antecedent “if something is ever
harmed” in the explanation is necessary because if something is not ever harmed, then it vacuously satisfies the
Fifth, there is the concept of moral status that appears in the thesis of this essay—what I shall call “serious moral status.” Serious moral status is a species of both the first and second sorts of moral status considered above. It includes everything that Warren and Harman’s concepts of moral status include, but it also includes much more as well. If something has serious moral status, then there is a strong moral presumption against harming it, a strong moral presumption against wronging it, and a strong moral presumption against even speaking ill of it, or “cursing” it in any way. If something has serious moral status, then it is owed respect, indeed owed justice, and there is a standing reason to benefit it whenever possible. Serious moral status is a place-holder for whatever morally salient features normal adult human persons possess.

One aspect of serious moral status sometimes goes by the name of “the right to life,” but it can be characterized without the language of rights as follows: if an entity has this aspect of serious moral status, then there is a strong moral presumption against killing that entity.

The concept of a strong moral presumption against killing is an intuitive concept that can be used across various normative theories. This concept is intuitive because it is designed to capture two widely shared pre-theoretical intuitions, intuitions that guide the thinking of most people and that also guide the thinking of most philosophers who attempt to construct adequate theories about the morality of killing: first, that killing other normal adult human persons is prima facie morally wrong; second, that this prima facie moral wrongness is strong enough to establish a presumption against killing other normal adult human persons. This concept is useful

---

7 An action is prima facie morally wrong just in case the action, on the face of it, is morally wrong. Admitting that actions of a certain type are prima facie morally wrong does not commit one to any further thesis about what the wrong-making properties of that type of action are.
to various normative theories because admitting that there is a strong moral presumption against killing an individual still leaves three important questions open. First, what is the precise strength of this presumption? Second, what conditions might override this presumption? Third, what provides the basis of this presumption? The main argument of this essay does nothing to answer the first two of these important questions, but it does attempt to answer the third. For the second premise of the main argument is tantamount to the claim that one basis of serious moral status, and hence the strong moral presumption against being killed, is the possession of a set of typical human capacities.

Before explaining what sorts of capacities are the relevant ones, it is important to explain why I focus on an aspect of moral status that is both intuitive and theoretically neutral. I focus on an aspect of moral status that is intuitive because I want to bring on board as wide an audience of philosophers and non-philosophers as possible. The intuition that it is, on the face of it, morally wrong to kill normal adult human persons, is a stable and widely shared intuition. As Fred Feldman writes,

“One of the most widely accepted and intuitively plausible moral principles is “Thou shalt not kill.” I take this to mean (or to imply) that it is morally wrong to kill people. It is hard to think of a moral principle with greater immediate credibility. Surely, if any moral principle is true, some version of this one is.”

The reason why I focus on an aspect of moral status that is theoretically neutral is that I do not want to prejudge certain important theoretical questions in moral philosophy. The concept of a strong moral presumption against killing, unlike the concepts of a right to life, or an absolute

---

prohibition against killing, or the sanctity of life, does not raise questions about the existence of rights, or the absoluteness of prohibitions, or the meaning of “sanctity”.

The connection between moral status and the morality of killing is widely recognized. For example, at the outset of his book *The Ethics of Killing: Problems at the Margins of Life*, Jeff McMahan claims that there are “four distinct categories into which we may sort most or all instances of killing for which there may be a reasonable justification,” and the third category, which constitutes the primary focus of his book, consists of “cases in which the metaphysical or moral status of the individual killed is uncertain or controversial.”9 For example, there is controversy about the claim that there is a strong moral presumption against killing undeveloped human organisms. This claim is controversial partly because many arguments against it entail that there is not even a strong moral presumption against killing certain classes of developed human organisms. For example, consider the following argument:

(1) An entity must possess a mental life of a certain sort of richness in order for there to be a strong moral presumption against killing it.

(2) Undeveloped human organisms do not possess this sort of mental life.

Therefore,

(3) There is not a strong moral presumption against killing undeveloped human organisms.

The key premise in this argument is (1), but this premise, when combined with other highly plausible premises, such as

---

9 Jeff McMahan, *The Ethics of Killing: Problems at the Margins of Life* (Oxford: Oxford University Press, 2002), vii-viii. The other three categories are (1) “cases in which killing would simply promote the greater good,” (2) “cases in which an individual has done something that has lowered the moral barriers to harming him, or compromised his status as inviolable, or made him liable to action that might result in his death,” and (4) “cases in which death would not be a harm to the individual but instead a benefit.”
Developed human organisms with certain forms of brain damage, dementia, or mental retardation do not possess this sort of mental life. Seems to entail that

There is not a strong moral presumption against killing developed human organisms with certain forms of brain damage, dementia, or mental retardation. McMahan and others are prepared to openly admit and discuss this sort of entailment. Consider the way he begins discussing his third category:

Among those beings whose nature arguably entails a moral status inferior to our own are animals, human embryos and fetuses, newborn infants, anencephalic infants, congenitally severely retarded human beings, human beings who have suffered severe brain damage or dementia, and human beings who have become irreversibly comatose.\(^{10}\)

This essay, then, is set within the general context of contemporary disagreement and debate over the ethics of killing, and within the narrower context of a certain development within this debate, a development in which philosophers seriously entertain the idea that there is not a strong moral presumption against killing certain classes of adult human organisms because of their diminished moral status. I argue, against this idea, that all human organisms have serious moral status, which includes a strong moral presumption against being killed, and that all human organisms have this serious moral status because they have a set of typical human capacities.

---

\(^{10}\) McMahan, vii.
4. What Typical Human Capacities Are

The concept of a set of typical human capacities is a technical concept that relies upon the simpler concepts of a capacity and a hierarchy of capacities.

A capacity is the metaphysical ground or truth-maker of a true conditional statement about a thing. For example, the capacities of sub-atomic particles are what make statements describing what would happen to these particles in different circumstances true. A capacity is the same thing as a power or disposition. There are different views about the nature of capacities. Some say they are ontologically primitive, and do not admit of further analysis. Others say they are ontologically derivative, because they logically supervene on ‘categorical’ properties plus the laws of nature. Which of these views is more plausible will not affect the truth of the main argument.

A hierarchy of capacities is a group of capacities with three interesting features. First, some members of the group are lower-order capacities. Second, some members of the group are higher-order capacities. Third, the higher-order capacities are just capacities to obtain the lower-order capacities. Examples from chemistry and biology illustrate this. Liquid water has the lower-order capacity to evaporate, while ice has the higher-order capacity to obtain this lower-order capacity. A mature oak tree has the lower-order capacity to support a tree house, while a sapling has the higher-order capacity to obtain this lower-order capacity.

When it comes to human organisms, ordinary language often presupposes these hierarchies even though it partially obscures them. For example, some would say that I do have the capacity to speak English, but I do not have the capacity to speak Chinese, because I can speak English right now if I needed to but it would take me a great deal of time and effort to learn to speak Chinese. We can capture what is going on here by saying that I have a lower-
order capacity to speak English, but I do not have a lower-order capacity to speak Chinese. I do, however, have a higher-order capacity to speak Chinese. More precisely, I have a higher-order capacity to obtain the lower-order capacity to speak Chinese.

This concept of a hierarchy of capacities helps explain the sense in which I still have a capacity even when I am temporarily “incapacitated” in some way. For example, when I am asleep, or under anesthesia, or comatose, or brain damaged, I do not have the lower-order capacity to think, but I do have the higher-order capacity to obtain this lower-order capacity.

In general, an individual has a higher-order capacity to do some activity whenever it has the ability to obtain the lower-order capacity to do that activity. In some cases, obtaining a lower-order capacity is relatively easy: for example, a sleeping or anaesthetized person eventually wakes up. In other cases, obtaining a lower-order capacity is more difficult: for example, learning Chinese takes time and effort on the part of the learner and on the part of teachers. In still other cases, obtaining a lower-order capacity requires extensive medical assistance: for example, recovering from a coma or brain damage requires the help of highly skilled professionals with sophisticated technology.

One of the most important claims of this essay is that, for an entity to have “a set of typical human capacities,” it is enough for that entity to have such capacities either at some lower-order level or at some higher-order level. “Typical human capacities,” in turn, are those capacities possessed at the lower-order level by any normal adult human person. I shall not seek to define which capacities are included in “a set of typical human capacities” any more precisely than this. So, for example, I shall not attempt to show that the set of typical human capacities includes the capacity to experience pleasure and pain, or the capacity for self-consciousness, or the capacity for rationality, or some combination of these capacities.
This essay is focused on showing that if an individual is human, it has a set of typical human capacities, and that if an individual has this set of typical human capacities, it has serious moral status, and hence there is a strong moral presumption against killing it.

Philosophers have often found it useful to make a distinction between active capacities and passive capacities. Examples of active capacities might include my capacities to raise my arm and to think about what I want to, whereas examples of passive capacities might include my capacities to see red and to experience pain. One of the distinctive aspects of this essay is that it argues that an entity’s passive higher-order capacities are just as relevant to that entity’s moral status as its active higher-order capacities. This is important to keep in mind throughout the essay, since most writers who attempt to forge a link between an entity’s higher-order capacities and its moral status hold that it is only active higher-order capacities that can be relevant to its moral status. I shall argue that allowing passive higher-order capacities to be relevant to an entity’s moral status solves a number of problems that are otherwise very difficult to solve. But, as we shall see, allowing passive higher-order capacities to be relevant in this way also generates many new problems that other accounts do not face.

Since the concept of a human being I am working with is designed to make room for the possibility that a human being might survive its own biological death as a human organism, it is important to explain how the concept of serious moral status relates to this possibility. For it seems very plausible both that (a) an account of moral status should allow us to retain our moral status if we become disembodied, and (b) an account of the strong moral presumption against killing cannot allow there to be a strong moral presumption against killing us when we are disembodied. After all, if we become disembodied, we are already dead. And it is not possible to kill something that is already dead.
One way of handling the tension between (a) and (b) is to replace the concept of a strong moral presumption against killing with the following definition of the sort of moral status in question, which allows for human beings to retain the same more status after death even if they can become disembodied:

For any X, X has moral status in the relevant sense if there is some type T, such that X is of type T, and there is a strong moral presumption against destroying anything of type T—where destroying something of type T includes, among other things, altering X so that it is no longer of type T.

On this definition, the concept of destroying takes the place of the concept of killing, and the relevant type T could be something like *Embodied Human Organism*: there is a strong moral presumption against altering me so that I am no longer an embodied human organism, since such alteration involves destroying an embodied human organism.

However, I am reluctant to adopt this definition, because it is not clear to me how “altering X so that it is no longer of type T” is really a legitimate instance of destroying something of type T. Take a teenager named Xavier. Now let Xavier turn 20. Xavier has been altered so that he is no longer a teenager—that is, he is no longer of the type *Teenager*. But I find it very odd to say that when he turns 20, a teenager has been destroyed. Or take a British politician named Xantippe, who happens to be a Tory. Now let Xantippe switch political parties. Xantippe has been altered so that she is no longer a Tory—that is, she is no longer of the type *Tory*. But again, I find it very odd to say that when she switches parties, a Tory has been destroyed. I am willing to grant that when Xavier turns 20, there is one less teenager in the world, and that when Xantippe switches parties, there is one less Tory in the world. And if that
is all one means by “destroy”, then the definition above might work. But in that case, I am reluctant to adopt the definition because it relies on an unconventional sense of “destroy”.

Perhaps the best way of avoiding the tension between (a) and (b) is to stipulate that the concept of serious moral status includes a strong moral presumption against performing an action that either makes a human disembodied or keeps a human disembodied. Then serious moral status would cover normal acts of killing, which count as performing an action that makes a human disembodied. But it would also cover any cases in which a disembodied soul is prevented from re-entering its original body. Although the latter sort of act is not technically killing, it seems very similar to killing. Let the term ‘skilling’ be synonymous with ‘performing an action that keeps a human disembodied.’ For a vivid example of skilling, imagine that you have a heart attack and become disembodied, floating above your body and viewing it from a somewhat remote distance, such as the ceiling of the room you are in when the heart attack happens. I rush onto the scene, but instead of performing cardio-pulmonary resuscitation (CPR) to make your disembodied soul return to your body, I instead use my ray-gun to instantly incinerate your body. I have not killed you—after all, you may still exist disembodied, floating near the ceiling of the room. But I have skilled you, since I have performed an action that kept you disembodied.

Once the concept of skilling is in place, I believe that the tension between (a) and (b) can be resolved by insisting that serious moral status should be understood, throughout the essay, as including a strong moral presumption against both killing and skilling. So, for example, the conclusion of the main argument should be understood to entail the following claim: “If an individual is human, there is a strong moral presumption against killing it when it can be killed, and a strong moral presumption against skilling it when it can be skilled.” If it turns out that
humans cannot survive the death of their biological organisms, then the second clause will be superfluous; but if it turns out that humans can survive the death of their biological organisms, then the second clause will be very informative.

Finally, we should note that the strong moral presumption against killing may or may not be applicable to non-biological entities, like advanced computers, that are functionally indistinguishable from biological entities. The reason is that it is only possible to kill something that is alive, and advanced computers are not alive. This might seem to be a significant disadvantage of my account, since on some views of the nature of the mind and mental states, advanced computers actually have a mental life much like the mental life of humans, and seem to experience emotions similar to the emotions humans experience when facing death. Remember the computer system from the movie *Short Circuit*, who pleaded with humans that they “No disassemble Johnny Five!”, or remember the young Indian girl, who is actually a young computer program, in the movie *Matrix 3: Revolutions*, who gets “murdered” by the evil Agent Smith. However, I am not especially concerned to address this problem, and for two reasons. On the one hand, I am not convinced by computational accounts of the mind, and hence remain agnostic about whether advanced computers would ever have the same capacities as a human being like Hillary Clinton. On the other hand, even assuming that advanced computers really did have minds, I am optimistic that a concept analogous to killing in the biological realm—perhaps “disassembly”—could be developed for such computers. The phrase “serious moral status” should be interpreted as including whatever concept analogous to killing applies to computers—if indeed there is such a concept. Likewise, if an advanced computer happened to have (or be) an immaterial soul—that is, if there really was a “ghost in the machine”—then a qualification about the difference between killing and skilling such a computer-soul would be in order.
5. What Persons Are

Typical human capacities, I claimed a moment ago, are those capacities possessed at the lower-order level by any normal adult human person. This immediately raises the question of what a person is. Since my answer to this question governs the way I approach the Adventure of Reagan’s Brain, and since it governs the way I frame several arguments below, it is important to be clear up front about what my approach involves. I believe that the concept of a person should be given a purely descriptive sense, and that this descriptive sense should be assigned to it, not based on moral considerations, but on considerations emerging from our considered judgment of when it is proper to apply the use of personal names like “Ronald Reagan” and the use of personal pronouns such as “me”. This section expands upon these claims.

The term “person” is used in a number of different ways both inside and outside philosophical discussions. One of the most important distinctions in the way “person” is used is what Joel Feinberg calls the distinction between normative (or moral) personhood on the one hand and descriptive (or commonsense) personhood on the other. Here is how Feinberg explains the normative use of “person”:

To be a person in the normative sense is to have rights, or rights and duties, or at least to be the sort of being who could have rights and duties without conceptual absurdity…when we attribute personhood in a purely normative way to any kind of being, we are attributing such moral qualities as rights or duties, but not (necessarily) any observable characteristics of any kind—for example, having flesh or blood, or belonging to a particular species.11

---

Examples of proposals for using “person” in this normative way include the following:

“X is a person” just means “X has the right to life”

“X is a person” just means “X ought to be treated as an end in itself”

On the other hand, here is how Feinberg explains the descriptive use of “person”:

There are certain characteristics that are fixed by a rather firm convention of our language such that the general term for any being who possesses them is “person.”…I shall call the idea defined by these characteristics “the commonsense concept of personhood.” When we use the word ‘person’ in this wholly descriptive way we are not attributing rights, duties, eligibility for rights and duties, or any other normative characteristics to the being so described. At most we are attributing characteristics that may be a ground for ascribing rights and duties.12

Examples of proposals for using “person” in this descriptive way include the following:

“X is a person” just means “X has reason”

“X is a person” just means “X has consciousness”

There is nothing incoherent about claiming that X is a person in some descriptive sense while denying that X is a person in some normative sense. But as Feinberg’s last sentence indicates, there are many examples of substantive proposals for the relationship between X’s descriptive personhood and X’s normative personhood.

I believe that our concept of a person should not be governed by moral considerations, but rather should be driven by our use of personal pronouns in situations where we try to imagine what can happen to us. This is because I can easily attempt to answer questions like “would that

individual be me?” while bracketing questions like “would that individual be a rights-holder, or an end-in-itself, or a possessor of such-and-such moral property?” Even an amoralist can entertain questions of the first sort. I therefore think of persons in purely descriptive terms.

It is instructive to compare my method for defining “person” with the method employed by Michael Tooley in *Abortion and Infanticide*. Tooley’s methodological starting point for defining ‘person’ is the desirability of obtaining a certain sort of result:

> It is…very important to have some term that refers only to…entities that have a right to life, and do so in virtue of their present properties, rather than in virtue of their potentialities. The term ‘person’ will here be understood in such a way that it applies to all and only entities of this sort.\(^{13}\)

His preferred method for defining “person” in order to achieve this result runs as follows:

> Instead of defining a person as an entity that possesses a right to life, and that does so in virtue of its present properties, rather than in virtue of its potentialities, one can first determine what properties, other than potentialities, suffice to endow an entity with a right to life. Then one can define the term ‘person’ as applying to all and only those things that have at least one of the relevant properties.\(^{14}\)

In attempting to justify this method for defining “person”, Tooley recognizes the objection that “it is right both that the term ‘person’, as ordinarily used, is generally a descriptive term, and that it is potentially misleading to define it instead in such a way that it is changed into an evaluative

---


\(^{14}\) Ibid., 35.
term.” Nevertheless, he believes the method is justified because “the assignment of descriptive content to the term ‘person’ is ordinarily guided by moral considerations.”

There are two implications of adopting Tooley’s method for defining “person”. First, when this approach is adopted, the term ‘person’ functions—as I believe it ordinarily does—as a purely descriptive term. The fact that something is a person does not by itself, therefore, imply any ethical conclusions. In particular, it does not follow from the fact that something is a person that it has a right to life.

The second implication, however, is this:

The fact that the descriptive content of the term ‘person’ has been selected with certain moral considerations in mind does mean, however, that if one’s moral views are in fact correct, then it will be the case that an entity is a person if and only if it has a right to life, and does so in virtue of its present properties, rather than in virtue of its potentialities.

My proposal for using “person” is different than Tooley’s, and for two reasons. First, I believe that his justification for using “person” the way he does is inconclusive. He recommends applying the term to all and only entities that have a right to life in virtue of their present properties. Thus for any entity we imagine in our philosophical thought-experiments, the way to know whether that entity is a person or not is to begin by asking the question: “does this entity have a right to life in virtue of its present properties?” If the answer to this question is “yes”, then the entity is a person, and we can go on to investigate what properties the entity has as a

---

15 Ibid.

16 Ibid., 35-36.

17 Ibid., 36.
way of determining what property or properties are necessary and sufficient for being a person. If the answer to this question is “no”, then the entity is not a person, and we can go on to investigate what properties the entity has as a way of determining what property or properties are not sufficient for being a person. He thinks this does justice to an alleged fact about our linguistic conventions, namely, that our assignment of descriptive content to “person” is ordinarily guided by moral considerations. But I believe that we can do more justice to our linguistic conventions by saying, with Feinberg, that the assignment of descriptive content to the term ‘person’ is sometimes guided by wholly moral considerations and sometimes guided by wholly non-moral considerations. Once this is done, we can then be explicit about which sort of considerations are going to guide our assignment of descriptive content to the term.

The second reason why my approach for defining “person” is different than Tooley’s emerges from a point I made a few paragraphs ago. When determining whether to use personal pronouns like “I” and “me”, I need not, and often am not, guided by moral considerations. Even if I was a purely selfish individual, caring nothing at all for rights, duties, or anyone but myself, I could still entertain and answer questions like “is this decision going to bring me the things that I want in ten years?” and “would that individual, ten years from now, really be me?” while bracketing questions like “would that individual, ten years from now, be a dirty rotten scoundrel?” and “would that individual, ten years from now, have a right to life?”

6. What We Are

Does an appeal to the use of personal pronouns automatically resolve philosophical debates about the nature of persons and personal identity? It would appear that the answer to this question is No. For ordinary language is sometimes ambiguous about what “we” are, and people
can disagree about how to best resolve this ambiguity. Imagine philosopher-kings one day rule the world, and set up a division of their executive branch whose job it is to enforce the “proper” use of personal pronouns such as “I” and “me”. The Personal Pronoun Police (PPP) get to decide on which of the following statements from ordinary language to ban or regulate:

(1) When I was born, I was a really ugly baby.
(2) After I die, I shall be buried in a cheap coffin.
(3) After I die, I shall experience wonderful things.
(4) If I lose my mind because of this brain disease, then please treat me tenderly.

One year, advocates of a certain sort of psychological account of personal identity run the PPP. They crack down on (1) and (4), replacing them with what they take to be more philosophically consistent paraphrases:

(1p) When the human organism with which I am currently associated was born, it was a really ugly baby.
(4p) If I lose my mind because of this brain disease, then please treat the resulting human organism tenderly.

Another year, substance dualists run the PPP, and so crack down on (2), replacing it with

(2s) After I die, my body shall be buried in a cheap coffin.

Yet another year, materialists run the PPP, and replace (3) with

(3m) After I die, nothing is left of me to experience wonderful things.

The imaginary case of the Personal Pronoun Police is designed to emphasize two points. First, as (1)-(4) illustrate, the use of personal pronouns in ordinary language is frequently ambiguous between different philosophical accounts of personal identity, and hence there may be no straightforward answer to the question “what account of personal identity does ordinary
language favor?” Second, no matter who is running the PPP from one year to the next, and hence no matter how ordinary language gets revised for the sake of consistency, there is a difference between consistency and accuracy. For suppose substance dualism is true, but materialists are successfully running the PPP. Everybody on earth can use the word “I” in such a way that (2) and (3m) are true. But this would leave untouched the truth that, after death, there is *something* which continues to think, remember, experience pleasure and pain, and so on.

How is one to proceed, given this disagreement about persons, personal pronouns, and personal identity? My proposed strategy for overcoming this disagreement is made clear by a pair of examples. First, imagine a teenager named Adam who believes that

(A) A person cannot survive the transition from age 19 to age 20.

You can imagine that this belief was formed any way you like. Perhaps Adam spontaneously formed this belief in a fit of misinterpretation and credulity on his 15th birthday when he overheard a trusted middle-aged neighbor make the offhand remark that “Well, a person can really live as a teenager, but once you turn twenty, well, then your life is pretty much over.” However his belief in (A) was formed, Adam now has it, and has it quite firmly. It is a central strand in his web of beliefs. He is thoroughly convinced of its truth. Since Adam also believes that he is a person, he easily draws the conclusion that he cannot survive the transition from age 19 to age 20.

Adam’s friends try to persuade him to give up this belief; but nothing seems to work. They tell him that nobody else shares the belief; Adam replies that nobody else knows the sober truth of the matter, and that he must courageously bear the burden of that truth alone. Friends of his who are themselves past the age of twenty assure Adam that they experienced no annihilation on their twentieth birthday; Adam replies that they are deluded, that the persons they seem to
remember being are different persons than the persons they now are, and that the persons that they seem to remember being no longer exist. They ask Adam to imagine how much fun he himself will have celebrating his 21st birthday, and how much suffering he will experience the morning after the celebration if he drinks too much alcohol; Adam replies that he “cannot look forward to the pleasures of, and cannot dread the pains of, that remote individual, since that remote individual will not be me.” Before midnight on his last day as a 19-year-old, Adam bids farewell to his friends, asking them to remember him fondly. After the clock strikes midnight, his friends cheer for him and jokingly ask him how it feels to have survived the transition. He makes the following reply: “Dear friends, I ask you, do not call me by the name of that good person; my name is Cain, and though I am thankful for what I inherited from that person you once called Adam, yet I am not he; the misfortune of his end was the good fortune of my beginning.”

What can possibly be said to Adam that would dislodge his belief in (A)? “Snap out of it”? Adam’s example is relevant because most of the positions taken by philosophers about personal identity through time are like Adam’s in the following way: even if you disagree with the position, it is very difficult to argue people out of the position if they are already convinced of it. This is because people who are already convinced of the position will have available answers for every attempt to persuade them out of the position, just like Adam had an available answer for his friends.

Next, imagine a second example, of a teenager named Ben who believes that:

(B) A person cannot survive the period from age 20 to 30.

You can imagine that this belief was formed any way you like. Perhaps Ben spontaneously formed this belief in a fit of misinterpretation and credulity on his 15th birthday when he
overheard a trusted middle-aged neighbor make the offhand remark that “Well, a person can really live as a teenager, and they can really live once they get past their twenties, but during those twenty-something years, you basically don’t have a life.” However his belief in (B) was formed, Ben now has it, and has it quite firmly. It is a central strand in his web of beliefs. He is thoroughly convinced of its truth. Since Ben also believes that he is a person, he easily draws the conclusion that he cannot survive the period from age 20 to age 30.

Ben’s friends try to persuade him to give up this belief, using all of the techniques that Adam’s friends used when trying to persuade Adam; but nothing seems to work. However, there is at least one strategy available for persuading Ben that is not available for persuading Adam. Whether it is persuasive to Ben depends on what other beliefs Ben has. The strategy runs like this:

(1) Ben exists at nineteen.

(2) Ben cannot have temporal gaps.

(3) Ben exists at thirty-one.

Therefore,

(4) Ben cannot cease to exist at twenty.

If Ben accepts (1) through (3), he must accept (4). And if he continues to believe that he is a person, he will realize that (4) is inconsistent with (B). Therefore he must give up (B).

When Ben is confronted with this strategy, it is always an option for him to revise his other beliefs to keep (B) intact. If a person is committed to a certain view about personal identity through time, they always have the option of scrapping other beliefs that threaten that view. One option Ben has of replying to his friends is this:

(1) Ben exists at nineteen.
(2) Ben cannot have temporal gaps.

(3*) Ben ceases to exist at twenty.

Therefore,

(4*) Ben cannot exist at thirty-one.

Alternatively, Ben could give up his beliefs about objects not having temporal gaps:

(1) Ben exists at nineteen.

(3*) Ben ceases to exist at twenty.

(3) Ben exists at thirty-one.

Therefore,

(2*) Ben can have temporal gaps.

I offer arguments throughout this essay that rely upon the sort of strategy Ben’s friends use. So this essay is directed to those who believe that an entity cannot have temporal gaps. This essay is also directed to those who believe certain things about when an entity exists at two distinct times. I am at present rather unsure of the best way to argue for these beliefs. But I think that the claim that an entity cannot have temporal gaps is plausible, and that my intuitions about when an entity exists at two distinct times are more plausible than the intuitions my rivals have about this.

Recall the example that began this chapter, The Adventure of Reagan’s Brain. To the question of “what happens to Reagan in this adventure?”, I answered, firstly, that Reagan survives the adventure, and secondly, that Reagan maintains his moral status throughout the adventure. I am not suggesting that those who deny the first part of my answer are inconsistent. I am just suggesting that their views are incorrect. The strategy for showing the incorrectness parallels the strategy of Ben’s friends:
(1) Reagan exists before the Alzheimer’s does its damage.

(2) Reagan cannot have temporal gaps.

(3) Reagan exists after the injection-plus-scan-in.

Therefore,

(4) Reagan cannot cease to exist when the Alzheimer’s does its damage.

Just as Ben could escape the conclusion of the argument above in various ways, I recognize that opponents of the argument in Reagan’s case can escape its conclusion in various ways. But I am confident that many people, both philosophers and non-philosophers alike, will be attracted to claims (1), (2), and (3) in the argument concerning Reagan, just as they are attracted to claims (1), (2), and (3) in the argument concerning Ben. Likewise, I am confident that many people, whether philosophically trained or not, will be sympathetic to the view that Reagan retains his moral status during the intermediate phase of his life, after the Alzheimer’s has done its damage, yet before the injection-plus scan-in. The remaining chapters in this dissertation are an attempt to build upon these ideas, to consider what these ideas imply about human organisms like Reagan and ourselves, and to defend these ideas against rival approaches.
CHAPTER TWO: HUMANS AND THEIR CAPACITIES

This chapter begins the defense of the idea that if an entity is human, it has a set of typical human capacities. The chapter is organized into three main sections. Sections one and two argue that there are several important distinctions among capacities, one of which revolves around the fact that some capacities come in natural hierarchies of lower-order capacities and higher-order capacities. Section three argues that attending to these hierarchies allows us to argue that all humans have a certain set of higher-order capacities in common.

There are three main problem areas for the claim that if an entity is human, it has a set of typical human capacities: the obvious diversity of capacities among normal humans, the nebulous sense in which undeveloped humans have capacities, and the apparent absence of certain capacities among abnormal humans. Although all three of these problem areas are discussed in this chapter, a fuller discussion of the last two problems is reserved for Chapters Four and Five.

1. Humans Have Capacities

On first glance, there does not appear to be anything philosophically controversial about the idea that humans have capacities. For example, I am a human organism with the capacity to walk, talk, jump, eat, play the trumpet, think, feel pain, remember events in my life, and do a thousand other things. There are some capacities that I have in common with almost all human organisms: for example, the capacity to walk. There are other capacities that I have and most other human organisms do not: for example, the capacity to play the trumpet. And there are
some capacities that I lack but other human organisms have: for example, the capacity to play the oboe. These sorts of claims seem to be completely uncontroversial.

There are at least two reasons why it is so natural to talk and think of ourselves in terms of capacities. First, capacities help explain how it is possible for a human organism to persist through certain sorts of changes. For example, how is it possible for me to persist through the change that occurs when I began thinking about philosophy this afternoon? How is it possible for me to not think about philosophy in the morning, think about philosophy in the afternoon, and still persist through this change? Why is it not the case that one individual was not thinking about philosophy in the morning, a completely different individual was thinking about philosophy in the afternoon, and this completely different individual replaced the first individual? Relying on capacities provides a neat way of answering these questions: in the morning, I had the capacity to think about philosophy. In the afternoon, I actualized this capacity. Capacities are like a kind of metaphysical glue that holds me together through changes that I undergo through time.

Second, capacities help explain why it is that certain sorts of changes happen. For example, why is it that I began thinking about philosophy in the afternoon, even though I was not thinking about philosophy in the morning? Why is it not utterly mysterious why this change of state happened to me? Why is the event of me thinking about philosophy not completely inexplicable, or random, or uncaused, or caused by something far away in another galaxy? Invoking my capacity to think about philosophy allows us to answer these questions: I thought about philosophy because (a) I had the capacity to think about philosophy if certain other conditions were met, and (b) those other conditions were met. Far from being a random event, my thinking about philosophy took place because my capacities were triggered or actualized in
some way. Capacities are like a kind of causal glue that helps explain why certain changes happen to me.

1.1. Ordinary Language Speaks about Capacities in Various Ways

Both ordinary language and the language of philosophy and science are suffused with statements about capacities. For example, you have the capacity to speak English, my pet rat has the capacity to experience pain, and salt has the capacity to dissolve in water. Such capacity-statements typically have a structure containing three elements: one which refers to the capacity itself, another which refers to the bearer of the capacity, and a third element which refers to what the capacity is a capacity for. Since the meaning of a given capacity statement is a function of these three elements, a word about each of them is in order.

First, there are various ways of characterizing the capacity itself. Modern discussions of capacities are sometimes carried out in terms of powers or dispositions. I shall understand the term ‘capacity’ to be synonymous with such terms, as well as with terms like ‘capability’, ‘ability’, ‘potential’, and ‘potentiality’. For example, your capacity to speak English is the same thing as your ability or power or potential to speak English. Although philosophers have sometimes used these various terms to make important distinctions among capacities, their ways of making such distinctions do not correspond to one another. I shall therefore use the previous

---


terms synonymously, and shall mark distinctions among capacities by supplying further qualifying words like ‘active’ or ‘passive’, ‘first-order’ or ‘second-order’, and so on.

Second, ordinary language recognizes various ways of characterizing the bearers of capacities. The paradigm capacities are borne by individuals: for example, your capacity to think is something you have as an individual. However, not all capacities are borne by individuals: for example, the capacity of salt to dissolve in water is something the salt has as a roughly defined aggregate of stuff.

Third, there are different ways of referring to what a bearer of a capacity has a capacity for, or a capacity to do, or a capacity to be. For example, dispositional terms get paired up with terms referring to actions (e.g., ‘thinking’ or ‘breaking’), properties (e.g. ‘hot’ or ‘hungry’), and individuals (e.g. ‘this human being’ or ‘this statue’). Dispositional terms also sometimes get paired up with phase sortals and substance sortals, which are count-nouns that serve as classificatory concepts for describing the world. As David Wiggins explains, the difference between phase sortals and substance sortals is the difference “between sortal concepts which present-tensedly apply to an individual \( x \) at every moment throughout \( x \)’s existence, e.g. human being, and those which do not, e.g. boy, or cabinet minister.”

Thus even when the bearer of dispositions is some individual \( x \), and even if the dispositional term is limited to ‘capacity’, many different sorts of capacity statements can be generated:

---

<table>
<thead>
<tr>
<th>Description</th>
<th>Representation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Action A (e.g., take in sunlight, think)</td>
<td>x has the capacity to A</td>
</tr>
<tr>
<td>Property F (e.g., five feet tall, hot)</td>
<td>x has the capacity to be F</td>
</tr>
<tr>
<td>Individual y (e.g., this statue, this human being)</td>
<td>x has the capacity to be y</td>
</tr>
<tr>
<td>Things picked out by phase-sortal P (e.g., hot thing)</td>
<td>x has the capacity to be a P</td>
</tr>
<tr>
<td>Things picked out by substance-sortal S (e.g., human being)</td>
<td>x has the capacity to be an S</td>
</tr>
</tbody>
</table>

### 1.2. Capacities May Be Active or Passive, Identity-preserving or Compositional, General or Specific

Two important distinctions among capacities are the distinction between active and passive capacities and the distinction between what might be called identity-preserving and compositional capacities. Both of these distinctions can be traced back to Aristotle’s discussion of potency. Aristotle wrote that all genuine potencies “are originative sources of some kind, and are called potencies in reference to one primary kind of potency, which is an originative source of change in another thing or in the thing itself qua other.” I shall label the potencies defined in terms of originative sources of change identity-preserving potencies because their actualization preserves the identity of their bearer.

Aristotle thought that the primary identity-preserving potencies are ‘active’ because their locus was “in the agent, e.g. heat and the art of building are present, one in that which can produce heat and the other in the man who can build.” Corresponding to these active potencies

---


were ‘passive’ ones, so-called because their locus was “in the thing acted on…[e.g.] that which is oily can be burnt, and that which yields in a particular way can be crushed.”  Such potencies “of being acted on” were still genuine because they are “the originative source, in the very thing acted on, of its being passively changed by another thing or by itself qua other”.  

Aristotle was also interested in the kind of potency some sort of matter has when considered in juxtaposition with something created out of the matter.  Such compositional potencies do not imply strict numerical identity between whatever is the actual and whatever is the potential.  Rather, they imply the ‘is’ of composition or constitution.  Just as there is a difference between the ‘is’ of identity and the ‘is’ of constitution, there is a corresponding difference between what might be called the ‘is potentially’ of identity and the ‘is potentially’ of constitution.  

The claim that humans have a set of typical human capacities should be understood in terms of identity-preserving potencies that include both active potencies (such as your capacity to feel pain) and passive potencies (such as your capacity to walk).  The material bits and particles that enter your lungs and that eventually constitute your brain tissues and legs do not have the capacity to feel pain and walk.  But you do have these capacities.  The fact that typical human capacities include “passive” potencies must be constantly kept in mind as the argument of this chapter unfolds, since many writers who emphasize the capacities of human organisms focus only upon the more “active” potencies.

---

6 *Metaphysics* 1046a22-26.

7 *Metaphysics* 1046a12-14.

8 *Metaphysics* 1048a37-b4.
There is a distinction among identity-preserving capacities that deserves to be noticed, which I shall call the distinction between general and specific (non-general, relational) capacities. Consider an example. You and I both have the capacity to think. The capacity to think is a general capacity. But your capacity to think your thoughts is a capacity specific to you, while my capacity to think my thoughts is a capacity specific to me. This would be true even if, contrary to fact, you and I had thought about the same sorts of things for every instant of our waking lives. Nobody else in the universe has the capacity to think my specific thoughts. Consider now a second example. You and I both have the capacity to remember our experiences. The capacity to remember is a general capacity. But your capacity to remember your experiences is a capacity specific to you, while my capacity to remember my experiences is a capacity specific to me. This would be true even if, contrary to fact, you and I had experienced the same sorts of things for every instant of our waking lives. Nobody else in the universe has the capacity to remember your specific experiences.

The claim that humans have a set of typical human capacities in common should be understood in terms of general capacities. Clearly, I do not have your specific capacities, and you do not have mine. Both specific capacities and general capacities are helpful for understanding personal identity through time, in the sense that they both help explain how I continue to exist when I begin thinking about philosophy. But neither the general capacities nor the specific capacities help explain why I am not identical to you. The general capacity to think cannot explain this, since we have that in common. The specific capacities we both have, to think only our thoughts and not each other’s thoughts, cannot explain this, since the fact that these capacities are specific already presupposes that I am not identical to you.
2. These Capacities Come in Hierarchies

A hierarchy of capacities is a group of capacities with three interesting features. First, some members of the group are lower-order capacities. Second, some members of the group are higher-order capacities. Third, the higher-order capacities are just capacities to obtain the lower-order capacities. This section explains why it is reasonable to view capacities as falling into such hierarchies, and argues that such hierarchies can be clarified by the concept of developmental distance.

2.1. Several Philosophers Have Recognized That Capacities Come in Hierarchies

A number of contemporary philosophers have noticed an ambiguity in our talk about capacities and have sought to overcome it by employing certain hierarchical distinctions. For example, Michael Tooley has at least two discussions in which he notices such an ambiguity and attempts to eliminate it by introducing a distinction. In one discussion, he notes that there is “a certain imprecision in everyday talk about capacities” and suggests that this imprecision can be resolved by a distinction between capacities “in a narrower and stricter sense” and capacities “in a broader sense”:

Suppose that someone asks whether Mary is capable of running a six-minute mile. One person might say: ‘Certainly, she’s capable of running a six-minute mile, but she would have to get back into serious training for a few weeks.’ What he is saying might equally well be put: ‘No, she’s not capable of running a six-minute mile at present, but she would be if she went back into serious training for a few weeks.’ The moral is that when we speak about what someone is capable of, we may, as the first sentence illustrates, be talking, not about what they are presently
capable of, but about what they could become capable of in a reasonably short period of time...Ordinary talk about capacities and capabilities is therefore somewhat imprecise. One must distinguish between capacities in a narrower and stricter sense, and capacities in a broader sense that includes potentialities for acquiring capacities in the narrower sense.  

In another discussion, however, Tooley makes a three-fold distinction between immediately exercisable capacities, blocked or suppressed capacities, and potentialities. The distinction between “immediately exercisable capacities, and blocked or suppressed capacities” is illustrated by his example of Mary, only this time her inability to run is due to alcohol:

To attribute an immediately exercisable capacity to something is to make a statement about how the thing would be behaving, or what properties it would have, if it were now to be in certain circumstances, or in a certain condition...Thus, to say that Mary is now capable of running a five-minute mile is to say that if Mary were now to try to run a five-minute mile, were appropriately dressed, were on a track where there is not too much wind, and so on, then she would succeed in running a mile in five minutes. This is what I mean by an immediately exercisable capacity...Suppose, however, that Mary is drunk. In one sense, she is no longer capable of running a five-minute mile. Yet one might still attribute that capacity to her. In such a case, one might speak of blocked capacities—the idea being that all of the ‘positive’ factors required for the

---

9 Tooley, 106-107.
immediately exercisable capacity are present, but there are also negative factors that prevent the exercise of the capacity.\textsuperscript{10}

Tooley then claims that “It is very important not to confuse the concept of a capacity—whether in the narrow sense that covers only immediately exercisable capacities, or in the broader sense that also includes blocked capacities—with the concept of a potentiality”:

To attribute a certain potentiality to an entity is to say at least that there is a change it could undergo, involving more than the mere elimination of factors blocking the exercise of a capacity, that would result in its having the property it now potentially has. It may also be to say that there are now factors within the entity itself that will, if not interfered with, cause it to undergo the relevant change.\textsuperscript{11}

Eric Olson also notices an ambiguity in our talk about capacities, but he attempts to eliminate it by introducing a distinction between “first-order” capacities and “second-order” capacities:

There are two different senses in which something might have the capacity to do something. In one sense, someone has the capacity to swim (for example) if she has learned how to swim and is not paralyzed, unconscious, or otherwise handicapped. If you put her in the water, she can swim. We might call this a “first-order capacity”. But there is also a sense in which even someone who has

\textsuperscript{10} Ibid., 149-150.

\textsuperscript{11} Ibid. These two discussions of Tooley’s are slightly confusing when put side by side, because he uses the labels “narrow” and “broad” differently in each discussion. In the first discussion, the “narrow” sense of a capacity is contrasted with a “broad” sense that includes potentialities, whereas in the second discussion, the “narrow” sense of a capacity only includes “immediate” capacities, and is contrasted with a “broad” sense that does not include potentialities, but only “blocked or suppressed” capacities.
not learned how to swim has the capacity to swim if she could learn how to do it. In this sense all human beings who aren’t somehow physically disabled have the capacity to swim. Butterflies, on the other hand, do not have the capacity to swim. They simply aren’t built for swimming. Someone has a “second-order capacity” to swim if she could acquire a first-order capacity to swim.  

If we accept Olson’s suggestion, then what Tooley calls “immediate capacities” may be re-named “first-order capacities”, what Tooley calls “blocked or suppressed capacities” may be named “second-order capacities”, and what Tooley calls “potentialities” may also be named “second-order capacities”. On Olson’s picture, since both Mary-when-out-of-shape and Mary-when-drunk could acquire a first-order capacity to run a speedy mile, both Mary-when-out-of-shape and Mary-when-drunk would have a “second-order” capacity to run a speedy mile.

I believe that something like the sort of numerical approach Olson introduces is a helpful supplement to Tooley’s approach, because it emphasizes the fact that Mary’s “potentiality” to run a speedy mile is really just a certain sort of capacity to run a speedy mile, and that this capacity is different only in degree, and not different in kind, from Mary’s capacity to run a speedy mile when she is drunk (or, for that matter, from Mary’s capacity to run a speedy mile when she is sober). However, I believe that Olson’s approach forces us to ignore some relevant differences between Mary-when-drunk and Mary-when-out-of-shape, since it simply labels both of them a “second-order” capacity. Part of what is needed, it seems, is a way of allowing for there to be more than two “orders” of capacities.

To get such a numerical approach that allows for more than two orders of capacities, we may look to the philosopher C. D. Broad, who sought to elucidate what he calls the “popular-

---

scientific” view that many of us hold by employing, among other things, the concept of a hierarchy of dispositions. The “popular-scientific” view itself is this:

We ascribe to a thing a certain inner nature, and we hold that its history is determined jointly by its inner nature and its external circumstances…Thus a Thing is conceived as a store of powers or dispositions.\(^\text{13}\)

Broad employed a number of distinctions among dispositions in order to clarify this view. He claims that things in nature have *orders* of dispositions:

A bit of iron which has been put inside a helix in which an electric current circulates acquires the power to attract iron-filings. A bit of copper, placed in similar circumstances, does not. Under certain other circumstances, e.g., if it be sharply hit or heated to a certain temperature, the bit of iron will lose the magnetic property. If we call the magnetic property a “first-order disposition”, the power to acquire this property when placed in a helix round which a current is circulating may be called a “second-order disposition” specific to iron. For it is a disposition to acquire the first-order disposition under certain circumstances, and it is common and peculiar to bits of iron. Similarly, the power to lose the magnetic property when heated or sharply hit will be a second-order disposition of iron.\(^\text{14}\)

This leads to his general definition of a higher-order disposition:

---

\(^{13}\) C. D. Broad, *Examination of McTaggart’s Philosophy* (Cambridge: Cambridge University Press, 1933), 1: 265.

\(^{14}\) Ibid., 266.
A disposition of the second order is, in general, a disposition to acquire or to lose, under assigned conditions, a disposition of the first order. In the same way we could define dispositions of the third or higher order.\textsuperscript{15}

Finally, Broad believes his account applies to minds as well as matter:

The power of learning to talk is a mental disposition of at least the second order, for it is a power to acquire a power of doing something. One of the peculiarities of minds in general, and of human minds in particular, is that they start with very few first-order powers, but rather with powers to acquire powers.\textsuperscript{16}

Before applying Broad’s distinctions to the ambiguity noticed by Tooley, it is important to pause and notice a question that Broad’s example of the bit of iron leaves open. Let A be the first-order disposition of being magnetic, B be the second-order disposition of acquiring this first-order disposition, and C be the second-order disposition of losing this first-order disposition. Let t be a time before the bit of iron has been placed in the solenoid, t\textsuperscript{*} be a time after the bit of iron has been placed in the solenoid, but before it has been heated, and t\textsuperscript{**} be a time after the bit of iron has been placed in the solenoid, and after it has been heated. The picture Broad seems to be committing himself to is, at the very least, this: at t, the bit of iron has B but not A; at t\textsuperscript{*}, the bit of iron has A and C; at t\textsuperscript{**}, the bit of iron again has B but not A. This is represented in the following chart, with question marks representing the questions that Broad’s example seems to leave open:

\textsuperscript{15} Ibid., 266-7.

\textsuperscript{16} Ibid., 267.
<table>
<thead>
<tr>
<th></th>
<th>t</th>
<th>t*</th>
<th>t**</th>
</tr>
</thead>
<tbody>
<tr>
<td>A: the first-order disposition of being magnetic</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>B: the second-order disposition of acquiring A</td>
<td>Yes</td>
<td>?</td>
<td>Yes</td>
</tr>
<tr>
<td>C: the second-order disposition of losing A</td>
<td>?</td>
<td>Yes</td>
<td>?</td>
</tr>
</tbody>
</table>

A more robust picture would put “Yes” where the question marks are, whereas a more austere picture would put “No” in these places. I think that Broad’s own position eventually favors the more robust picture, but I need not show that here. What is important for the present discussion is that I favor the more robust picture, and will consequently formulate arguments below concerning humans assuming that the more robust picture is true. If the more austere principle turns out to be true, the arguments below could be reformulated without loss to accommodate it.

The main reason I favor the more robust picture is that I think the more austere picture relies upon the following principle, which I think is incorrect:

(1) Whenever an entity X has a disposition D at t to acquire (lose) a disposition D* at t*, then once X acquires (loses) D* at t*, X no longer has D at t*.

I think (1) is incorrect because it is simply a version of the following principle:

(2) Whenever an entity X has a disposition D at t to acquire (lose) a property P at t*,

then once X acquires (loses) P at t*, X no longer has D at t*.

(2) seems false. In general, whenever a disposition is manifested, the disposition still continues to exist. For example, imagine David Hume (the philosopher) and David Robinson (the professional basketball player) are arguing over whether Robinson has the capacity, or disposition, to slam-dunk a basketball. “Show me the money,” says Hume. “Prove you can
dunk.” Robinson then jumps up and slam-dunks a basketball. It seems that Robinson has the capacity, or disposition, to slam-dunk even when he is in the very act of slam-dunking. Otherwise, it would not be possible to prove that he has the disposition by manifesting it. But it is. Robinson had the disposition, at t, to slam-dunk at t*. When he actually slam-dunked at t*, he still had the disposition, at t*, to slam-dunk. Indeed, at t*, Robinson is showing off that very disposition. Hence principle (2) is false. But this basic line of argument can be recast even when we move to principle (1) and higher-order dispositions. For example, imagine Hume and Robinson are arguing over whether Robinson has the disposition to slam-dunk, but Robinson is currently recovering from an ankle injury. “I cannot dunk it right now, because I’m injured,” says Robinson, “but give me six weeks, and I will be capable of dunking it.” Robinson has a higher-order disposition, at t, to acquire the lower-order disposition six weeks later at t* to slam-dunk. When Robinson actually acquires the lower-order disposition at t* to slam-dunk, he still has the higher-order disposition, at t*, to acquire the lower-order disposition to slam-dunk. Indeed, at t*, the fact that Robinson has the lower-order disposition proves that he has the higher-order disposition.

In any event, with Broad’s suggestion in hand, we might now be able to say that what Tooley calls “blocked or suppressed” capacities, and what he calls “potentialities,” are really just different sorts of “higher-order” capacities, and that one of the differences between them (although certainly not the only one) is a difference in just how high their order is.

George Molnar has argued for a numerically characterized hierarchy of capacities in a way similar to Broad. Molnar argues that one of the senses “in which we need a hierarchical taxonomy of properties” is to be found in the way dispositional properties are arranged:
We can talk of a behavioral disposition of an object, for example *being magnetized*, as a first-order power. We can also talk of an object’s capacity to acquire a first-order power, for example *being magnetizable*, as a second-order power. We can also call an object’s aptness to lose a first-order power, for example metal’s *tendency to fatigue*, a second-order power…For the sake of clarity, from now on I will refer to…powers to acquire (or to lose) a power as ‘iterated powers’.  

Let’s sum up this discussion of a hierarchy of capacities by recalling the example where a mature oak tree has the lower-order capacity to support a tree house. It seems that if a sapling has a higher-order capacity to support a tree house, then so does an acorn. But the acorn’s higher-order capacity to support a tree house seems to be higher in order than the sapling’s. A more informative way of describing the difference between the acorn’s and the sapling’s capacity to support a tree house involves assigning numbers to the orders of capacities before explicitly relating these orders to one another. Let us call the mature oak tree’s capacity to support a tree house a first-order capacity. Let us call the sapling’s capacity to support a tree house a second-order capacity—recognizing, as we must, that numbering it ‘second’ is somewhat arbitrary since there are any number of stages between the sapling and the mature oak. Finally, let us call the acorn’s capacity to support a tree house a third-order capacity—again, recognizing that numbering it ‘third’ is somewhat arbitrary since there are any number of stages between the acorn and the sapling. (Although the exact numbers we assign to the orders are somewhat arbitrary, what is not arbitrary is that we assign a higher number to the sapling’s capacity to support a tree house than we assign to the mature oak tree’s capacity to support a tree.

---

17 Molnar, 32-33.
house, and that we assign a higher number to the acorn’s capacity to support a tree house than we assign to the sapling’s capacity to support a tree house.) We can now characterize the difference between the sapling and the acorn as follows:

(*) While the sapling’s second-order capacity is a capacity to obtain the mature oak tree’s first-order capacity, the acorn’s third-order capacity is a capacity to obtain the sapling’s second-order capacity.

Given this understanding of a hierarchy of capacities, where every additional order of the hierarchy makes reference to the orders beneath it, a “first-order” or “lowest-order” or “immediate” capacity is simply the capacity where a given hierarchy starts.

2.2. Hierarchies of Capacities Can Be Clarified by the Concept of Developmental Distance

The concept of a hierarchy of capacities can be clarified by the concept of the developmental distance an individual must cross before it has a certain immediate capacity. Such developmental distance can be characterized either in terms of a certain number of temporal units or in terms of a certain number of developmental steps.

Consider the following two examples. First, assume that I acquired the immediate capacity to think on my 3rd birthday. Assume also that I had to form 10 new “nested” neural pathways—in which the formation of the second pathway requires the first pathway already to be formed, and the formation of the third pathway requires the second pathway already to be formed, and so on—in between my 2nd and 3rd birthdays en route to acquiring this immediate capacity to think. A characterization of the developmental distance I had to cross, on my 2nd birthday, before I had the immediate capacity to think, could take one of two forms:

\[ \text{DDIS}_{\text{time}} (\text{me on my 2nd birthday}) \ (\text{the immediate capacity to think}) = 1 \text{ year} \]
Now consider a second example. Assume that I injured my brain on my 22\textsuperscript{nd} birthday, and that I eventually re-acquired the immediate capacity to think on my 23\textsuperscript{rd} birthday. Assume also that I had to form 10 new “nested” neural pathways in between my 22\textsuperscript{nd} and 23\textsuperscript{rd} birthdays en route to re-acquiring this immediate capacity to think. A characterization of the developmental distance I had to cross, on my 22\textsuperscript{nd} birthday, before I had the immediate capacity to think, could take one of two forms:

\[ \text{DDIS}_{\text{steps}} (\text{me on my 22\textsuperscript{nd} birthday}) (\text{the immediate capacity to think}) = 10 \text{ pathways} \]

\[ \text{DDIS}_{\text{time}} (\text{me on my 22\textsuperscript{nd} birthday}) (\text{the immediate capacity to think}) = 1 \text{ year} \]

The steps-based way of characterizing developmental distance is connected to the concept of a hierarchy of capacities. Roughly, a single developmental step corresponds to a single “order” of capacities in a given hierarchy of capacities. Consequently, in the pair of examples above, I have a 10\textsuperscript{th}-order capacity to think on both my 2\textsuperscript{nd} birthday and on my 22\textsuperscript{nd} birthday. As I move one developmental step closer to having the immediate capacity to think, I thereby move one order lower in my hierarchy of capacities related to thinking. The final developmental step I take in acquiring the immediate capacity to think is the step from the second-order capacity to think to the first-order capacity to think.

Since there are different ways of individuating developmental steps, there are therefore different ways of individuating orders in a hierarchy of capacities. In the pair of examples above, instead of focusing on neural pathways, we could have focused on some smaller step, such as cell divisions, or some larger step, such as specific systems of neural pathways. Consequently, the number of orders in a given hierarchy of capacities is a function of just how
fine-grained one wants the hierarchy to be. A hierarchy can have a handful of orders, or billions of orders, depending on how it is described.

Although there is not a necessary connection between a given time-based way of calculating developmental distance and a given steps-based way of calculating developmental distance,\(^{18}\) we are usually on safe ground in assuming them to mirror one another. In the pair of examples above, we would have been on safe ground assuming that me-on-my-2\(^{nd}\)-birthday and me-on-my-22\(^{nd}\)-birthday were the same number of developmental steps away from having the immediate capacity to think, even if we did not know for sure the exact details of what those developmental steps were. In general, if x and y are both human organisms, and if x and y are the same developmental distance from having the immediate capacity to think, on some time-based way of calculating developmental distance, then it is usually safe to assume that x and y are the same developmental distance from having the immediate capacity to think, on some steps-based way of calculating developmental distance.

Consequently, even if we did not know for sure the exact details of what the developmental steps were that I had to cross between my 2\(^{nd}\) and 3\(^{rd}\) birthdays on the one hand and my 22\(^{nd}\) and 23\(^{rd}\) birthdays on the other, and even if all we knew was the times it would take for me-on-my-2\(^{nd}\)-birthday and me-on-my-22\(^{nd}\)-birthday to acquire the immediate capacity to

\(^{18}\) Two examples illustrate this. First, it is possible for two individuals, x and y, to be the same developmental distance from having the immediate capacity to think, on the time-based way of calculating developmental distance, yet be different developmental distances from having the immediate capacity to think, on the steps-based way of calculating developmental distance. This would happen if x and y both take 1 year to develop to the point where they have the immediate capacity to think, but x takes 10 steps during this 1 year and y takes 20 steps during this 1 year. Second, and conversely, it is possible for x and y to be different developmental distances from having the immediate capacity to think, on the time-based way of calculating developmental distance, yet be the same developmental distance from having the immediate capacity to think, on the steps-based way of calculating developmental distance. This would happen if x and y both take 10 steps to develop to the point where they have the immediate capacity to think, but x takes 1 year to take these 10 steps while y takes 2 years to take these 10 steps.
think, we would have been on safe ground assuming that me-on-my-2\textsuperscript{nd} birthday and me-on-my-22\textsuperscript{nd} birthday had equivalent “higher-order” capacities to think.

The fact that that an individual can have a higher-order capacity to do some activity, even though it may be a year or more before it has the immediate capacity to do that activity, can be combined with the distinctions between active and passive capacities, identity-preserving and compositional capacities. Consider, for example, the following concern expressed by Michael Tooley:

To characterize potential persons as entities that have a passive potentiality for becoming persons would have the consequence that random collections of matter that could, with sufficient knowledge and technological advances, be transformed into human organisms, would have to be classified as potential persons.\textsuperscript{19}

Tooley’s concern here can be reformulated by replacing the language of potential persons with the language the higher-order capacity to think:

To characterize those with a higher-order capacity to think as entities that have a passive higher-order capacity to think would have the consequence that random collections of matter that could, with sufficient knowledge and technological advances, be transformed into thinking things, would have to be classified as entities with the higher-order capacity to think.

This concern amounts to the claim that, by including a thing’s passive capacities in its set of higher-order capacities, the result is an unwelcome proliferation, and indeed, explosion, of capacities in the world. Virtually everything will turn out to have the higher-order capacity to do

\textsuperscript{19} Tooley, 168.
everything. For example, random collections of matter will turn out to have the higher-order capacity to think.

If what it means for an entity to ‘become’ a person is understood along the same lines as what it means for a lump of bronze to ‘become’ a statue, then Tooley’s quote, and its reformulation in terms of higher-order capacities, is surely correct, although the clause about “sufficient knowledge and technological advances” is not necessary. For the relevant transformations already occur every day without much technology and without much knowledge: all it takes is a living human organism and a certain amount of time. Human organisms are constantly getting transformed, acquiring new parts through processes like inhalation, nutrition, and hydration, and losing current parts through processes like exhalation, excretion, and perspiration. These mereological transformations (from the Greek *meros*, “part”) are ubiquitous; they are the very stuff of life itself.

Consider an example of a mereological “switch” in which a human organism and an island of material stuff gradually exchange all their simplest parts with one another between times \( t \) and \( t^* \). Label the simplest parts constituting the human organism at \( t \) the \( ps \), and the simplest parts constituting the island at \( t \) the \( qs \). Label the human organism constituted by the \( ps \) at \( t \) Lew, the island constituted by the \( qs \) at \( t \) Honolulu, the human organism constituted by the \( qs \) at \( t^* \) Kareem, and the island constituted by the \( ps \) at \( t^* \) Caribbea. Imagine that the rate of exchange among \( ps \) and \( qs \) was such that whenever a \( q \) leaves Honolulu and becomes a part of Lew, a \( p \) leaves Lew to take the place of that \( q \). This example can be diagrammed by letting solid black lines represent the flow of \( ps \) and \( qs \) and by letting dotted lines represent the “continuity” of the island and the man:
This diagram illustrates several distinct sorts of potentiality:

1. Honolulu is potentially Kareem.
2. Lew is potentially Caribbea.
3. Honolulu is potentially Caribbea.
4. Lew is potentially Kareem.

The sense of potentiality found in (1) and (2) is what was earlier labeled compositional potency. These claims rely on the constitutive sense of ‘is’: (1) must be paraphrased as, and means no more and no less than, the claim that the $qs$ that constitute Honolulu have the potential to constitute Kareem. The same thing is true of (2).

Assume that Kareem has the immediate capacity to speak Chinese, but that Lew does not. If the concept of a higher-order capacity is explicated using the sense of potentiality found in (1) and (2), then Honolulu has the higher-order capacity to speak Chinese. But this is absurd: an island of material stuff surely does not have the higher-order capacity to speak Chinese. This suggests that the concept of a higher-order capacity should not be explicated using the sense of potentiality found in (1) and (2).

The sense of potentiality found in (4) is not compositional potency but what was earlier labeled identity-preserving potency. This claim relies on the ‘is’ of identity: (4) must be
paraphrased as, and means no more and no less than, the claim that Lew has the potential to become Kareem while remaining the self-same individual over time.

The concept of a higher-order capacity should be explicated using the sense of potentiality found in (4). Return again to the assumption that that Kareem has the immediate capacity to speak Chinese, but that Lew does not. If the concept of a higher-order capacity is explicated using the sense of potentiality found in (4), then Lew does, while Honolulu does not, have the higher-order capacity to speak Chinese. For Lew does, but Honolulu does not, have the potential to become Kareem while remaining the self-same individual over time.

The metaphysical picture I am adopting here is one made clear with more simple examples. Even if the saying is true that you are what you eat, still, you are not what you eat until you eat it, and even then, what you eat is not literally you—it merely composes you: even while it composes you, it does not do all the things you do, and before you ate it, it did not have the higher-order capacities to do all the things you do. Even if God made Adam out of a pile of dust we name Dusty, still, Dusty did not have the higher-order capacities to do all the things Adam did. Even if the saying were true that little boys are made of snips and snails and puppy dog tails, still, snips and snails and puppy dog tails do not have the higher-order capacities to do all the things little boys do (and the same thing goes for little girls, who are supposedly made out of “sugar and spice and all things nice”).

Claim (3) is controversial because philosophers do not agree on what it takes for an island of material stuff to persist through time. If (3) is interpreted like I have interpreted (4), then many will claim (3) is false. For it is not true that Honolulu has the potential to become Caribbea while remaining the self-same individual over time. Islands are not individuals in the

---

same way that organisms are, and islands do not persist in the same way that organisms do. If (3) is to be true at all, it must be paraphrased as, and mean no more and no less than, the claim that Honolulu was able to ‘become’ Caribbea by a gradual replacement of its parts. Of course, if that is all (3) means, then the same thing could be said of (4): for Lew is able to ‘become’ Kareem by a gradual replacement of its parts. Both human organisms and islands have higher-order capacities to be constituted by new sets of particles. But even if I agree with this point, the higher-order capacity to be constituted by new sets of particles is not the higher-order capacity I have in mind with (4).

3. As Long As You Are Human, You Have a Certain Set of Higher-Order Capacities

You are human. You are a human organism. But you have thousands of properties right now besides the property of being human. For example, you have a certain shape right now, a certain mass right now, and a certain mental state (thinking) right now. Let us call all such properties your immediate properties. Among your immediate properties are your immediate capacities. For example, you have the immediate capacity to think right now, the immediate capacity to smile right now, and the immediate capacity to breathe right now. You have these and thousands of other immediate capacities right now. I would like to explore some of the ways in which you could lose some of these immediate capacities, as well as some of the ways in which you could lose some of your other immediate properties. In particular, I would like to focus on what sorts of capacities you hold on to if all of your immediate properties are removed besides the property of your humanity. I am going to argue that, even when we strip away your other immediate properties in this way, you still retain a certain set of higher-order typical human capacities.
Consider the following three properties:

1. the property of having the immediate capacity to think
2. the property of having the higher-order capacity to think
3. the property of being human

Now consider the following question: could you have the higher-order capacity to think even when you do not have the immediate capacity to think? There are at least three approaches for arguing that the answer to this question is “yes”. The first approach focuses on times of your life when you have not yet thought: for example, when you were a preconscious fetus in your mother’s womb. Unfortunately, this approach is not persuasive to people who are ambivalent about the idea that you ever existed as a preconscious fetus in your mother’s womb. For if you do not exist at a certain time, then you cannot have any capacities (such as higher-order capacities) at that time. The second approach focuses on times of your life when you are no longer have the immediate capacity to think: for example, when your consciousness becomes lost in the later stages of a terminal brain disease. Unfortunately, this approach is not persuasive to people who are ambivalent about the idea that you would continue to exist in the later stages of such diseases. For, once again, if you do not exist at a certain time, then you cannot have any (higher-order) capacities at that time.

A third approach focuses on times of your life when you lose the immediate capacity to think but eventually regain it: for example, when you are temporarily unconscious due to being asleep, anesthetized, or comatose. We can call such times “temporary changes” in your immediate capacity to think. It seems that you still possess the higher-order capacity to think during these temporary changes. Fortunately, this approach is persuasive to almost everyone, since almost everyone thinks that you would continue to exist during these temporary changes.
Those who do not think you exist during the times of these temporary changes would have to deny one of the following two ideas, each of which is very difficult to deny: (1) You exist both before the time of the “temporary change” and after the time of the “temporary change”; (2) It is not possible for an individual to have temporal gaps in its existence; or, put differently, for any individual x and any time t, it is not possible for x to exist both before t and after t unless x exists during t.

Let us call this third approach the temporary change argument. The temporary change argument can be formulated more precisely by considering what happens to you as you move through three times, t₁, t₂, and t₃. Although you are human and have the higher-order capacity to think at each of these times, you have the immediate capacity to think only at t₁ and t₃:

<table>
<thead>
<tr>
<th></th>
<th>t₁</th>
<th>t₂</th>
<th>t₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you have the higher-order capacity to think?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Do you have the immediate capacity to think?</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Are you human?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

The temporary change argument can be set out as follows:

(1) You possess the higher-order capacity to think at t₂.

Therefore,

(2) You do not need to have the immediate capacity to think at a given time in order to possess the higher-order capacity to think at that time.

(3) If there is no other feature of yours that we could base your possession of the higher-order capacity to think on at t₂, then we should conclude that, as long
as you remain human at a given time, you have the higher-order capacity to think at that time.

(4) There is no other feature of yours that we could base your possession of the higher-order capacity to think on at $t_2$.

Therefore,

(5) As long as you remain human at a given time, you have the higher-order capacity to think at that time.

This argument’s most controversial premise is (4). It will no doubt be objected that there are other available features of yours that we could base your possession of the higher-order capacity to think on: for example, the property of having a functioning cerebral cortex.

But this objection is mistaken. Replace the question “Are you thinking?” in the above chart with the question “Do you have a functioning cerebral cortex?”. Imagine that you receive an injury to your cerebral cortex, which causes that organ to cease functioning for a period of time before it resumes functioning. You retain the higher-order capacity to think during this period of time, even though you do not have a functioning cerebral cortex. It may be a “passive” higher-order capacity rather than an “active” one, if the damage to the cortex requires a good deal of external assistance, such as surgery, to fix. But it is still a bona fide higher-order capacity. Consequently, the temporary change argument can be recast at step (2) by simply replacing “think” with “possess a functioning cerebral cortex.”

Perhaps it will be objected that the relevant property is not having a functioning cerebral cortex, but simply having a cerebral cortex. But this objection is also unsatisfactory, since the chart and the argument can be revised even when “functioning cerebral cortex” is replaced with “cerebral cortex”. Imagine again that you receive an injury to your cerebral cortex, which causes
that organ to cease functioning for a period of time. However, in this case, surgeons remove the cerebral cortex from your body for a period of time in order to study and repair it. (Although scientists are currently not able to do this sort of thing, there is no reason in principle to suppose that it cannot be done.) When your damaged cerebral cortex is out of your body, it seems that you do not have a cerebral cortex at all. Yet you still have the higher-order capacity to think during this time.

One might object to this thought experiment on the grounds that you still have a cerebral cortex during the period of time in which it is out of your body. After all, it is your cerebral cortex, even if it is in a laboratory across the room, or across the world. But the thought experiment can be revised to accommodate this objection. Imagine that the surgeons, after removing the cerebral cortex from your body and studying it, decide that it is too damaged to repair. They decide to make you a new cerebral cortex from stem cells, or from whatever the latest techniques in regenerative medicine happen to be. The surgeons destroy your original cerebral cortex in an incinerator, go on a vacation for two weeks, and then begin manufacturing a new cerebral cortex for you when they return. Clearly, during the two weeks when your surgeons are on vacation, you do not have a cerebral cortex at all. Yet you still have the higher-order capacity to think during this time.

Some will object at this point that my use of the personal pronoun “you” in the temporary change argument is being used far too liberally. For the states that make for personal identity, this objection goes, might not be preserved when the cerebral cortex is damaged, and are certainly not preserved when the cerebral cortex is destroyed. Consequently, this objection concludes, it is false that “you” have the higher-order capacity to think when your cerebral cortex is destroyed. For “you” do not exist when your cerebral cortex is destroyed.
I disagree with this objection, partly for reasons that have already been discussed. Imagine you knew, in advance, that the cerebral cortex damage was going to occur, and that you knew, in advance, the various scenarios that might unfold when the surgeons try to repair it—including the destruction-vacation-reconstitution scenario. It seems to me perfectly rational for you to look forward to being “made whole” again, to look forward to thanking your surgeons once the whole thing was over, and to anticipate telling them how much you appreciate the fact that they made sure you were taken good care of even when they were on vacation. I recognize that this approach disagrees with the intuition some philosophers have when considering thought-experiments involving upper brain transplants, since the intuition is that wherever your upper brain goes, there you also go. But I am not convinced by that intuition. The upper brain is just like upper liver: it is a mere part of you that can be damaged, destroyed, and re-constituted before being reinserted into you.

There are two noteworthy implications that follow directly from the fact that as long as you are human, you have the higher-order capacity to think. First, it follows that as long as you are already human, you already have the higher-order capacity to think, even before your very first moment of having the immediate capacity to think. Variations on the temporary change argument, such as the case of the doctors incinerating the cerebral cortex while they go on vacation for two weeks, make the point that a person (“you”) can exist while it is not having any experiences at all, and indeed while it has no physical traces of having ever had any experiences at all. There is no reason to think that this point is any less true before your first experience than it is after your first experience. Second, it follows that as long as you are still human, you still have the higher-order capacity to think even after your very last moment of having the immediate capacity to think. These two implications correspond to the first two approaches,
considered above, for arguing that you have the higher-order capacity to think even at times when you do not have the immediate capacity to think. Recall how the first approach emphasized the time of your fetal development, and the second approach emphasized the time of your of terminal psychological deterioration. As long as you are human during these times, you have the higher-order capacity to think during these times.

The temporary change argument is a philosophical tool for testing and sharpening your beliefs about what it takes for you to persist through time. It is somewhat different than the more popular philosophical tools that are used for this task. Most philosophical approaches to your identity over time begin with the assumption that you exist right now, and then proceed straightaway into an examination of different thought experiments that focus on what sort of events could bring about the end of your existence, or on what sort of events could have sparked the beginning of your existence. The temporary change argument begins with the assumption that you exist—or at least appear to exist—both right now and at some later time, and then proceeds into an examination of different thought experiments that focus on what sort of events could happen in between now and this later time. Thus while most philosophical approaches focus on the beginning and end of our existence, the temporary change argument focuses upon the middle, and then works its ways out.

Still, there are a number of questions that need to be answered. Just how far back in our biological life can we truly speak of ourselves possessing the higher-order capacity to think? Also, what about humans who have always lacked the genetic equipment that helps build the brain structures used for thinking? These questions will be answered, respectively, in Chapters Four and Five. At present, it is important to explain how we can get from the higher-order capacity to think to a set of typical human capacities.
The structure of the temporary change argument can be repeated for any two properties you have in addition to the property of being human. But the structure of the temporary change argument can also be repeated for the entire set of immediate capacities you have. Let A be the entire set of immediate capacities that you possess right now. Let B be the set of higher-order capacities that corresponds to A. A temporary change argument can show that as long as you remain human at a given time, you have B at that time. If we drop the qualifiers “at a given time” and “at that time”, the temporary change argument seems to show that, as long as you remain human, you have a certain set B of higher-order capacities.

The temporary change argument, as stated thus far, only focuses on higher-order capacities that correspond to the immediate capacities you possess right now. But let C be the set of immediate capacities that I possess right now. Let D be the set of higher-order capacities that corresponds to C. A temporary change argument can show that as long as I remain human, I have D at that time. But you and I are different. C and D are different. There is still an important gap between the idea that every human has some set or other of higher-order capacities, and the idea that there is some set of typical human capacities that every human has in common. What might close this gap?

I think this gap can be closed by what I shall call “the normalizing approach.” You, me, Hillary Clinton, and Shaquille O’Neal are each what we might call normal adult human persons. There seem to be certain immediate capacities that all normal adult human persons have in common, and although I do not intend to give an exhaustive list of them, the immediate capacity to think, the immediate capacity to communicate using language, the immediate capacity to pray, and the immediate capacity to understand humor would likely be among them. In any case,
possession of these and other immediate capacities is precisely what makes someone a “normal”
adult human person.

Let $G$ be the set of general immediate capacities possessed by all normal adult human
persons. $G$ will only include “general” capacities that distinct individuals could have in
common, such as the capacity to think and remember. $G$ will not include “specific” capacities
that distinct individuals could not have in common, such as the capacities I have to think my
specific thoughts and to remember my specific experiences. Let $H$ be the set of higher-order
capacities that corresponds to $G$. The normalizing approach claims that, as long as you remain
human, you have a certain set $H$ of higher-order capacities. $H$ is the set of typical human
capacities.

There are still several questions that need to be carefully explored before this normalizing
approach can be fully endorsed. For example, focus again on the capacity to think. Just how far
back in our biological life can we truly speak of ourselves possessing the higher-order capacity
to think? Can we speak truthfully about having the higher-order capacity to think as infants, or
fetuses, or embryos? Questions like these will be taken up in more detail in Chapter Four. Also,
just what sorts of changes can we permit to happen to us before we deny that we still have the
higher-order capacity to think? Can we speak truthfully about having the higher-order capacity
to think after terrible brain diseases have taken their toll, or in individuals with severe genetic
deficiencies? Questions like these will be taken up in more detail in Chapter Five.
CHAPTER THREE: WHY CAPACITIES MATTER

This chapter defends the claim that if an entity has a set of typical human capacities, it has serious moral status. Since the strong moral presumption against killing is the aspect of serious moral status that I wish to focus on, I will sometimes speak only in terms of this presumption. The main argument given for this claim is a temporary change argument. After various objections to this argument are considered, the chapter closes by explaining why the present view is to be preferred over an important alternative view, which claims that the only entities that have a right to continued existence are actual, continuing subjects of experience.

1. Another Temporary Change Argument

The structure of the argument in this section is as follows. There are times when there is a strong moral presumption against killing you, but the only possible thing we could base this strong moral presumption on is your higher-order capacities. Since we do base this strong moral presumption against killing you on something, it follows that this presumption must be based on your higher-order capacities.

For the sake of simplicity, the discussion of this section will be formulated in terms of a single capacity, and this capacity will be the capacity to think. But since the problem discussed in this section is a general problem, the discussion could be formulated in terms of any single capacity (e.g. the capacity to experience pleasure and pain) or any set of capacities (e.g. the set of capacities possessed by any normal adult human person).

Consider the following question: is there a strong moral presumption against killing you when you do not have the immediate capacity to think? There are three approaches for arguing
that the answer to this question is “yes”. One approach focuses on times of your life when you have not yet attained the immediate capacity to think: for example, when you were an infant. Unfortunately, this approach is not persuasive to people who are ambivalent or undecided about whether there is a strong moral presumption against killing human infants. A second approach focuses on times of your life when you no longer have the immediate capacity to think: for example, when you become psychologically incapacitated in the later stages of a disease like Alzheimer’s. Unfortunately, this approach is not persuasive to people who are ambivalent or undecided about whether there is a strong moral presumption against killing those who are in the later stages of such diseases. A third approach focuses on times of your life when you go through what can be called temporary changes: for example, when you are temporarily unconscious due to being asleep, anesthetized, or comatose. Unfortunately, this third approach is sometimes left imprecise and undeveloped by its adherents. As a result, this third approach is often simply ignored or brushed aside by its detractors. It is often viewed as a conceptual wrinkle that can be easily ironed out while still preserving the central importance of the immediate capacity to think.

Let us call this third approach the temporary change argument. The temporary change argument is actually far more promising than its detractors realize. This remainder of this section seeks to formulate the temporary change argument more precisely and to argue that three of the most plausible strategies for replying to it are unsatisfactory.

The temporary change argument can be formulated more precisely by considering your immediate capacity to think as you move through three times, \( t_1 \), \( t_2 \), and \( t_3 \). Although you have a higher-order capacity to think at each of these times, you only have the immediate capacity to think at \( t_1 \) and \( t_3 \):
The temporary change argument begins with the assumption that you have serious moral status (and hence there is a strong moral presumption against killing you) at \( t_2 \). From this assumption, the temporary change argument draws a conclusion and makes a suggestion. Negatively, it concludes that your possession of the immediate capacity to think at a given time \( t \) is not what grounds the strong moral presumption against killing you at \( t \). Positively, it suggests that your possession of the higher-order capacity to think at \( t \) is what grounds the strong moral presumption against killing you at \( t \).

An initial objection should be mentioned at the outset. It might be thought that, although temporary change scenarios are sound ways of establishing theses of type (1), they are not a good way of establishing theses of type (2):

1. Possession of property P is not a necessary condition for an entity having serious moral status.
2. Possession of property Q is a sufficient condition for an entity having serious moral status.

The problem with using temporary change scenarios to establish theses of type (2), according to this objection, is that while Q is present in the temporary change situation, other properties may be present as well. If property R is present as well, then the temporary change scenario does
itself not enable one to determine whether Q is sufficient, or R is sufficient, or both are sufficient.

I believe that this objection can be overcome by recognizing that theses of type (2) can be established by an inference to the best explanation, and that temporary change scenarios are part of what make the explanation best. Let us say that temporary changes with humans lead us to recognizing the existence of two properties that might be sufficient for serious moral status. The first property, Q, is the property of having a set of (higher-order) typical human capacities. The second property, R, is the property of being human, which, as Chapter One claimed, amounts to having the same basic genotype as you and I. Which of these two properties—Q or R—is sufficient for an entity having serious moral status?

An answer to this question will emerge if we consider cases of what might be called “shumans” who go into a temporary change. Shumans, let us say, are phenotypically indistinguishable from humans, but they have ZNA instead of DNA in their cells, so they have a different basic genotype than you and I. Shumans have all the same sorts of immediate capacities, and all the same sorts of higher-order capacities, as humans. Let us say that temporary changes with shumans lead us to recognizing the existence of two properties that might be sufficient for serious moral status. The first property, Q, is, once again, the property of having a set of (higher-order) typical human capacities. The second property, S, is the property of being shuman, which amounts to having the same basic genotype as typical shumans with ZNA. Now, then, if S was sufficient for serious moral status, this would allow us to explain why shumans have serious moral status, but it would not help us to explain why humans have serious moral status. Conversely, if R was sufficient for serious moral status, this would help us to explain why humans have serious moral status, but it would not help us to explain why shumans
have serious moral status. However, if Q were sufficient for serious moral status, this would help us to explain why both humans and shumans have serious moral status during their temporary changes. The fact that Q would have this explanatory power is what makes it the best explanation for why both humans and shumans have serious moral status during their temporary changes. (Other candidate properties that humans and shumans both share during their temporary changes would also need to be seen to be inadequate: for example, the property T of being composed of physical simples is not sufficient for having serious moral status, since it is easy to think of entities with property T that do not have serious moral status, such as rocks.)

Nevertheless, there are at least three other strategies for meeting the temporary change argument that will now be carefully considered. One strategy for meeting the temporary change argument is to ground the strong moral presumption against killing you at $t_2$ in the fact that you did have the immediate capacity to think at $t_1$. According to this first strategy, the strong moral presumption only exists in the case of the temporary change in virtue of the fact that the temporary change is a change from a previous state of a certain sort. A second strategy is to ground the strong moral presumption against killing you at $t_2$ in the fact that you will have the immediate capacity to think at $t_3$ if allowed to live. According to this second strategy, the strong moral presumption only exists in the case of the temporary change in virtue of the fact that the temporary change is temporary. A third strategy is to combine these first two strategies; that is, to ground the strong moral presumption against killing you at $t_2$ in the fact that either you did have the immediate capacity to think at $t_1$, or you will have the immediate capacity to think at $t_3$ if allowed to live. According to this third strategy, the strong moral presumption only exists in the case of the temporary change in virtue of the fact that the temporary change is either temporary or a change from a previous state of a certain sort. If one of these three strategies is
satisfactory, then it will not be necessary to accept the positive suggestion mentioned a moment ago, namely, that your possession of the higher-order capacity to think at a given time $t$ is what grounds the strong moral presumption against killing you at $t$. Nevertheless, it will now be argued that each of these three strategies is unsatisfactory.

1.1. A First Strategy for Replying

The first strategy for meeting the challenge of the temporary change is to ground the strong moral presumption against killing you at $t_2$ in the fact that you did, actually, have the immediate capacity to think at $t_1$. To see why the first strategy is unsatisfactory, consider the following two cases.

First, consider a case where two human organisms, A and B, are identical twins and are nurtured and developed in a highly refined science laboratory from conception onwards: they are exactly similar in their genetic constitution, environmental stimuli, and so on, throughout their entire biological lives. A and B are grown up like this for many years, but neither of them are allowed to develop the immediate capacity to think. Each is highly developed enough to the point where a push of a certain button will allow her to develop into the state of having the immediate capacity to think. Now imagine that, on a certain day, both A and B are going to get their respective buttons pushed. But a double malfunction occurs. A’s button works for a moment, but then stops working, while B’s button gets stuck and does not work at all. The result is that A, for a moment, is allowed to develop the immediate capacity to think, but then lapses back into the state she was in before the button was pushed, whereas B is not allowed to develop the immediate capacity to think—not even for a moment. The net effect of this double
malfunction is that neither A nor B have the immediate capacity to think, but A did actually have
the immediate capacity to think—at least for a moment.

The first strategy would hold that there is now a strong moral presumption against killing
A, but not a strong moral presumption against killing B. But this is hard to believe. Imagine
walking into the lab shortly after this malfunction had happened, without knowing how it
happened: that is, even though you know that one of these two had her special moment, you do
not know whether it was A or B. A scientist tells you that there is a strong moral presumption
against killing one of these two human beings but not the other. You would be quite perplexed.
After all, A and B will both develop the immediate capacity to think at the same time if they are
just allowed to. It seems reasonable to think that if there is a strong moral presumption against
killing A, there is also a strong moral presumption against killing B. The mere fact that A had
once possessed the immediate capacity to think should not bear the moral weight that the first
strategy insists it bear.

It might be objected that this example relies upon the implausible assumption that the
first strategy would be satisfied with the merest instant of possessing the immediate capacity to
think. However (this objection continues), many writers hold that it is crucial whether actual
thinking has occurred, and indeed actual thinking of a special kind: namely, self-awareness over
time, accompanied by some pro-attitude, such as desire or care, that attaches to what continues
over time. So (this objection concludes) our intuitions about this case would be very different
indeed if A spent a period of time thinking about herself, recognizing that she is the same
enduring one living, wanting that living to go on, and then sinking back into unconsciousness.

In reply, I do not think that amending this example so that A has actual self-awareness
and pro-attitudes will vindicate the first strategy. However, in order to let this objection be as
strong as possible, consider a very different sort of case that does not involve developing identical twins. Imagine A is a normal, healthy adult with a rich and satisfying life, endowed with the immediate capacity to think, and also endowed with self-awareness and the desire to go on living. Now imagine A undergoes a temporary change so that she is currently at $t_2$, and at $t_2$ she gets replicated in one of Derek Parfit’s famous replication booths:¹ A is preserved intact and is not destroyed, but her perfect replica B is instantly produced across the laboratory. B has precisely the same sort of molecular structure that A had, and is functioning at precisely the same level as A. Furthermore, B has exactly the same capacities as A. Both A and B lack the immediate capacity to think, and both A and B will have the immediate capacity to think, along with self-awareness and the desire to go on living, at the same time if they are just allowed to.

Once again, however, the first strategy would hold that there is a strong moral presumption against killing A, but not a strong moral presumption against killing B. But this is hard to believe. Imagine walking into the lab shortly after the replication had happened, without knowing how it happened: that is, even though you know that one of these two is a replica, you don’t know whether it is A or B. A scientist tells you that there is a strong moral presumption against killing one of these two human beings but not the other. You would be quite perplexed. After all, A and B will both develop the immediate capacity to think at the same time if they are just allowed to. It seems reasonable to think that if there is a strong moral presumption against killing A, there is also a strong moral presumption against killing B. The mere fact that A had once possessed the immediate capacity to think, along with actual self-consciousness and pro-attitudes, should not bear the moral weight that the first strategy insists it bear.

In these cases involving identical twins and replicas, we can admit that there are some morally relevant differences between A and B without admitting that a strong moral presumption against killing applies only to A. For example, in the replica case, imagine that A had worked hard and put her earnings into a savings account before slipping into a temporary coma and getting replicated. Now imagine that A and her replica B both come out of their temporary comas, and both claim to own the money in the savings account. I believe that A has a stronger claim to the money than B, since A actually saved the money whereas B merely has pseudo-memories of saving the money. However, not all morally relevant properties are dependent like this upon the actual history of an individual. In particular, moral status properties, such as the strong moral presumption against killing, do not seem to be so dependent upon an individual’s actual history.

The upshot of these two cases is this. There can still be a strong moral presumption against killing you at $t_2$ even if it is not true that you $did$ have the immediate capacity to think at $t_1$. Thus the first strategy for replying to the temporary change argument is unsatisfactory.

1.2. A Second Strategy for Replying

The second strategy for meeting the challenge of the temporary change is to ground the strong moral presumption against killing you at $t_2$ in the fact that you $will$ have the immediate capacity to think at $t_3$ if allowed to live. To see why the second strategy is unsatisfactory, it is important to see how it is similar to what have been called deprivation accounts of the wrongness of killing.

Deprivation accounts of the wrongness of killing emerge from the combination of three claims:
(1) The wrongness of a particular act of killing is solely a function of the misfortune of the particular death caused by that particular act of killing.

(2) The misfortune of a particular death is solely a function of the goods that this particular death deprives the entity that dies of.

(3) The goods that a particular death deprives the entity that dies of are the goods that the entity would have had, if, contrary to fact, the entity had not died that particular death.

Such accounts of the wrongness of killing are very similar to this second strategy for replying to the problem of the temporary change. While deprivation accounts emphasize the future goods you would have enjoyed had you not been killed, this second strategy emphasizes the future state you would have been in had you not been killed. (Indeed, if one thinks that having the immediate capacity to think is itself a good one would have had, then the second strategy is itself a deprivation account of the wrongness of killing. But one need not think this in order to see the relevant similarity.)

Deprivation accounts of the wrongness of killing face what has been called a problem of overdetermination. For claims (2) and (3), taken together, entail something very odd whenever the following conditional holds true: if, contrary to fact, you did not die the particular death you did die, then you would have died some other way at the same time you actually did die the particular death you did die. Claims (2) and (3), taken together, entail that, whenever the just-mentioned conditional holds true, your particular death does not deprive you of anything and is therefore not a misfortune. When this entailment is combined with claim (1), the result is that, whenever your death is overdetermined, someone can kill you without doing anything wrong.

---

2 See McMahan, 117-120.
To see why this is a problem, consider the following example. A man lives in a place where racial tensions are at the boiling point, and where race riots are known to lead to lynchings that kill innocent people. He is walking down a deserted country road one night. Out of nowhere, a boy runs past him who is being chased by a mob of angry people shouting “Lynch him!” The man realizes that, no matter what he does to try and prevent it, the crowd will catch the boy and lynch him before any help arrives. In the past, the man has tried—in vain—to stop such crowds by pleas, by arguments, and by the use of force. Nothing ever works. So this time, the man decides to join the crowd; he himself catches up to the boy and tackles him to the ground, he himself wraps the noose around the boy’s neck, and he himself pulls on the rope that hangs the boy to death from the limb of a nearby tree.

Most of us want to say that this man has done something seriously wrong. But the deprivation account of the wrongness of killing cannot say this. For even if the man had not killed the boy, the boy would have died at the same time he actually died, because someone else in the crowd would have done the same sort of thing the man did, at the same time the man did it. Therefore, the man’s action did not deprive the boy of any of the goods he would have had if the action had not been performed. According to the deprivation account of the wrongness of killing, the man did not do anything wrong.

This case illustrates how the inevitability of certain sorts of outcomes is irrelevant to the wrongness of certain sorts of actions. It also illustrates the inadequacy of grounding a strong moral presumption against killing in the fact that the individual being killed will be in a certain state in the future if the killing is not performed. But these illustrations are relevant to the problem of the temporary change.
Consider someone undergoing a temporary change in a situation such that they will be killed no matter what you try to do to prevent it. For example, imagine a female college student becoming so drunk at a fraternity party that she blacks out and will not recover for several hours, and no matter what you do, she is going to be killed while unconscious. Or imagine a woman undergoing anesthesia for a sort of surgery, and no matter what you do, she is going to be killed while under anesthesia. Or imagine a girl in a temporary coma for a period of time, and no matter what you do, she is going to be killed while in this temporary coma. Surely, in each of these cases, there is still a strong moral presumption against killing the individual during her temporary change. The fact that someone is going to kill her does not give you any reason to think that you may now kill her.

The upshot is this. There can still be a strong moral presumption against killing you at \( t_2 \) even if it is not true that you will have the immediate capacity to think at \( t_3 \) if allowed to live. Thus the second strategy for replying to the temporary change argument is unsatisfactory.

### 1.3. A Third Strategy for Replying

The third strategy for replying to the temporary change argument is to combine the first two strategies; that is, to ground the strong moral presumption against killing you at \( t_2 \) in the fact that either you did have the immediate capacity to think at \( t_1 \), or you will have the immediate capacity to think at \( t_3 \) if allowed to live. But it is hard to see how this is going to help. To see the inadequacy of this combinatorial approach, all one needs to do is to present a counterexample which combines the features of the previous counterexamples.

Recall again the case where A is a normal, healthy adult with a rich and satisfying life, endowed with the immediate capacity to think, and also endowed with self-awareness and the
Now imagine A undergoes a temporary change so that she is currently at \( t_2 \): perhaps she gets so drunk at a fraternity party that she blacks out and will not recover for several hours, or perhaps she undergoes anesthesia for a sort of surgery, or perhaps she is in a temporary coma for a period of time. In any event, at \( t_2 \) she gets replicated in one of Derek Parfit’s famous replication booths: A is preserved intact and is not destroyed, but her perfect replica B is instantly produced across the laboratory. The replica produced, B, has precisely the same sort of molecular structure that A had, and is functioning at precisely the same level as A. B also has exactly the same sorts of capacities that A has. Both lack the immediate capacity to think, and both will have the immediate capacity to think, along with self-awareness and the desire to go on living, at the same time if they are just allowed to.

But now imagine that no matter what you do, B is going to be killed. Perhaps there is a team of doctors who are determined to remove all of B’s vital organs in order to transplant them to patients around their hospital. Whatever the reason, someone is going to kill B and nothing you can do will prevent this from happening.

The third strategy would hold that there is a strong moral presumption against killing A, but not a strong moral presumption against killing B. But this is hard to believe. Imagine walking into the lab shortly after the replication had happened, without knowing how it happened: that is, even though you know that one of these two is a replica, you don’t know whether it was A or B. A scientist tells you that there is a strong moral presumption against killing one of these two human beings but not the other. You would be quite perplexed. After all, both will develop the immediate capacity to think at the same time if they are just allowed to. It seems reasonable to think that if there is a strong moral presumption against killing A, there is also a strong moral presumption against killing B. The mere fact that A had once possessed the
immediate capacity to think, along with actual self-consciousness and pro-attitudes, should not bear such moral weight. Likewise, the fact that someone is going to kill B does not give you any reason to think that you may now kill B. The fact that B will inevitably be killed, before she comes out of her temporary change, is not relevant.

The upshot is this. There can still be a strong moral presumption against killing you at \( t_2 \) even if it is both not true that you did have the immediate capacity to think at \( t_1 \), and not true that you will have the immediate capacity to think at \( t_3 \) if allowed to live. Thus the third, combinatorial strategy for replying to the temporary change argument is unsatisfactory.

1.4. Implications of the Temporary Change Argument

The temporary change argument shows that your possession of the immediate capacity to think at a given time is not the basis for the strong moral presumption against killing you at that time. But since there is a strong moral presumption against killing you at \( t_2 \), and since this strong moral presumption must be based upon something, what else can it be based upon? Since the replies given by the three strategies are unsatisfactory, it seems that a very natural place to turn is the fact that you possess a higher-order capacity to think at \( t_2 \). Although the next section will consider an alternative view, and although the next few paragraphs will expand upon this claim, let us provisionally assume for the moment that the fact that you possess a higher-order capacity to think is sufficient for there to be a strong moral presumption against killing you. This is the provisional result of the temporary change argument.

There are several extensions of this provisional result of the temporary change argument that bear directly upon the claim this chapter is seeking to defend. The first extension involves the conceptual move from the sort of higher-order capacity to think that one possesses when
asleep, anesthetized, or comatose, to the sort of higher-order capacity to think that is yet “higher” in terms of its order within the same hierarchy. Although the structure of the temporary change was originally formulated by referring to an “immediate” capacity and a “higher-order” capacity, this original formulation is ambiguous because a “higher-order” capacity could be a “second-order” capacity, or a “third-order” capacity, or a “fourth-order” capacity, and so on. Once this ambiguity is recognized, it is easy enough to eliminate: we can insist, for example, that what we mean by the immediate/higher-order contrast is merely the first-order/second-order contrast:

<table>
<thead>
<tr>
<th></th>
<th>t₁</th>
<th>t₂</th>
<th>t₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you have serious moral status?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Do you have the <em>first-order</em> capacity to think?</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Do you have the <em>second-order</em> capacity to think?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

However, once we recognize that there are different ways to be a higher-order capacity to think, it now becomes possible to formulate a temporary change using the second-order/third-order contrast:

<table>
<thead>
<tr>
<th></th>
<th>t₁</th>
<th>t₂</th>
<th>t₃</th>
</tr>
</thead>
<tbody>
<tr>
<td>Do you have serious moral status?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Do you have the <em>second-order</em> capacity to think?</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
</tr>
<tr>
<td>Do you have the <em>third-order</em> capacity to think?</td>
<td>Yes</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

This reformulated temporary change is structurally isomorphic to the original temporary change. But this means that everything said above about the original temporary change will apply to this reformulated temporary change as well. The original temporary change showed how the
possession of the \textit{first}-order capacity to think was not necessary for there to be a strong moral
presumption against killing you. The reformulated temporary change will show how the
possession of the \textit{second}-order capacity to think is not necessary for there to be this presumption.

In fact, the structure of a temporary change can be utilized on any two adjacent capacities
in a hierarchy of capacities:

\begin{center}
\begin{tabular}{|l|c|c|c|}
\hline
 & t_1 & t_2 & t_3 \\
\hline
Do you have serious moral status? & Yes & Yes & Yes \\
\hline
Do you have the \textit{n}^{th}-order capacity to think? & Yes & No & Yes \\
\hline
Do you have the \textit{(n+1)}^{th}-order capacity to think? & Yes & Yes & Yes \\
\hline
\end{tabular}
\end{center}

Thus the structure of the temporary change argument can be iterated as long as there is a higher-
order capacity to think. It would seem, then, that the temporary change argument is able to show
that the fact that you possess \textit{some} higher-order capacity to think is sufficient for there to be a
strong moral presumption against killing you. This is the first extension of the temporary change
argument.

The second extension builds upon the fact that the higher-order capacity in question need
not be the restricted to the higher-order capacity to think. Recall that the capacity to think was
originally chosen for the sake of simplicity, as a convenient illustration. But any immediate
capacity of yours is such that you could experience a temporary change in that capacity. The
same thing can be said for any set of immediate capacities. Recall the example, from Chapter
Two, where G is the set of general immediate capacities possessed in common by every normal
adult human person, and where H is the set of higher-order capacities that corresponds to G. The
temporary change argument seems to be able to show that, if a you possess a certain set H of
higher-order capacities at a given time, then there is a strong moral presumption against killing you at that time.

The third extension builds upon the fact that there is no reason to limit the insights gained from the temporary change argument to you: the temporary change argument could apply to any organism, whether it is human (like you) or whether it is one of those “shumans” with ZNA instead of DNA in its cells. A strong moral presumption against killing any entity is generated by that entity’s possession of a set of typical human capacities. But this is precisely the claim we were seeking to prove at the beginning of this chapter: if an individual possesses a set of typical human capacities, there is a strong moral presumption against killing it.

There are two other important extensions of the results of the temporary change argument. Speaking for convenience again in terms of the capacity to think, these extensions concern the times of your life before you have the immediate capacity to think for the first time and after you have the immediate capacity to think for the last time. To see this, notice how two things follow directly from the fact that your possession of a higher-order capacity to think is sufficient for there to be a strong moral presumption against killing you. First, it follows that as long as you already have a higher-order capacity to think, there is already a strong moral presumption against killing you even before your very first moment of possessing the immediate capacity to think. Second, it follows that as long as you still have a higher-order capacity to think, there is still a strong moral presumption against killing you even after your very last moment of possessing the immediate capacity to think.

These two implications correspond exactly to the failures of the first and second strategies for replying to the temporary change. These two implications also correspond to the first two approaches, considered at the beginning of this section, for arguing that there is a strong
moral presumption against killing you even at times when you do not have the immediate capacity to think. Recall how the first approach emphasized the time of your infancy, and the second approach emphasized the time of your terminal psychological deterioration. As long as you have a higher-order capacity to think during these times, there is a strong moral presumption against killing you during these times.

But these other important extensions raise the same interesting questions that were raised at the end of Chapter Three. Just how far back in our biological life can we truly speak of ourselves possessing the higher-order capacity to think? Can we speak truthfully about having the higher-order capacity to think as infants, or fetuses, or embryos? Questions like these will be taken up in more detail in Chapter Four. Also, just what sorts of changes can we permit to happen to us before we deny that we still have the higher-order capacity to think? Can we speak truthfully about having the higher-order capacity to think after terrible brain diseases have taken their toll, or in individuals with severe genetic deficiencies? Questions like these will be taken up in more detail in Chapter Five.

2. An Alternative Account: Actual, Continuing Subjects of Experience

The view that has been defended in this chapter can be expressed as follows:

(C) If an entity has a set of typical human capacities, there is a strong moral presumption against killing it.

If the defense of (C) is to be successful, then one of the alternative views that needs to be refuted is the following one:

(P) The only entities that have a right to continued existence are actual, continuing subjects of experiences.
I would like to briefly explain (P) and highlight how (P) differs from (C). Then, I would like to explain why one of the reasons sometimes given for accepting (P)—namely, that it adequately handles thought-experiments that involving what may be called “complete reprogramming” of the upper brain—is actually a reason for favoring (C) over (P).

2.1. The Account Itself

Chances are, everyone reading this sentence is an actual, continuing subject of experiences. But to understand what it means to be an actual, continuing subject of experiences, it is important to emphasize that a continuing subject of experiences is different than a momentary subject of experience. Suppose that there is an organism A that can experience pleasure, but that has no thoughts at all, including no memories of past experiences, no intentions for obtaining future experiences, and so on. One moment, A is enjoying the pleasure of eating. The next moment, A is enjoying the pleasure of standing in the shade. The next moment, A is enjoying the pleasure of resting. One might be forgiven for thinking that A itself is a continuing subject of experiences. But this is one of the things that are denied by those who employ the concept of actual, continuing subjects of experience. According to them, there is no continuing subject of experiences and other mental states associated with A, but only a series of psychologically isolated momentary subjects of experience.

Suppose next that there is an organism B that has thoughts, apparent memories, and intentions regarding the future, but that has these mental states in a way that appears completely random from one moment to the next. One moment, B is beginning the thought “I think, therefore, I am.” But before B can finish this thought, B’s mental state suddenly switches to the momentary belief that it enjoyed visiting Beijing. The next moment, B’s mental state switches to
the intention to read Paul’s letter to the church at Rome. And so on. As with organism A, there is no continuing subject of experiences and other mental states associated with B, but only a series of psychologically isolated momentary subjects of experience.

When, then, does an organism have a continuing subject of experience associated with it? The short answer is: when the organism has certain sorts of mental states that are causally connected to one another in the right sorts of ways. And what this usually amounts to, for those who take this approach, is that the organism needs to have certain sorts of brain states that are causally connected to one another in the right sorts of ways. Of course, it is possible, in principle, to believe that some angels or immaterial minds could be, or could at least be associated with, actual, continuing subjects of experiences. But those who speak of actual, continuing subjects of experiences usually hold the view that experiences themselves, momentary subjects of experiences, and actual, continuing subjects of experiences all depend upon certain parts of the brain. The result is that, if an actual, continuing subject of experiences is associated with a given human organism, then once the relevant brain regions of that human organism are damaged or destroyed, the actual, continuing subject of experiences ceases to exist.

The view that actual, continuing subjects of experiences have a right to continued existence is able to handle some of the temporary change scenarios envisioned above at least as well as the view that possessing a set of higher-order typical human capacities suffices to give an entity serious moral status. After all, when you are asleep, anesthetized, or comatose, the relevant brain states that give rise to an actual, continuing subject of experiences presumably remain intact. However, some of the other scenarios envisaged above can only be seen as a direct challenge to the idea that only actual, continuing subjects of experiences have a right to continued existence. This is because, in order to be an actual, continuing subject of experiences,
it is necessary to have actually had some of the right sorts of experiences already, in the past, and to have those past experiences causally connected to other sorts of experiences in the right sorts of ways. So, for example, in the replica cases, since the original, but not the replica, has an actual, continuing subject of experiences associated with it, it follows that the original, but not the replica, has the right to life.

One of the more important things to emphasize in this context is this: the mere fact that a human organism has certain sorts of brain states does not by itself entail that the organism has an actual, continuing subject of experiences associated with it. Imagine that the replica cases envisioned above had not involved a temporary incapacitation of the original A: A gets replicated, let us say, while she is wide awake and enjoying her life very much. The replica produced, B, does not have an actual, continuing subject of experiences associated with it until B actually begins having the relevant sorts of experiences, and these experiences are causally related in the right sorts of ways. So even after B has her first experience (such as the thought “am I the original or am I the replica?”), this only means that a momentary subject of experiences is associated with B; a continuing subject of experiences does not exist until it can be built up out of at least two momentary subjects of experiences, causally related in the right sorts of ways.

2.2. Complete Reprogramming and Other Challenges to the Account

One of the reasons sometimes given for accepting (P) is that it handles the following sort of case very well, and indeed much better than its rivals:

Suppose…that there are technological developments that allow the brain of an adult human to be completely reprogrammed, so that the organism winds up with memories (or rather, apparent memories), beliefs, attitudes, and personality traits
completely different from those associated with it before it was subjected to reprogramming. (The pope is reprogrammed, say, on the model of David Hume.) In such a case, however beneficial the change might be, most people would surely want to say that someone had been destroyed, that an adult human being’s ‘right to life’ had been violated, even though no biological organism had been killed.³ Many people have the intuition that such “complete reprogramming” of a normal adult human organism’s brain is morally on a par with killing that human organism. (P) apparently explains these intuitions, but these intuitions receive no explanation at all given (C), since reprogramming neither kills any human organism, nor does it change the typical human capacities that the human organism in question has.

I believe that the most direct way of replying to this argument is to show that careful consideration of cases that involve complete reprogramming actually provides one with several reasons for favoring (C) over (P). I will now consider several related cases, explaining why (C) handles these cases better than (P). For convenience, I will speak in terms of a “right to life” instead of “serious moral status”, “a strong moral presumption against killing,” and “a right to continued existence”.

Case 1: Malicious Reprogram, No Experience Yet: Imagine a malicious reprogrammer reprograms the brain of George W. Bush at some time t so that it (the brain) gives rise at t* to the exact same types of beliefs and desires as John Kerry’s brain. But at t* the reprogrammed brain has not given rise to any actual experiences. (C) maintains that the human organism with this reprogrammed brain still has a right to life at t*. But (P) denies this. (P) is mistaken; the human organism does still have a right to life at t*.

³ Tooley, 102-3.
Case 2: *Malicious Reprogram, No Experience Yet, Benevolent Reprogram*. Case 2 is the same as Case 1, only now there is a benevolent reprogrammer who can reverse the effect of the malicious reprogrammer at t* so that the brain gives rise at t** to the exact same types of beliefs and desires as George W. Bush originally had at t. (C) maintains (1) that the human organism with the reprogrammed brain still has the right to life at t*, (2) that there is excellent reason to let the benevolent reprogrammer reprogram the brain at t*, (3) that the human organism with the re-reprogrammed brain still has the right to life at t**. But (P) denies all three of these claims, maintaining instead (1) that the human organism with the reprogrammed brain does not have the right to life at t*, (2) that it is a matter of indifference whether the benevolent reprogrammer reprograms the brain at t*, and (3) that the human organism with the re-reprogrammed brain still does not have the right to life at t**. But all three of these holdings of (P) are mistaken.

Case 3: *Malicious Reprogram, Experience, Benevolent Reprogram*. Case 3 is the same as Case 2, only now, let the human organism with the reprogrammed brain have some appropriately related experiences between t* and t+, which is the time at which the benevolent reprogrammer arrives on the scene. (C) maintains that there is excellent reason to let the benevolent reprogrammer reprogram the brain at t+. But (P) claims that it would be morally wrong for the benevolent reprogrammer to reprogram the brain at t+; this, (P) claims, would be morally akin to murder. But this is a mistake; reprogramming at t+ is not morally akin to murder: in fact, it is the morally proper thing to do, since it is simply reversing the effect of the malicious reprogrammer, and giving back to George W. Bush the distinctive personality which he had at t.

Case 4: *Double Malicious Reprogram, No Experience Yet*. Case 4 is the same as Case 1, only this time the malicious reprogrammer reprograms the brains of both George W. Bush and John Kerry at t so that they (the brains) give rise at t* to the exact same types of beliefs and
desires that the other person’s brain gave rise to at t. The original Bush brain gives rise at t* to
the beliefs and desires the original Kerry brain had at t; the original Kerry brain gives rise at t* to
the beliefs and desires the original Bush brain had at t. But at t* neither reprogrammed brain has
given rise to any actual experiences. (C) maintains that the human organisms with these
reprogrammed brains still have a right to life at t*. But (P) denies this. (P) is mistaken; the
human organisms do still have a right to life at t*.

We could easily construct additional cases that parallel Case 2 and Case 3 in the same
sorts of way Case 4 paralleled Case 1. A case paralleling Case 2 may be called Double Malicious
Reprogram, No Experience Yet, Benevolent Reprogram. The case paralleling Case 3 may be
called Double Malicious Reprogram, Experience, Benevolent Reprogram. But I think that it
would be more instructive to leave behind for the moment malicious reprogrammers bent on
altering the brains of George W. Bush and John Kerry, and focus instead on a case where a
human organism might really want to have its brain reprogrammed.

So, then, Case 5 may be called Repentant Man, Benevolent Reprogrammer: Imagine a
mature adult human organism comes to the conviction that most of his central beliefs about
himself and his place in the world are false, and that most of his deepest desires are wicked. He
wants to change his personality so that he becomes like Gandhi, but he realizes that the project of
such self-transformation will take him more time than his biological organism has left to live.
However, this man knows a reprogrammer who can reprogram his brain so that it gives rise to
the sorts of beliefs and desires that Gandhi had. (C) has no objection to this procedure: if the
repentant man asked the reprogrammer to do it, it would be morally right for the reprogrammer
to reprogram the repentant man in the way he asks. But (P) seems to take a different view of this
case. In the first place, (P) would insist that the repentant man is really requesting something
akin to annihilation, or assisted suicide. According to (P), whether the repentant man realizes it or not, “he” is going to be gone forever once the reprogramming happens. Perhaps (P) will not condemn the repentant man’s decision, since (P) takes no stand on whether an actual, continuing subject of experiences has the right to wave its right to continued existence. But (P) will insist that the repentant man does not continue to exist past the point of reprogramming. In the second place, however, according to (P), the human organism that does exist right after the reprogramming does not have the right to life until it actually begins having some experiences. But both of these implications of (P) seem strained.

Finally, it is worth noting that (C) is better able than (P) to handle other sorts of cases, which involve, not the reprogramming of a human organism’s brain, but the temporary destruction of a human organism’s upper brain. Case 6 may be called Voluntary Operation. Imagine a mature adult human organism finds out from his doctors that he has a degenerative disease in his upper brain that will eventually result in his biological death, and that the only way to prevent this from happening, since the disease has spread throughout his upper brain, is to agree to an operation that involves both destroying, and then reconstituting, the upper brain itself. One way this might work is as follows: a structural snapshot of his upper brain is ‘scanned out’ into a computer file, but the process of ‘scanning out’ necessarily involves the complete de-programming of the upper brain tissues themselves; the computer file is altered so that the parts of the file corresponding to the degenerative disease are removed; the altered structural snapshot is ‘scanned in’ to his de-programmed upper brain so that the brain has the exact same structure that it had at first, minus the disease. Another way this might work is this: a surgical vacuum cleaner is used to painlessly suck out and transport all the upper brain tissues from the man’s skull to a ‘cleaning tank’ where the diseased tissue is separated from the healthy tissue via a
process of high-speed spinning, and the healthy tissue is then put back into the man’s skull with the same structure as it had before, minus the damage done by the disease. In any event, (C) has no objection to these procedures: the man being cured is there at the beginning of the operation, there in the middle of the operation, and there at the end—and he retains his right to life throughout the operation. But (P) denies that the man is there in the middle of the operation, denies that the man is there at the end of the operation, denies that the human organism in the middle of the operation has the right to continued existence, and denies that the human organism at the end of the operation has the right to continued existence. (P) seems to me to be mistaken in all four of these denials.

The same sort of case can be constructed without the voluntariness of the patient, and without the disease to eliminate. Consider Case 7: Involuntary Experiment: Imagine the following alien abduction story. It is a “kinder, gentler” alien abduction story, partly because it all happens while the abductee is asleep, so there is no experiential suffering on his part. A kind and gentle alien removes, and completely destroys, the upper brain of a normal adult human organism, but leaves the rest of the organism alive. It puts all the parts of the upper brain into an alien experiment and totally disaggregates them, right down to the simplest atoms. But then the kind and gentle alien re-assembles those parts, creating a new upper brain, indistinguishable from the original one in every way: the new upper brain not only has the same structure as the original upper brain, but has, in addition, the exact same parts in all their exact old places: atom #4321 is next to atom #4322, just as before, and so on. Finally, the alien implants the ‘new’ upper brain where the ‘old’ upper brain was, and flies away. All this happens without waking up the human organism in question. According to (C), the man being experimented on is there at the beginning of the experiment, there in the middle of the experiment, and there at the end—and
he retains his right to life throughout the operation. But (P) denies that the man is there in the middle of the experiment, denies that the man is there at the end of the experiment, denies that the human organism in the middle of the experiment has the right to continued existence, and denies that the human organism at the end of the experiment has the right to continued existence. (P) seems to me to be mistaken in all four of these denials.

The basic points that emerge from these cases are as follows. Sometimes reprogramming a human organism’s brain is positively the right thing to do. Furthermore, the fact that there is not an “actual, continuing subject of experiences” associated with an organism does not mean that it is morally acceptable to kill that organism. Consequently, these cases support (C) over (P).

I believe a short summary of this chapter would be helpful at this point. Whereas the temporary change argument constructed in Chapter Three attempted to show that being human is sufficient for having a set of typical human capacities, the temporary change argument of this chapter attempted to show that having a set of typical human capacities is sufficient for having serious moral status. One might naturally ask: why not just skip the detour about typical human capacities, and instead construct a temporary change argument that attempts to show that being human is sufficient for having serious moral status? The answer to this natural question is that focusing upon typical human capacities allows us to explain more: it allows us to explain both why humans retain their serious moral status when undergoing various sorts of temporary changes, and why “shumans” (with capacities like ours, but with ZNA instead of DNA) retain their serious moral status when undergoing various sorts of temporary changes. Finally, I considered and rejected several alternative strategies for attempting to explain our serious moral status in a way that accommodates our beliefs about temporary changes.
CHAPTER FOUR: CAPACITIES AND THE ARGUMENT FROM POTENTIAL

Chapter Two began the defense of the first step of the main argument of this essay, that if an entity is human, it has a set of typical human capacities. Chapter Three began the defense of the second step of the main argument, that if an entity has a set of typical human capacities, it has serious moral status. This chapter builds upon this account by explaining the connection between capacities and the “Argument from Potential” (AFP). The next chapter explains the connection between capacities and the “Argument from Marginal Cases” (AMC). As we shall see below, the AMC and the AFP overlap in important respects.

1. The Argument from Potential and the Argument of This Chapter

The general strategy of the “Argument from Potential,” as its name suggests, is to argue from the claim that an entity has some potential or other to the claim that this entity has some moral status or other. The term “potential” here can mean a variety of things, depending on the philosopher using it, but at the very least this term typically means mere potential and thus implies “not actual”: for example, a potential “person” is typically not thought to be an actual person, and a potentially rational entity is typically not thought to be an actually rational entity.\(^1\) The AFP argues that the mere fact that an entity has a certain potential generates certain

\(^1\) Of course, there is nothing in the concept of potential that strictly demands that it be construed as “mere” potential: after all, an entity could be both actually \(P\) and potentially \(P\) at the same time. Indeed, as I argued in Chapter Two with David Hume and David Robinson, the fact that something is actually \(P\) usually is good evidence for the idea that it is potentially \(P\): if you ask me whether I have the potential to lift a one hundred pound barbell over my head, it usually settles the matter for me to reply “yes, of course I can—in fact, I’m doing it right now.”
obligations on the part of others to treat the entity in certain ways, whether or not that potential is now being actualized or has ever been actualized before.

Although this is the general strategy of the AFP, particular versions of it vary. At least three main things account for the differences between the versions. First, as just mentioned, different versions of the AFP employ different meanings of the term “potential”. This term can mean one or more of the following: bare logical possibility; probability or likelihood; an active causal propensity to change in certain ways; a passive receptivity to undergo certain sorts of changes.

Second, different versions of the AFP rely on different understandings of what the (given notion of) potential is said to be a potential for. Some focus on a given attribute: the potential for consciousness or sentience or rationality. Others focus on the potential for being a certain kind of thing: a potential “person” or a potential human being or a potential right-holder or a potential bearer of interests.

Third, a given version of the AFP is partly a function of just what moral relevance the (given notion of) potential is said to have. The consequentialist versions typically focus on the value of what a given entity’s potential is a potential for, and thus typically argue that any obligation to “respect” an entity’s potential grows out of a more basic obligation to promote certain valuable outcomes. The nonconsequentialist versions typically focus on the moral standing of the entity that possesses the potential, and thus typically argue that certain sorts of potential give a being certain sorts of interests (such as the interest in realizing its potential) or rights (such as a right to life).

The AFP is both important and controversial. The AFP is important because of its history and centrality in philosophical debates about the morality of abortion. Michael Tooley
has even claimed that when the conservative position on abortion “is thought through in a critical fashion, it appears to be the case that it [the conservative position] stands or falls with the answer to be given to the question of the moral status of potential persons.”

He reaches this claim by arguing that the moral status of the fetus is the central issue in the abortion debate and that attempts to defend the moral status of the fetus without relying on the concept of a potential person do not succeed. Similarly, Jim Stone has claimed that “a strong fetal claim to protection rises or falls with the appeal to the fetus’s potentiality, for nothing else can justify it.”

Although the AFP is also important to the moral evaluation of infanticide, contraception, and the use of human embryos in medical experiments, it originally emerged out of debates about abortion. This may explain, at least in part, why the AFP is so controversial: in particular, it is both widely criticized and widely loathed. As Michael Wreen once put it, “Potentiality is taking a bit of a beating in ethical circles these days, and the concept, it would be no exaggeration to say, currently enjoys about as much popularity as leprosy or the bubonic plague.” In addition to the fact that a variety of argumentative assaults have been carried out against the AFP, there is also the fact that the AFP can trigger some of the most visceral emotional responses among philosophers. As Elizabeth Harman recently put it, “I used to be

---

2 Tooley, 169.

3 For example, in section 3, pp. 40-49, Tooley argues against the idea that a woman’s right to control what goes on in her own body (an idea reflected, for instance, in Judith Jarvis Thomson’s argument from unplugging the violinist) is the central moral issue in the abortion debate.

4 For example, in section 4, pp. 50-86, Tooley argues against the moral relevance of species membership.


naively terrified of acknowledging any moral significance for potentiality.” 7 Now, lest these quotes mislead, it is worth noting that Wreen ends up arguing that potentiality is not vulnerable to most of the common assaults upon it, and Harman ends up arguing that her terror was premature. Indeed, her quote comes in her paper’s very last section, titled “How I Learned to Stop Worrying and Love Potentiality.” 8 Although Harman’s eventual endorsement of the qualified relevance of potentiality exemplifies the open-mindedness of some philosophers, these quotes still illustrate that a defender of the AFP has her work cut out for her.

It is not the purpose of this chapter to explain and evaluate all of the available versions of the AFP that have emerged in the considerable philosophical literature on this topic in the past 30 years. Rather, the purpose of this chapter is to explain how the main argument of this essay, which connects a thing’s higher-order capacities to the strong moral presumption against killing that thing, intersects with this literature.

This chapter argues that at least one version of the AFP can be defended by attending to an organism’s hierarchy of capacities. In particular, it will be argued that even the most undeveloped human organisms have the same set of general, higher-order, typical human capacities as you have. Since these typical human capacities are sufficient to generate a strong moral presumption against killing you, they are also sufficient to generate a strong moral presumption against killing these undeveloped human organisms. But since not all “precursors”

---


8 Ibid., 193.
to undeveloped human organisms have the same set of typical human capacities as you, it is an open question whether there is a strong moral presumption against killing these precursors.9

Two tasks are involved in explaining how capacities are relevant to the AFP. The first task is that of showing how it makes sense to think that undeveloped human organisms possess the capacities that generate a strong moral presumption against killing. But the second task is that of showing how one can agree with this first task without endorsing certain absurdities that the AFP is alleged to commit one to, and without embracing certain fallacies that the AFP is alleged to commit.

The following two sections correspond to these two tasks. Section 2 returns to the claim, presented briefly at the end of Chapter Three, that the present account implies that there is a strong moral presumption against killing human infants. In particular, 2.1 expands upon this claim in the context of normal human infants, and 2.2 explains why this claim can be extended to apply to normal human fetuses, embryos, and zygotes. Section 3 explains why the present account does not involve either absurdities or fallacies. In particular, 3.1 explains why the present account is not committed to the idea that there is a strong moral presumption against killing human precursors like gametes or body cells, and 3.2 explains why the present account does not commit the logical error of conflating the mere potential qualification for a moral status property (such as rights) with the actual possession of such a qualification.

---

9 The reason for saying not all precursors rather than no precursors is that, as we shall see below, in cases of embryonic fission and fusion, some precursors to undeveloped human organisms are themselves undeveloped human organisms.
2. Potential and Higher-Order Capacities

2.1. Normal Human Infants

In order to argue that human infants possess a set of capacities that generates a strong moral presumption against killing them, we could adopt a strategy that begins with the intuition that there is a strong moral presumption against killing human infants. We could then argue that, since this strong moral presumption must be based upon something, and since the only plausible thing that it could be based upon is the capacities of the infants, it follows that human infants possess a set of capacities that generates a strong moral presumption against killing them. An advantage of this strategy is that it appeals to an intuition that is widely shared by philosophers. But a disadvantage of this strategy is that it cuts no philosophical ice with those who do not share the intuition.\(^\text{10}\)

But there is another way to argue that human infants possess the relevant set of capacities. Towards the end of Chapter Three, it was briefly argued that one of the implications of a proper understanding of the problem of the temporary change was that there is a strong moral presumption against killing human infants. However, the argument still left unanswered at least two important questions:

Q1. Why should someone believe that a human infant actually has any higher-order capacity to think at all?

\(^{10}\)Michael Tooley argued, in chapter 10 of *Abortion and Infanticide*, that the intuition should be set aside for at least three reasons. First, appeals to moral intuitions are plausible only if the appeal is to principles that are (what Tooley calls) basic moral principles (or derivable from basic moral principles), but the principle appealed to by this intuition is not. Second, the intuition in question is not unanimously shared in our own society and was not generally shared by earlier societies. Third, the intuitions of people over the past 2000 years have been heavily shaped by the religions of Judaism and (especially) Christianity, so that a person can reasonably rely upon such intuitions only if he takes the teachings of one of these religions to be true.
Q2. Why should someone believe that a human infant possesses the sort of higher-order capacity to think that generates a strong moral presumption against killing its possessor?

Those asking Q1 could agree with the claim that your possession of a higher-order capacity to think is sufficient for there to be a strong moral presumption against killing you, even before your very first moment of possessing the immediate capacity to think. But they could disagree with the claim that you possessed this higher-order capacity when you were an infant, on the grounds that you never were an infant. Perhaps the commonsense idea that you once were a human infant is simply false. After all, this commonsense idea is flatly incompatible with certain accounts of what constitutes your identity over time. Those asking Q1 might doubt that the infant from which you developed had your set of higher-order capacities because they doubt that the infant from which you developed was you. (For example, someone attracted to the view that “you” refers to a continuing subject of experiences could argue, first, that if one wants to give conditions for identity over time in the case of a continuing subject of experiences in terms of higher-order capacities, the higher-order capacities in question cannot be purely general higher-order capacities, and second, that the fetus that ultimately gives rise to a given continuing subject of experiences does not possess the non-general, specific, relational capacities that are essential to the identity over time of the relevant subject of experiences.)

Those asking Q2 could agree with the claim that your possession of a higher-order capacity to think is sufficient for there to be a strong moral presumption against killing you, even before your very first moment of possessing the immediate capacity to think. But they could disagree with the claim that you possessed this higher-order capacity when you were an infant, on the grounds that the order of this capacity when you were an infant was too high to generate a
strong moral presumption against killing. Perhaps there is an order in the hierarchy of capacities—perhaps the 3rd or 4th order, perhaps the 30th or 40th—where the strong moral presumption against killing is no longer generated. After all, even if it is true that you once were an infant, and even if it is true that a strong moral presumption against killing you is generated by the higher-order capacity to think you possess *when sleeping*, it still needs to be established that this presumption is generated by the higher-order capacity to think you possessed *when an infant*. For the capacity to think you possessed when an infant is of a different and indeed much higher order than the capacity to think you possess when sleeping. Those asking Q2 might believe that *some* higher-order capacities to think generate the presumption, while *other* higher-order capacities to think do not.

These questions will now be answered by a further argument, whose basic thrust can be summarized in this paragraph and whose premises can be set out and defended in the next few paragraphs. Recall that the temporary change argument showed that for there to be a strong moral presumption against killing an entity, it is sufficient for that entity to possess a certain higher-order capacity to have the immediate capacity to think. Previous possession of the immediate capacity to think was seen to be irrelevant. But as we imagine the temporary change getting more and more serious, the order of the capacities we must appeal to gets higher and higher. Since previous possession of the immediate capacity to think is irrelevant, it follows that as we heighten the order of capacities sufficient to generate a strong moral presumption against killing an individual *who has possessed but then lost* the immediate capacity to think, we thereby heighten the order of capacities sufficient to generate a strong moral presumption against killing an individual *who has not already possessed* the immediate capacity to think. Eventually, we reach a point where the order of the capacity to think possessed by the individual in the middle of
a temporary change is the same as the order of the capacity to think possessed by an infant: for example, just as an infant only has a (say) 100\textsuperscript{th}-order capacity to think, so too we can envisage an adult in the middle of a temporary change who only has a 100\textsuperscript{th}-order capacity to think. But since possession of this capacity is enough to generate a strong moral presumption against killing the adult, possession of this capacity is also enough to generate a strong moral presumption against killing the infant.

More formally, the argument can be set out as follows:

(1) Some adults are the same, in terms of the order of their capacity to think, as some infants.

(2) For the adults in premise (1), having their particular order of the capacity to think is sufficient to generate a strong moral presumption against killing them.

(3) For any individuals x and y, for any activity A, and for any order N of the capacity to A, if x’s having the N-th order capacity to A is sufficient to generate a strong moral presumption against killing x, then y’s having the N-th order capacity to A is sufficient to generate a strong moral presumption against killing y.

Therefore,

(4) For the infants in premise (1), having their particular order of the capacity to think is sufficient to generate a strong moral presumption against killing them.

It seems that this argument is valid; acceptance of premises (1), (2), and (3) requires acceptance of (4). But is this argument sound? Are its premises true? Premise (1) is perhaps the most difficult premise to defend. But a defense of premise (1) will answer Q1 above: for if you could become one of the adults in premise (1), then you could have been one of the infants in premise
(1). Premise (2), since it depends on premise (1), inherits any of the difficulties of premise (1).

But a defense of premise (2) will answer Q2 above: for if the sort of higher-order capacity possessed by the adults generates a strong moral presumption against killing them, then, as long as premise (3) is true, the same sort of higher-order capacity will generate a strong moral presumption against killing the infants. Premise (3), although a bit technical, is carefully formulated so as to be as uncontroversial as possible. Hence a separate defense of premise (3) will not be given.

A natural objection can be raised against premise (1). Admittedly, this objection goes, it is not hard to think of cases involving adults who do not have the immediate capacity to think: for example, a sleeping adult must first wake up before she can truly be said to have this immediate capacity. But in our world, this objection continues, most adults who lack this immediate capacity need very little in order to regain it: for example, most adults when sleeping have a 2nd-order, or at most a 3rd-order, capacity to think. Comparable things can be said if the adults without the immediate capacity to think are drunk, or sedated, or under anesthesia. Nevertheless, the objection goes, it is surely more difficult to think of cases involving an adult whose order of the capacity to think is even remotely comparable to, much less the same as, the order of the capacity to think possessed by a human infant. The objection concludes that premise (1) must be a long reach at best and simply mistaken at worst.

But this natural objection can be answered by focusing on cases of human adults whose present inability to think is more serious than the sort of “inability” to think brought on by sleep, alcohol, or anesthesia. Take, for example, the case of a normal human adult A that suffers such a severe form of brain damage that he must go through a long period of rehabilitation in order to regain the immediate capacity to think. To fix ideas, let A be almost exactly like Thomas
Nagel’s example of “an intelligent person [who] receives a brain injury that reduces him to the mental condition of a contented infant…[for whom] happiness consists in a full stomach and a dry diaper.” The only difference between A and Nagel’s example is that, while it is an open question whether Nagel’s brain-damaged adult can ever outgrow his unfortunate condition, it is certain that A can. A can be fully rehabilitated over time. To grasp how much time it will take for A to recover, imagine that, at the time of his brain damage, A was the parent of an infant B. A’s brain damage is so severe that it will take A the same amount of time to regain his immediate capacity to think as it takes B to obtain that immediate capacity for the first time. It seems that the order of A’s capacity to think is as high as the order of B’s capacity to think. If this claim is doubted, the concept of “developmental distance” introduced in Chapter Two can be used to back it up. The way A and B have been described, A and B are the same developmental distance, in terms of length of time, away from having the immediate capacity to think. But it is not hard to imagine A and B being the same developmental distance, in terms of developmental steps, away from having the immediate capacity to think. But to imagine this is to imagine A and B having the same order of the capacity to think, since each developmental step maps directly on to an order of the hierarchy of capacities leading up to the first-order capacity to think. The case of A therefore supports premise (1) in the above argument.

Once premise (1) has been defended, it is not hard to see that premise (2) can be defended as well. Surely, there is still a strong moral presumption against killing A during the period of his temporary change. Otherwise, certain sorts of temporary injuries or setbacks can cause the strong moral presumption against killing you to disappear. Furthermore, it can be shown (via the arguments of Chapter Three) that this presumption is generated by A’s possession of his higher-

---

order capacity to think, and not, for example, by the fact that A once possessed the immediate
capacity to think. This case therefore supports premise (2) in the above argument.

2.2. Beyond Normal Human Infants

It might be thought at this point that, although this strategy works for some human
infants, it cannot work for all of them, and in particular it cannot work for abnormal infants
whose abnormality consists of a brain defect that prevents the infant from developing the
immediate capacity to think. Since this thought will be carefully evaluated, and eventually
rejected, in the next chapter, there is no need to evaluate it in this sub-section. Instead, the focus
of this sub-section is on a different question. Can the argument from the temporary change be
extended to human organisms less developed than human infants? This sub-section argues that it
can. Although there might be skeptical concerns that parallel those embodied by Q1 and Q2, it
will now be argued that the argument sketched above would still be sound even if (1) and (4)
were amended by replacing the word ‘infants’ with the word ‘fetuses’ or ‘embryos’ or ‘zygotes’.

The basic idea is that, as we imaginatively alter the temporary changes of the adult
organisms, we eventually reach a point where the order of the capacity to think possessed by the
organism in the middle of a temporary change is the same as the order of the capacity to think
possessed by a human fetus, or embryo, or zygote. For example, just as a fetus only has a (say)
1000th-order capacity to think, so too we can envisage an adult in the middle of a temporary
change who only has a 1000th-order capacity to think. But since possession of this capacity is
enough to generate a strong moral presumption against killing the adult, possession of this
capacity is also enough to generate a strong moral presumption against killing the fetus. And the
same sorts of considerations can be extended to a human embryo or zygote.
A defense of an amended premise (1) can be offered along the same lines as the defense of the original premise (1). Consider the case of a normal human adult organism C that suffers such a severe form of brain damage that he must go through a long period of rehabilitation in order to regain the immediate capacity to think. As before, C can be fully rehabilitated over time. To grasp how much time it will take for C to recover, imagine that, at the time of his brain damage, C was the parent of a fetus D. C’s brain damage is so severe that it will take C the same amount of time to regain his immediate capacity to think as it takes D to obtain that immediate capacity for the first time. It seems that the order of C’s capacity to think is as high as the order of D’s capacity to think. Once again, this type of claim can be backed up with the concept of “developmental distance” that was introduced in Chapter Two. The case of C therefore supports premise (1) in the above argument, even when that argument is amended to refer to fetuses.

Once the amended premise (1) has been defended, it is not hard to see that premise (2) can be defended as well. Surely, there is still a strong moral presumption against killing C during the period of his temporary change. Otherwise, certain sorts of temporary injuries or setbacks can cause the strong moral presumption against killing you to disappear. Furthermore, it can be shown (via the arguments of Chapter Three) that this presumption is generated by C’s possession of his higher-order capacity to think, and not, for example, by the fact that C once possessed the immediate capacity to think. The case of C therefore supports premise (2) in the above argument, even when that argument is amended to refer to fetuses.

Now precisely the same sorts of defenses can be constructed when the original argument is amended to refer to embryos or zygotes. All one needs to do is to construct cases involving a human adult organism E and his embryonic offspring F, or a human adult organism G and his zygotic offspring H. Rather than spell these amendments out in repetitive detail, the remainder
of this section will focus upon a few of the stronger objections to the line of argument just developed.

2.3. Objections and Replies

2.3.1. You Can Only Survive Certain Sorts of Upper Brain Damage: Tooley

Perhaps the most direct objection to the argument developed above, whether one is thinking about infants, fetuses, or what have you, is an objection that emerges from chapters 5 and 6 of Michael Tooley’s *Abortion and Infanticide*. In chapter 5, Tooley argues that a *person* is an entity that possesses at least one of the relatively permanent, non-potential properties that make it intrinsically wrong to destroy that entity, and that do so independently of that entity’s intrinsic value.\(^{12}\) His eventual account of these properties has the important implication that an individual can be a human organism at a certain time without being a person at that time. Then, in chapter 6, Tooley argues that:

\[
X \text{ is a potential person if and only if X has all, or almost all, of the properties of a positive sort that together would be causally sufficient to bring it about that X gives rise to a person, and there are no factors present within X that would block the causal process in question.}^{13}
\]

The relevant objection to the argument of the present chapter is found towards the middle of Tooley’s chapter 6. He says that the only argument he has encountered in support of the claim that “the destruction of potential persons is intrinsically wrong”, aside from arguments showing

---

\(^{12}\) Tooley, 87.

\(^{13}\) Ibid., 168.
that there is a prima-facie obligation to actualize “possible” persons, is an argument that runs as follows:

Consider a normal adult human being that is in a coma due to brain damage. Unless the term ‘capacity’ is used loosely, we cannot speak of such an individual as having capacities for self-consciousness, for rational thought, etc. It has only the potentiality of re-acquiring those capacities. Nevertheless, one certainly does not think that it is morally permissible to kill such an individual. If it is possible for the individual to recover, it is just as wrong to kill him as it is to kill a normal adult human being who is not in a coma. So potentialities do have moral weight. And if they count in this context, if their presence serves to make it seriously wrong to destroy something, why should they not count equally in the case of potential persons?\(^\text{14}\)

This argument is quite similar to the argument developed above. Although this argument involves a coma while the argument developed above did not, this difference is not important for present purposes.

Tooley’s case against this argument consists of two points. His first point is that if the mentioned potentialities were sufficient to make it wrong to kill the organism that is in a coma, then they must be understood as “passive” potentialities in cases where brain damage can only be repaired by others. But this is too broad, Tooley argues, since it entails that it is wrong to destroy living human skin cells, which, because of cloning technologies, have the relevant “passive” potentialities too. Tooley’s concern here will be dealt with in a more specific way regarding cloning below. The basic idea is that allowing passive potentialities to bear moral weight is

\(^{14}\) Ibid., 203.
unproblematic as long as the distinction between identity and material constitution is upheld. The passive potentiality \( a \) has to become \( b \) in an identity-preserving way, and the passive potentiality \( a \)'s matter has to constitute \( b \), are different sorts of potentialities. A living human adult in a profound coma has the former sort; a living human skin cell has only the latter sort.

Tooley’s second point is that the mentioned potentialities are not sufficient to make it wrong to kill the organism that is in a coma. His reasoning here is worth quoting in full:

For the injury may have deprogrammed the organism’s brain, with the result that while it will shortly revive, and enjoy self-consciousness, etc., it will not have any of the memories, beliefs, attitudes, personality traits, and so on, of the person previously associated with that human organism. An organism may continue to possess certain general potentialities after the associated person has been destroyed. In addition to general potentialities, there must be states of certain sorts standing in appropriate causal relations to corresponding earlier states—the types of states being those upon which personal identity depends. Accordingly, it is simply not true that general potentialities suffice to make it wrong to destroy a human organism in a coma, even though the organism was once a person. And if general potentialities are not sufficient in this case, the present argument has failed to provide one with any grounds for thinking that general potentialities suffice to make it wrong to destroy organisms that are not yet persons.\(^\text{15}\)

It is worth pausing for a moment to ponder the implications of what Tooley is claiming in this second point. Tooley had written at the outset of *Abortion and Infanticide* that “it is very difficult indeed to arrive at a defensible position on abortion unless one is prepared to come to

\(^{15}\) Ibid., 204-205.
terms with the difficult issue of the moral status of infanticide." What the present passage illustrates is that it is difficult to arrive at a defensible position on either abortion or infanticide unless one is prepared to come to terms with the issue of the moral status of killing human organisms that undergo certain sorts of temporary changes. Tooley’s reasoning here confirms one of this chapter’s main thrusts: the strong moral presumption against killing certain adult human organisms who undergo specific types of temporary changes, and the strong moral presumption against killing certain undeveloped human organisms, stand or fall together.

Tooley argues that both presumptions fall together. The reason is that, in the case that he is discussing, the states that he thinks are the basis for personal identity have been completely destroyed, so that the person who existed previously no longer exists, even though the same human organism exists. Tooley would then be able to claim, when confronted by a “temporary change” argument, that the person (“you”) no longer exists once the relevant change has occurred. The change in question, on his view, is not temporary but permanent, because it destroys or annihilates a person who can never come back.

However, Tooley’s reasoning can be challenged by challenging his claim that “In addition to general potentialities, there must be states of certain sorts standing in appropriate causal relations to corresponding earlier states—the types of states being those upon which personal identity depends.” Imagine that a normal adult human organism suffers an accident that causes brain damage so serious that the supposedly relevant causal links for personal identity through time are severed, even though the “general potentialities” remain. However, imagine further that the organism eventually comes to possess the very same types of brain states as it possessed before: the organism’s favorite symphony before the accident eventually becomes its

16 Ibid., 2.
favorite symphony after the recovery; the scent of chocolate which the organism found so exhilarating before the accident eventually becomes just as exhilarating after the recovery; the friendships, the religious commitments, the idiosyncrasies of belief and behavior all eventually become the same after the recovery as they were before the accident. The organism, as it were, lives the same bits of its life over again, but experiences those bits as though it were experiencing them for the very first time. Now, in a case such as this, is the most natural interpretation that the original person—Jim, let’s say—ceases to exist, and that a new person, indistinguishable in every way from Jim, has come to be associated with the same organism that Jim was associated with? Or is the most natural interpretation that the same person—namely, Jim—has recovered his original personality traits? It seems that the latter interpretation is preferable, and thus the causal links Tooley suggests are necessary for personal identity through time turn out to be unnecessary. For the case was described in such a way that, even though it is precisely those causal links that have been severed, the same person nevertheless remains.

I believe that there are at least two additional reasons, which even someone sympathetic with Tooley’s approach might be able to accept, for believing that it is the same person—namely, Jim—that exists after the organism’s recovery. Both of these reasons emerge from other passages in *Abortion and Infanticide*. First, when Tooley considers the question “does an individual have to have desires at a given time in order for it to be the case that some things are in its interest?”17, his answer is “No” for the following reason:

Suppose that there were a disease that completely eradicated all desires in normal adult human beings, but only for a time. If John has contracted the disease, and is now in a desireless state, but will return to normal, and will then enjoy a life that

17 Ibid., 117.
he deems worth living, one surely wants to say that John’s continued existence is in his interest. So individuals can have interests at a given time without having any desires at that time.\textsuperscript{18}

This first passage is relevant because it describes a case in which John undergoes what can only be described as a temporary change in all of his desires. John has a certain set of desires at $t_1$. John exists in a desireless state at $t_2$. John regains his desires at $t_3$.

Now, desires are among the traits that give personalities their distinctive shape and structure. John’s desires are arguably just as central and important to John as his memories, or beliefs, or attitudes. Thus if John can have interests even while he is undergoing a temporary change with respect to his desires, then it seems like he could have interests even while he is undergoing a temporary change with respect to his memories, or beliefs, or attitudes. Suppose that there were a disease that completely eradicated all memories (and/or beliefs, and/or attitudes) in normal adult human beings, but only for a time. If John has contracted the disease, and is now in a memory-less (and/or belief-less, and/or attitude-less) state, but will return to normal, we surely want to say that John’s continued existence is in his interest. But for something to be in John’s interest at a given time, it must be the case that John exists at that time. Therefore, it seems that John can still exist even if he is now in a memory-less (and/or belief-less, and/or attitude-less) state.

The second relevant passage comes in a section of the book whose aim is to lay out a case against the idea that something’s capacities determine whether or not it is a person. Part of this case consists of the claim that possession of certain capacities is not \textit{necessary} to make something a person. Tooley considers two types of brain damage that make it “impossible for

\textsuperscript{18} Ibid.
the organism to enjoy any consciousness at all”: in the first type, “the damage might involve the complete destruction of those structures that are the positive, constitutional basis of consciousness and rational awareness”\textsuperscript{19}; in the second type, the damage might leave those structures intact, but damage other parts of the brain so that the structures in question are isolated, with the result that it is impossible for the capacities for consciousness, and for rational awareness, to be exercised as long as the damage goes unrepaired.\textsuperscript{20}

Tooley claims that even in the first type of brain damage, it is “unacceptable” to conclude that “if one were to kill the organism, or allow it to die, one would not be guilty of having destroyed a person.”\textsuperscript{21} The “crucial point,” he says, is that it might very well be possible to repair such damage, and that the result of doing so might be an organism that not only was capable of rational awareness, but that had the memories, beliefs, attitudes, personality traits, and so on, characteristic of the person who previously existed.\textsuperscript{22}

Tooley then asks a question: “Would the resulting individual be identical with the individual who existed prior to the damage, or merely be a replica?” And the answer he gives is this: The view that he would be a replica does not seem plausible. A person who revives from a coma is not a replica of the person who existed previously. Yet a coma may very well involve brain damage that temporarily destroys the

\textsuperscript{19} Ibid., 153.

\textsuperscript{20} Ibid.

\textsuperscript{21} Ibid.

\textsuperscript{22} Ibid.
The possibility Tooley is envisioning here is similar in certain respects to the case of Jim presented above. And the interpretation Tooley gives to this possibility is very similar to the interpretation I have suggested above in the case of Jim. Tooley’s idea is that even in cases where brain damage destroys the constitutional basis for rational awareness, the resulting person is identical with the person who existed before the damage occurred. The idea I proposed with Jim was that even in cases where brain damage destroys certain causal connections, the resulting person is identical with the person who existed before the damage occurred.

The remainder of this second relevant passage consists of the conclusion Tooley draws from the possibility he envisions, where brain damage destroys the constitutional basis for rational awareness:

Once it is granted that the resulting person is identical with the person who existed previously, the argument can be put as follows. If something can be a person only if it possesses a capacity for rational awareness, then it follows that in destroying an organism that has suffered the sort of damage described above, one cannot be destroying any person. But if the damage is repairable, and if the result would be the revival of the person who previously existed, then the destruction of the organism, by making impossible any such revival, thereby destroys the person.

---

23 Ibid.
in question. Hence it cannot be the case that there is a person only where there is
a capacity for rational awareness.\footnote{Ibid., 153-154.}

This remainder of the second passage contains an important ambiguity. Although Tooley does
admit that the person after the recovery is the same as the person before the injury, it is not clear
that he wishes to commit himself in this passage to the view that there is any person, much less
the same person, which presently exists during the time between the injury and the recovery.
Nothing Tooley has said commits him to the view that the person endures through this period of
time, and there may be signs that he does not hold this view: for example, notice how he uses the
phrase ‘the revival of the person who previously existed’ rather than the phrase ‘the revival of
the person who presently exists’. Perhaps Tooley believes that the person does not persist
through this period of time, but rather, as it were, vanishes when the injury occurs only to
reappear when the recovery occurs. If so, then his view about destroying this person requires the
assumption that someone can destroy a person at time t even if that person does not presently
exist at t. But this is a rather difficult assumption to make. Imagine that we could revive
(resurrect?) Socrates by waving a certain magic wand, and that this was the only way to revive
(resurrect?) him. The assumption we are considering implies that, if we destroy the wand, it
follows that, since we have made impossible the revival of Socrates, we have thereby destroyed
Socrates. But this seems mistaken. After all, Socrates was already destroyed, by hemlock. Our
destroying the wand does not destroy Socrates a second time. Since attributing this assumption
to Tooley is not the most charitable way of reading him, the discussion to follow will proceed as
though he does not hold that assumption, and that he does hold the view that the person persists
through the period of time between the injury and the recovery.
If the person does indeed persist through the period of time between the injury and the recovery, and if this is the basis for Tooley’s argument that the capacity for rational awareness is not necessary for being a person, then there is an important qualification that needs to be made about the conclusion of his argument. Although the argument does seem to show that what we have called the “immediate” capacity for rational awareness is not necessary to be a person, it seems to show this by relying on what we have called a higher-order capacity for rational awareness: the organism Tooley describes has undergone a temporary change in its immediate capacity for rational awareness, but it still has a higher-order capacity for rational awareness, as is shown by the possibility of recovery. Therefore, Tooley’s argument in this second passage does not show that what we have called a higher-order capacity for rational awareness is not necessary to be a person. It only shows that what we have called an immediate capacity for rational awareness is not necessary to be a person.

When this second passage is considered alongside the first passage, it appears that one should not make a metaphysical distinction between cases where the brain damage to be repaired involves the constitutional basis for rational awareness, and cases where the brain damage to be repaired involves the constitutional basis for specific desires of the person, as in the case of John. That is, just as persons can persist even though the constitutional basis for their rational awareness has been temporarily destroyed, so too persons can persist even though the constitutional basis for their desires has been temporarily destroyed.

But once it is admitted that persons can persist even when both the constitutional basis for their desires and the constitutional basis for their rational awareness have been temporarily destroyed, it seems hard to deny that persons can persist even when the constitutional basis for the causal links connecting earlier mental states to later mental states has been destroyed. Recall
the earlier example (from Chapter Three) where a human adult has a certain sort of brain disease, and the only way of treating this disease requires the temporary destruction of all these supposedly relevant causal connections. If you were this human adult organism, is it not quite coherent to envision yourself surviving this operation?

In short, then, I believe that these two passages from *Abortion and Infanticide* actually help to motivate the interpretation I have been urging in the case of Jim. Jim himself, and not a mere replica, exists even after the supposedly relevant causal links have been severed. Hence these causal links are not necessary for personal identity.

Once we realize that these causal links are not necessary for personal identity, in situations where the “memories, beliefs, attitudes, personality traits, and so on” after the recovery are *the same as* before the accident, we also realize that these causal links are not necessary for personal identity, even in situations where the “memories, beliefs, attitudes, personality traits, and so on” after the recovery are *different than* before the accident. It might be objected at this point that a significant similarity of personality traits are necessary for personal identity through time, even if causal links are not necessary. But this is not true. Imagine that it takes 10 years for Jim to recover his personality traits. Surely Jim himself exists during this recovery period: after all, it is his recovery. But during this recovery period, Jim’s organism does not have most of the personality traits it had before the accident. In short, Jim still exists even when these personality traits do not. Therefore these personality traits are not necessary for Jim’s personal identity. Since Jim can still exist *during* a recovery phase without these personality traits, why should we doubt that Jim can still exist *after* a recovery phase without these personality traits?
2.3.2. You Could Not Have Been an Embryo: McMahan

Some philosophers argue that, even if it makes sense to think of your identity stretching back to one of the later stages of your mother’s pregnancy, when you were a fetus, it does not make sense to think of your identity stretching back farther than this, when even the slightest traces of a mental life are absent. Consider, for example, the following thought experiment from Jeff McMahan’s book *The Ethics of Killing*. McMahan attempts to cast doubt on the idea that “one was a tiny cluster of cells” by asking us to consider “whether one could ever become such an entity”:

Imagine that in some of us the process of biological development were somehow reversed. Those to whom this happened would begin to grow younger, in biological terms. Eventually they would revert to being babies and thereafter would have to be placed in artificial wombs in order to survive. As their brains reverted to the infantile and fetal stages of their development, their mental lives would become increasingly rudimentary and would eventually disappear altogether when their brains ceased to be capable of supporting consciousness. Suppose now that one were to face this prospect. It is instructive to ask oneself when in this process of biological regression one would cease to exist.\(^{25}\)

After presenting the example, McMahan claims that

For my part, I find it impossible to believe that I would still be around when what we may neutrally designate as my organism had been reduced to a microscopic network of cells from which any possibility of consciousness had vanished.\(^{26}\)

\(^{25}\) McMahan, 29.

\(^{26}\) Ibid.
Modifying his example so that it involves a temporary change can challenge McMahan’s intuition. Imagine that, in addition to our biologically based cycles of falling asleep and falling awake, human organisms also had biologically based cycles of shrinking down, in a manner like McMahan suggests, and growing back up, in which the steps of shrinking are reversed. To keep other aspects of our existence constant in this modified example, imagine also that, just as there is always psychological similarity between oneself before and after a period of being asleep, so too there is always psychological similarity between oneself before and after a period of being shrunk. If this were how we lived out our biological lives, it would not only be possible to believe that we still existed during one of these shrunken moments, but it would be quite likely that we would be unable to shake this belief, just as we are unable to shake the belief that we still exist during one of our sleeping moments. And since this is so, it should not be impossible to believe that we exist during the first of such shrunken moments.

2.3.3. You Could Not Have Been a Zygote: Singer and Kuhse

Some philosophers argue that, while it may make sense to think of your history stretching back to your embryonic stage, it makes no sense to think of your history stretching back prior to the formation of the “primitive streak” at approximately 14 days after fertilization. The reason the moment of primitive streak formation is thought to be so important is that it is considered to be the moment when three types of identity-undercutting possibilities are closed off: the possibility of fission (e.g. twinning), the possibility of fusion (e.g. the production of chimeras), and the possibility of separating out a totipotent cell from an “early human embryo” (in other words, from a morula or blastocyst).
These three possibilities are sometimes invoked to undermine the claim that genetic similarity and/or causal continuity are adequate bases for organismic identity through time, or to undermine the claim that morulas and blastocysts are even bona fide individuals. But to see how these three possibilities are invoked to undermine the claim that, just as you were once an infant, so too you were once a morula or blastocyst, consider a trio of thought experiments from Peter Singer and Helga Kuhse. With the possibility of fission, the thought experiment is this:

A man and a woman have intercourse, fertilization takes place, and a genetically new zygote, let’s call it Tom, is formed. Tom has a specific genetic identity—a genetic blueprint—that will be repeated in every cell once the first cell begins to split, first into two, then into four cells, and so on. On day 8, however, the group of cells which is Tom divides into two separate identical cell groups. These two separate cell groups continue to develop and, some nine months later, identical twins are born. Now, which one, if either of them, is Tom? There are no obvious grounds for thinking of one of the twins as Tom and the other as Not-Tom; the twinning process is quite symmetrical and both twins have the same genetic blueprint as the original Tom. But to suggest that both of them are Tom does, of course, conflict with numerical continuity: there was one zygote and now there are two babies.27

With the possibility of fusion, the thought experiment is this:

A man and a woman have intercourse and fertilization takes place. But this time, two eggs are fertilized and two zygotes come into existence—Mary and Jane. The zygotes begin to divide, first into two, then into four cells, and so on. But,

then, on day 6, the two embryos combine, forming what is known as a chimera, and continue to develop as a single organism, which will eventually become a baby. Now, who is the baby—Mary or Jane, both Mary and Jane, or somebody else—Nancy?28

With the possibility of separating out a totipotent cell from a morula or blastocyst, the thought experiment is this:

It is now believed that early embryonic cells are totipotent; that is, that, contrary to the ‘identity thesis’, an early human embryo is not one particular individual, but rather has the potential to become one or more different individuals. Up to the 8-cell stage, each single embryonic cell is a distinct entity in the sense that there is no fusion between the individual cells; rather, the embryo is a loose collection of distinct cells, held together by the zona pellucida, the outer membrane of the egg. Animal studies on four-cell embryos indicate that each one of these cells has the potential to produce at least one fetus or baby.29

These three phenomena are sometimes invoked in discussions of the metaphysics of early human development, as a way of showing that no human babies can trace their histories back prior to the moment when the primitive streak is formed. But these phenomena do not show this, since even in cases where fission, fusion, and/or totipotent separation occur, the relevant phenomena occur prior to the formation of the primitive streak. It would seem that the most these phenomena could ever show is that some babies cannot trace their histories all the way

---

28 Ibid., 70.

29 Ibid., 71.
back to a moment of fertilization, but must be content to trace their histories back to a moment of fission, or a moment of fusion, or a moment where a totipotent cell is separated out.

Nevertheless, many argue that these three phenomena show much more than this. Their basic argument is this:

(1) No human individuals can trace their histories back prior to the time when events like fission, fusion, and totipotent cell separation can occur.

(2) Such events can occur up until the time when the primitive streak is formed.

Therefore,

(3) No human individuals can trace their histories back prior to the time when the primitive streak is formed.

But this argument is defective: if we suspend belief for a moment about the second premise, we will quickly realize that the first premise is false. Even if it was possible for me, 10 years ago, to undergo the sort of “brain bisection” described by Thomas Nagel,\(^\text{30}\) or the sort of “division” described by Derek Parfit,\(^\text{31}\) this possibility of fission would not undercut the claim that I can trace my history back farther than 10 years.\(^\text{32}\) The same thing is true with fusion. Even if it were possible for you and I, 10 years ago, to undergo the sort of “brain fusion” described by science fiction novels, this possibility of fusion would not undercut the claim that you and I can trace our

\(^{30}\) Nagel, 147-164.

\(^{31}\) Parfit, 245-248.

\(^{32}\) This reply goes back at least as far as when Michael Lockwood considered Mary Warnock’s statement in a television interview that “before fourteen days the embryo hasn’t yet decided how many people it is going to be.” He writes: “It is, however, independently clear that she [Warnock] is not personally inclined to set much store by that fact, morally speaking. And rightly so, in my opinion: any philosopher who took that as grounds for denying that early human embryos are human beings would, it seems to me, have to deny that ordinary adult human organisms are human beings, if Sperry and others turned out to be correct in claiming that the effect of cutting the corpus callosum, which links the two hemispheres of the cerebral cortex, is to transform one human being into two.” See “Warnock Versus Power (And Harradine): When Does Potentiality Count?” Bioethics 2, no. 3 (1998): 190.
own individual histories back farther than 10 years. And the same thing is true with the separation of a totipotent cell. Even if it was possible for me, 10 years ago, to create a new human individual by merely separating one of my body cells from the rest of my body, this possibility does not undercut the claim that I can trace my history back farther than 10 years. In short, when it comes to adult human beings, like you and I, the mere logical possibility of fission, fusion, or totipotent separation at some time \( t \) does not undercut the claim that the adults can trace their identity through time back before \( t \). Why should it be any different with early embryonic human beings, like morulas and blastocysts?

At this point, some philosophers fall back upon the claim that morulas and blastocysts are not yet genuine individuals because of the three possibilities mentioned above. They can then argue that, since you can only trace your history back to other individuals, it follows that you cannot trace your history back to a morula or a blastocyst. The key claim that these philosophers fall back upon is typically argued for as follows:

1. No human individuals can exist so long as events like fission, fusion, and totipotent cell separation can occur.

2. Such events can occur up until the time when the primitive streak is formed.

Therefore,

3. No human individuals can exist prior to the time when the primitive streak is formed.

But this argument is defective for the same reason the other argument was defective: the first premise is false. Even if it is possible for me to undergo fission right now, this possibility of fission does not undercut the claim that I am an individual right now. Similar things can be said
regarding fusion and totipotent cell separation. Since these phenomena do not undercut the individuality of human adults, why should it be any different for human morulas and blastocysts?

2.3.4. Even Given Identity, It Begs the Question to Assume Serious Moral Status: Boonin

Some philosophers, though willing to admit that you could have been an infant or fetus (or embryo or zygote), still deny that there was a strong moral presumption against killing you at that time. When confronted with the case of an adult human organism that has serious brain damage, these philosophers, though willing to admit that you could survive such damage, still deny that there would be a strong moral presumption against killing you at that time. For example, David Boonin writes, at the outset of his book *A Defense of Abortion*, that

> In the top drawer of my desk, I keep another picture of [my son] Eli. This picture was taken on September 7, 1993, 24 weeks before he was born. The sonogram image is murky, but it reveals clearly enough a small head tilted back slightly, and an arm raised up and bent, with the hand pointing back toward the face and the thumb extended out toward the mouth. There is no doubt in my mind that this picture, too, shows the same little boy at a very early stage in his physical development. And there is no question that the position I defend in this book entails that it would have been morally permissible to end his life at this point.³³

Although an examination of the details of Boonin’s own position is beyond the scope of this essay, a move he makes when defending that position is directly relevant to the points being made in this chapter. Boonin considers a counterexample to his view of “…an imaginary case in which a temporarily comatose adult has had the entire contents of his brain destroyed so that

---

there is no more information contained in his brain than is contained in that of the preconscious fetus.” Boonin argues that, in a case such as this,

It seems right that my position does not imply that such an individual has the same right to life as you or I. But... a critic of abortion cannot appeal to such a case as a means of rejecting my position because we cannot assume ahead of time that killing such individuals is seriously immoral. 34

I believe that the claim Boonin is making here is mistaken. We are well within our rights to assume, ahead of time, that killing such individuals is seriously immoral. Any one of us could become such a temporarily comatose adult. If a cure for our condition were available, it would not be a matter of moral indifference whether we were killed or cured. So the assumption that killing such individuals is seriously immoral is an assumption we are entitled to use in framing our moral views. If a given moral view cannot accommodate this assumption, this is a mark against the view and not a mark against the assumption.

Perhaps a short summary of this section would be helpful. It has been argued that even the most undeveloped human organisms possess the set of capacities that generates a strong moral presumption against killing them. This conclusion was reached by arguing that, as an adult human organism undergoes temporary changes that are more and more serious, the order of the capacities we must appeal to in generating a strong moral presumption against killing this human organism gets higher and higher. We eventually reach a point where the adult in the middle of a temporary change has an order of capacities that is just as high as the order of capacities possessed by the most undeveloped human organisms. Finally, a few of the more promising attempts to block this line of argument were seen to be unsatisfactory.

34 Ibid., 78.
3. Higher-Order Capacities and the Argument from Potential

It is one thing to show even the most undeveloped human organisms possess the relevant set of capacities throughout their existence. It is another thing to defend a version of the AFP. For the AFP, it is alleged, entails various sorts of absurdities and commits various sorts of fallacies. Therefore, this section argues that such allegations do not threaten the version of the AFP represented by this chapter.

3.1 Human Germ Cells and Somatic Cells

The AFP is often accused of leading to the absurd conclusion that human gametes—whether sperm, or egg, or some combination of sperm and egg prior to conception—possess the sort of properties that generate what this essay has called a strong moral presumption against killing. A similar allegation has been made against the AFP regarding human somatic cells—whether the clonable somatic cells of an adult human being, or the “totipotent” somatic cells of a developing human embryo.

If the strong moral presumption against killing were merely grounded in a more general sort of moral presumption against stopping causal sequences whenever those sequences would have contributed to producing certain sorts of valuable outcomes, or certain sorts of valuable entities, then there would indeed be a strong moral presumption against killing human gametes. For many human gametes are indeed members of such causal sequences, and killing these gametes is a way of stopping these sequences. A similar thing could be said about certain human somatic cells. But this essay has argued that certain higher-order capacities of an entity are what ground the strong moral presumption against killing that entity, quite apart from what causal
sequences the entity is a part of. Hence, in order to show that the present account entails the absurd conclusion that there is a strong moral presumption against killing human gametes, or human somatic cells, one needs to show that these gametes, or somatic cells, possess the relevant higher-order capacities.

The most direct way of showing this would be to show that the gametes, or the somatic cells, are identical to a thing that possesses the relevant higher-order capacities. Because of this, and because the previous section argued that a human zygote possesses the relevant higher-order capacities, it will now be argued that no human gametes (whether considered separately or considered jointly) and no human somatic cells (whether clonable or totipotent) are identical to any human zygote. The basic claim behind this argument is that, whenever we are confronted with a situation where a zygote appears to be identical to one of its “precursors”, the situation is better understood as one where all the material that used to constitute the precursors now constitutes the zygote, even though the precursor is not identical to the zygote.

Few would claim that a sperm cell is identical to the zygote it helps to produce. Perhaps the reason for this is due to simple matters of size. Imagine that a statue was composed of 99% bronze and 1% silver. It is hard to believe that the silver that went into the statue is identical to the statue itself. Perhaps considerations of these sorts are what lead us to resist the idea that a sperm cell is identical to the zygote it helps to produce.

However, more must be going on than considerations of relative size in order to deny the claim that an egg cell is identical to the zygote it helps to produce. After all, the overwhelming bulk of the material constituting the zygote constituted the egg a short time earlier. This biological fact makes it seem easier to refer to the zygote as a “fertilized egg” rather than an “enveloped sperm”. Indeed, one occasionally hears the suggestion that, because the zygote is
referred to as a “fertilized egg”, the zygote literally is an egg. Perhaps these facts about biology and our linguistic conventions are responsible for way philosophers invoke the individual egg cell as a reason for rejecting the moral relevance of an entity’s “potential”. For example, when Mary Anne Warren considers John Rawls’ claim that the “capacities” that underwrite human equality must be understood in a way that coheres with our intuitions regarding human infants, she claims that his claim leads to absurd results:

The hypothesis that the potential to develop one’s own ends and a sense of justice is sufficient for moral personhood enables Rawls to gather normal infants and young children into the fold. However, it appears to do this at the cost of also admitting fertilized or unfertilized human ova—which also have the potential, under the right circumstances, eventually to develop the capacities in question.35

Although Warren may be right in suggesting that Rawls’ account admits fertilized human ova into the fold, her suggestion that it brings unfertilized human ova into the fold relies on the assumption that the unfertilized ova is identical to the entity that develops the capacities in question. But this assumption is mistaken, since the zygote is not identical to the egg. Just as a statue whose composition is 99% bronze and 1% silver is not for that reason identical to the lump of bronze that went into the sculptor’s shop, so too a zygote, whose composition consists of materials, most of which came from an egg, is not for that reason identical to the egg. It is not merely considerations of relative size that prevent us from identifying one entity a with another entity b, but considerations of the difference between a continuing to exist as b and a contributing all of its material to b.

35 Warren, 105, emphasis mine.
The idea that the zygote is not identical to either sperm or egg is confirmed by attempting to construct a temporary change for the sake of gametes that parallels the temporary change constructed above for the sake of zygotes. Recall the modification to Jeff McMahan’s thought experiment, where your body can shrink down to a zygote and then grow back into a fully formed adult organism. Now imagine that, instead of making the zygote the turn-around point in this biological cycle, we allow the zygote to split apart into a sperm and an egg, and then we allow the sperm and egg to come back together again and re-fuse, and then we allow the zygote to grow up again, so that the adult organism after the shrinking-fissioning-fusioning-growing process has the same personality traits, etc. as the adult organism before this shrinking-fissioning-fusioning-growing process. Now ask: do you think that either the sperm or the egg in the middle of this process was really you? I believe that most people, when confronted with this prospect, would deny that they were either the sperm or the egg in the middle of this process. Consequently, when confronted with the apparent fact that “you” seem to exist at the end of this process, one must either reject this apparent fact (since individuals cannot have temporal gaps) or else explain how “you” can continue to exist during the fissioning-fusioning part of the process without existing as either the sperm or the egg.

Some of the same things need to be said about parthenogenesis. Parthenogenesis (from the Greek words meaning “virgin birth”) takes places when an ovum, without the actual fertilization of a sperm, nevertheless begins to start dividing in the way a zygote divides. Parthenogenetic development is thought to be rather rare for humans; although it can be triggered in a laboratory by what biologists call “noxious stimulation,” the process of dividing has never been observed to continue for more than a few days. But in some organisms parthenogenesis is the standard way of reproduction. Whether in other organisms or in humans, however, the
relevant point is the same: parthenogenetic development should not lead us to assume that an individual egg cell that exists before parthenogenesis is strictly identical to an individual cell that exists after parthenogenesis. Just as a lump of bronze is not identical to the bronze statue that it gives rise to, even if that statue is composed of all and only the bronze that composed the lump, so too an egg is not identical to the zygote that it gives rise to via parthenogenesis, even if that zygote is composed of all and only the material that composed the egg.

The last few paragraphs do not show that there is something incoherent about the claim that the sperm (or the egg) possesses the set of higher-order capacities that generates a strong moral presumption against killing. They merely show that this claim is not entailed by the account of higher-order capacities given above. For the account of higher-order capacities given above does not merely rely upon whether the time it takes for a given entity a to “give rise to” a given entity b is equivalent to the time it takes for you to emerge from a certain sort of temporary change. In addition, the account of higher-order capacities given above relies upon whether a and b are identical to one another.

Even if neither the sperm nor the egg, considered separately, has the relevant higher-order capacities, it might be thought that the sperm and egg, considered jointly, do have the relevant higher-order capacities. For it might be thought that

(1) The sperm and egg, considered jointly, do indeed constitute some sort of entity or thing.

and that

(2) This entity or thing is identical to the zygote.

At this point it should come as no surprise that the present account, even if it agreed to (1), is going to disagree with (2) for the reasons given a moment ago: just because a zygote is
composed of all and only the material that composed a certain entity, that does not mean that the
zygote is identical to that entity. But is there any reason for thinking that (1) is true? Alistair
Norcross once flirted with denying the claim that “a combination of sperm and ovum, understood
as a mereological sum, is not a thing”\textsuperscript{36}:

Perhaps the most obvious answer to the charge that a combination of sperm and
ovum is not a thing is simply to deny it. I am inclined to pursue this option. To
the extent that I am prepared to admit that a zygote or a fetus is a thing, I would
claim that a combination of sperm and ovum is also a thing.\textsuperscript{37}

But this denial is tantamount to putting mereological sums on the same metaphysical footing as
genuine wholes. For example, it puts the mereological sum consisting of your various body
organs (or cells, or chemicals, or whatever part-type we choose to focus on with our
mereological microscopes) on the same metaphysical footing as you. Indeed, since mereological
sums are no respecter of the separateness of persons, Norcross’ denial even puts you on the same
metaphysical footing as a mereological sum consisting of your heart, my lungs, and so on.

Because we have good reasons for denying that mereological sums in general are on the same
metaphysical footing as genuine wholes, and because Norcross offers no reason for his denial
apart from his own inclination, it seems that we are still justified in resisting (1).

To sum up this discussion of gametes: one of the reasons the gametes are not
metaphysically on a par with the zygote is that neither gamete can claim to be strictly identical
with the zygote. Although the gametes, considered jointly before fertilization, have the potential
to produce a zygote, and although the gametes are indeed spatiotemporally continuous with the


\textsuperscript{37} Ibid.
zygote, it is not true that either gamete is the same organism as the zygote. The difference between killing the gametes before fertilization has occurred and killing the zygote after fertilization has occurred is the same difference as that between preventing a human organism from beginning to live and preventing a human organism from continuing to live.

Now exactly the same sorts of considerations are going to be relevant to a proper evaluation of the metaphysics of clonable adult cells and totipotent embryonic cells. The main idea is that, regardless of what cell or group of cells one begins with, and regardless of what event triggers the genesis of a zygote, “cells can be altered intrinsically, by events which scientists are becoming ever more competent to identify, from undetached constituent parts of the organism, to detached embryos [zygotes] in their own right.”

3.2. Potential Presidents and Potential Persons

As was mentioned at the outset of this chapter, the AFP is often formulated using the language of rights. When formulated this way, the AFP is often opposed by the example of the potential president: just as the mere fact that an individual is a potential president does not give that individual the right to command the military, so too the mere fact that an individual is a potential person does not give that individual the right to life. There are two reasons for ending this chapter with a discussion of this example. First, although this example alerts us to a possible mistake to avoid in formulating the AFP, it can be shown that the formulation of the AFP represented by this essay does not make that mistake. Second, although this example has been widely discussed by philosophers, the discussion has completely ignored the following interesting feature of the office of the American presidency: an individual can be a president

during one period of time, cease being a president during a second period of time, and then
become a president, again, during a third period of time. Attending to this feature of the
presidency is important for adequately evaluating both the rights-based versions of the AFP and
the version of the AFP represented by the argument of this essay.

3.2.1. The Potential President Example

The following claim is often invoked in discussions about the morality of abortion,
infanticide, embryonic stem cell research, and a number of other issues that touch upon the
beginning stages of human life:

(1) POTENTIAL RELEVANT TO MORAL STATUS: an entity’s potential is relevant to
its moral status

One of the more common ways of fleshing out this claim relies upon the concept of a potential
person and the concept of a right to life, as follows:

(2) POTENTIAL PERSONHOOD GENERATES A RIGHT TO LIFE: the fact that an entity
is a potential person is sufficient for it to possess a right to life

This latter claim, in turn, is commonly opposed by the example of the potential president, which
received its initial formulation by the Australian philosopher Stanley Benn:

[My argument] is not the argument that infants are potential persons, and have
rights as such. For if A has rights only because he satisfies some condition P, it
doesn’t follow that B has the same rights now because he could have property P at
some time in the future. It only follows that he will have rights when he has P.
He is a potential bearer of rights, as he is a potential bearer of $P$. A potential president of the United States is not on that account Commander-in-Chief.\textsuperscript{39}

Benn’s example of the potential president has received a fair amount of philosophical attention over the last thirty years, and variations on the example abound—for example, Peter Singer used a distinctively British version of the example in his \textit{Practical Ethics}, noting that “Prince Charles is a potential King of England, but he does not now have the rights of a king.”\textsuperscript{40} However, the discussion, as a whole, has completely ignored the following interesting feature of the office of the American presidency: an individual can be a president during one period of time, cease being a president during a second period of time, and then become a president, again, during a third period of time. This section explains why attending to this feature of the presidency is important for adequately evaluating (2).

Two important preliminary points are discussed in the next few paragraphs. First, it is important to have a clearer understanding of the operative concept of a person that appears both in (2) and in the example of the potential president. Second, it is important to have a clearer understanding of why one of the standard responses to the example of the potential president is both correct yet limited in scope.

(2) uses the term “person”. But the term “person”, as Chapter One discussed, is used in a number of different ways both inside and outside philosophical discussions. Recall how one of the most important distinctions in the way “person” is used is what Joel Feinberg calls the distinction between normative (or moral) personhood on the one hand and descriptive (or commonsense) personhood on the other. (2) uses “person” in a descriptive sense, not a


\textsuperscript{40} Peter Singer, \textit{Practical Ethics}, 2d ed. (Cambridge: Cambridge University Press, 1993), 153.
normative sense. The claim is *not* that the fact that an entity is a potential normative person is sufficient for it to possess a right to life; rather, the claim is that the fact that an entity is a potential descriptive person is sufficient for it to possess a right to life. (2), as stated, leaves open the question of precisely which characteristics—reason, and/or consciousness, etc.—are constitutive of being a “person” in a descriptive sense.

The example of the potential president also uses “person” in a descriptive sense, not a normative sense. As Benn originally put it,

I characterize a person…as someone aware of himself, not just as process or happening, but as *agent*, as making decisions that make a difference to the way the world goes, as having projects that constitute certain existing or possible states as “important” and “unimportant,” as capable, therefore, of assessing his own performances as successful or unsuccessful.41

One need not be committed to Benn’s proposal of the characteristics constitutive of being a “person” in order to rely upon his example of the potential president. However, one certainly does need to keep to *some* descriptive use of “person”—and to stay away from *all* normative uses of “person”—in order to rely upon Benn’s example. Otherwise, the entire structure of the example, and its relevance to (2), collapses.

One of the standard responses to Benn’s example of the potential president is to claim that, although he is correct both in his description of the example and in the general principle he extracts from the example, he is mistaken in thinking that either the general principle or the example are relevant to (2). Although I believe this standard response to be correct as far as it

---

41 Benn, 141.
goes, there is nevertheless another way of employing Benn’s example that appears to cause
problems for the defender of (2).

First, consider why the standard response is correct. Benn imagines the defender of (2)
arguing as follows (dropping the bit about A and B, and inserting the rider ‘at t’):

(i) X has the right to life at t only because X is a person at t.

Therefore,

(ii) X has the right to life at t because X could be a person at some time after t.

The conclusion, (ii), is basically the same claim as (2). But, as Benn correctly notes, the
argument from (i) to (ii) is just a particular version of the following argument form:

(iii) X has rights at t only because X satisfies some condition P at t.

Therefore,

(iv) X has rights at t because X could have P at some time after t.

But this argument form is clearly invalid, as can be seen by filling it in with the concept of a
president and the concept of a right to command the military:

(v) X has the right to command the military at t only because X is a president at t.

Therefore,

(vi) X has the right to command the military at t because X could be a president at
some time after t.

Benn quite correctly claims that what follows from (v) is not (vi) but

(vii) X will have the right to command the military at t when X is a president at

More generally, what follows from (iii) is not (iv) but

(viii) X will have rights at t when X has P at t.
Therefore, what follows from (i) is not (ii) but

(ix) X will have the right to life at t when X is a person at t.

If the defender of (2) really were arguing from (i) to (ii), then the example of the potential president would be an appropriate criticism. As Joel Feinberg put it,

It is a logical error…to deduce actual rights from merely potential (but not yet actual) qualification for those rights. What follows from potential qualification…is potential, not actual, rights; what entails actual rights is actual, not potential, qualification. As the Australian philosopher Stanley Benn puts it,

“A potential president of the United States is not on that account Commander-in-Chief [of the U.S. Army and Navy].” This simple point can be called “the logical point about potentiality.”

However, a standard response to Benn’s example is that Benn has misunderstood the defender of (2). After all, this standard response goes, the defender of (2) would not accept (i) in the first place. Still less would she attempt to argue for (ii) on the basis of (i). As Michael J. Wreen notes,

Proponents of [the view that all potential persons have a right to life] would agree that mere potential possession of the qualifications for a right is not sufficient for actual possession of that right. But the question, they would add, is what the actual qualifications for possession of a right to life are. They hold that potential personhood is that—or, better, one such—actual qualification, or sufficient

---

condition, and so quite correctly claim that the “logical point about potentiality” that Feinberg mentions counts nought against them.43

I believe this standard response is correct as far as it goes. And this standard response is relevant to the argument of this essay, even though the argument of this essay does not make use of the concepts of rights, persons, or potential persons. The argument of this essay is not: because an individual’s possession of a certain set of immediate capacities generates a strong moral presumption against killing that individual, *it follows that* an individual’s possession of the capacity to have these immediate capacities generates such a presumption. Rather, the argument of this essay is: since there are cases where an individual’s possession of certain higher-order capacities *by itself* generates a strong moral presumption against killing that individual, it follows that possession of these higher-order capacities is always sufficient to generate such a presumption. Hence, the argument of this essay does not make the logical mistake that Feinberg borrows from Benn’s example of the potential president.

However, there is another way of employing Benn’s example that appears to cause problems for (1) (and, by extension, for the argument of this essay). In particular, the example of the potential president does more than merely illustrate the fallacy of reasoning from (i) to (ii), or “the logical point about potentiality”. In addition, the example of the potential president advances a rhetorically attractive way of thinking about persons that creates a presumptive case for rejecting (2) out of hand. The next few paragraphs explain in more detail why the example of the potential president appears to create a presumptive case for rejecting (2) out of hand. The following section explains why this appearance is actually misleading.

---

43 Wreen, 138.
Benn’s example of the potential president relies upon two assumptions about presidents and two parallel assumptions about persons. While the assumptions about presidents are unobjectionable, the assumptions about persons are more substantive. It is these substantive assumptions that are primarily responsible for creating a presumptive case for rejecting (2).

Benn’s first unobjectionable assumption is that the term “president” is what logicians call a *phase sortal* rather than a *substance sortal*:

(3) **Benn’s First Unobjectionable Assumption:** The term “president” is a phase sortal

This assumption is rarely made explicit, but it is important to do so. *Sortals* are count-nouns that serve as classificatory concepts for describing the world, and they can be divided up into *substance* sortals and *phase* sortals. As David Wiggins explains, the difference between substance sortals and phase sortals is between sortal concepts which present-tensedly apply to an individual \(x\) at every moment throughout \(x\)’s existence, e.g. *human being*, and those which do not, e.g. *boy*, or *cabinet minister*.\(^{44}\)

Part of the logic of phase sortals is illustrated by Wiggins’ example of the boy: there is nothing incoherent about an individual falling under a phase sortal \(P\) during one period of time, and yet failing to fall under \(P\) during a later period of time. Another part of the logic of phase sortals is illustrated by Wiggins’ example of the prime minister: there is nothing incoherent about an individual failing to fall under a phase sortal \(P\) during one period of time, and then coming to fall under \(P\) during a later period of time. This is why it makes perfect sense to treat the term

“president” as a phase sortal. There is nothing incoherent about an individual failing to be a president during one period of time, and then being a president during a later period of time.

Benn’s second unobjectionable assumption is that having the right to command the military at a given time requires being an actual president at that time:

(4) **Benn’s Second Unobjectionable Assumption:** having the right to command the military at a given time requires being an actual president at that time

When these two unobjectionable assumptions are combined, the result is that whenever an individual does not fall under the phase sortal “president”, that individual does not have the right to command the military. And since, when an individual is merely a potential president, that individual does not fall under the phase sortal “president”, it follows that, when an individual is merely a potential president, that individual does not have the right to command the military.

There are two substantive assumptions that make the example of the potential president relevant to discussions about potential persons. These two substantive assumptions are almost exactly the same as the two unobjectionable assumptions just considered, with the only differences being the substitution of the word “person” for the word “president” and the substitution of the phrase “right to life” for the phrase “right to command the military” in the relevant places:

(5) **Benn’s First Substantive Assumption:** The term “person” is a phase sortal

(6) **Benn’s Second Substantive Assumption:** having the right to life at a given time requires being an actual person at that time

When these two substantive assumptions are combined, the result is that whenever an individual does not fall under the phase sortal “person”, that individual does not have the right to life. And
since, when an individual is merely a potential person, that individual does not fall under the phase sortal “person”, it follows that, when an individual is merely a potential person, that individual does not have the right to life. If Benn’s two substantive assumptions are true, then (2) must be rejected as false.

3.2.2. Re-electing the Potential President

Are Benn’s two substantive assumptions true? Some philosophers have attacked (5) by arguing that the term “person” is not a phase sortal at all, but is rather a substance sortal. This attack often takes the form of a complaint that there is something fundamentally incoherent in talk about a “potential person”, or that to speak of a “potential person” is to make some kind of category mistake.

However, there are two reasons why, for the sake of the present argument, I am willing to grant (5). The first reason is simply a matter of usage: it seems to me that the term “person” is sometimes used as a phase sortal, sometimes used as a substance sortal, and sometimes used with no clear commitment either way. Among philosophers and non-philosophers alike, these different uses of “person” often pass one another unnoticed, like ships in the night. While this practice of multiple usage may be regrettable, it is no good to deny that it goes on.

The second reason for granting (5) is simply a matter of consistency: it would be self-defeating for a defender of (2) to complain that Benn uses “person” as a phase sortal. After all, (2) uses “person” as a phase sortal too. Those philosophers who insist that “person” is a substance sortal would argue, no doubt, that this is an excellent reason to reject (2) in favor of a more philosophically sophisticated formulation of (1). These philosophers may be right; still, that is another discussion for another time.
The question I intend to investigate is this: once (5) is granted, how plausible is (6)?

Here is where it becomes relevant that an individual can be a president during one period of time, cease being a president during a second period of time, and then become a president, again, during a third period of time.

One source of this feature of the American presidency is the provision in the American constitution that allows for presidents to run for re-election to a second term of office. Since nearly all first-term presidents have attempted to get re-elected for a second term, and since these re-election attempts are very much in the public view, it is somewhat surprising that philosophers have overlooked this feature of the presidency.

However, there may be an explanation available for why philosophers have overlooked this feature of the presidency: it is very unusual for a first-term president to even make the attempt at re-election for a second term, after having taking a break from the presidency, and it is even more unusual for such an attempt at re-election to succeed. This is not surprising, since the reasons why a first-term president would take a break from the presidency to begin with—illness, age, unpopularity, scandal, defeat, and death—are often excellent reasons for not running for the office in a future election.

Still, there is nothing incoherent about this feature of the American presidency. With (3) in mind, this feature of the American presidency illustrates how there is nothing incoherent about an individual falling under a phase sortal $P$ during one period of time, failing to fall under $P$ during a second period of time, and then falling under $P$ again during a third period of time.

Imagine the following scenario. Al is president from 3000-3004, ceases being president in 3005, and then gets re-elected to be president from 3009-3012. Put a bit differently, Al is an actual president from 3000-3004, a potential president from 3005-3008, and an actual president
from 3009-3012. The important point for present purposes is that from 3005-3008, Al is not an actual president but a potential president. Now enter (4). Since having the right to command the military at a given time requires being an actual president at that time, and from 3005-3008 Al is not an actual president, it follows that from 3005-3008 Al does not have the right to command the military. The entailment illustrated by Al’s case may be called THE INNOCUOUS

ENTAILMENT:

(7) THE INNOCUOUS ENTAILMENT: if an individual temporarily fails to fall under the phase sortal “president” at a given time, that individual loses the right to command the military at that time

In order to see how this example of re-electing the potential president has a parallel in the case of persons, we need only combine (5) with the truth, mentioned just a moment ago, that there is nothing incoherent about an individual falling under a phase sortal $P$ during one period of time, failing to fall under $P$ during a second period of time, and then falling under $P$ again during a third period of time.

Imagine the following scenario. A human organism named Bill possesses the properties that constitute being a person from 3000-3004, loses those properties in 3005, and then regains those properties from 3009-3012. Put a bit differently, Bill is an actual person from 3000-3004, a potential person from 3005-3008, and an actual person from 3009-3012. The important point for present purposes is that from 3005-3008 Bill was not an actual person but a potential person. Now enter (6). Since having the right to life at a given time requires being an actual person at that time, and since from 3005-3008 Bill was not an actual person, it follows that from 3005-3008 Bill does not have the right to life. The entailment illustrated by Bill’s case may be called THE EMBARRASSING ENTAILMENT.
(8) **THE EMBARRASSING ENTAILMENT:** if an individual temporarily fails to fall under the phase sortal “person” at a given time, that individual loses the right to life at that time.

To see why this entailment is embarrassing, take any given definition of “person” in a descriptive sense: for example, X is a person =df. X has reason. Now imagine that Bill has reason from 3000-3004, loses his reason from 3005-3008 due to a brain injury, and gets his reason back again in 3009. Benn’s two substantive assumptions, when combined, entail that Bill does not have the right to life from 3005-3008.

(8) can be reached even if the definition of person is changed to emphasize something other than reason: for example, X is a person =df. X has consciousness. Imagine that Bill has consciousness from 3000-3004, loses his consciousness from 3005-3008 due to a brain injury (he exists, let’s say, in a temporary coma) from 3005-3008, and gets his consciousness back again in 3009. Once again, Benn’s two substantive assumptions, when combined, entail that Bill does not have the right to life from 3005-3008.

Finally, (8) can be reached even if the definition of person is changed to “actual, continuing subject of experiences” as that phrase was defined in Chapter Three. Imagine that, in 3005, the human organism named Bill has his upper brain completely annihilated, even though Bill continues to exist. In fact, just to make the case even more dramatic, we can assume Bill has his *whole* brain completely annihilated, even though Bill continues to exist. (This assumption is disutable, since there is a legitimate concern about whether an organism can actually remain alive if its whole brain, including its brain stem, is destroyed. But we can bracket this concern by assuming that Bill remains alive due to the presence of an electronic ‘brain stem.’) Then, in 3009, a new whole brain is created, and put in Bill’s head, and programmed with completely
different psychological states. Once again, Benn’s two substantive assumptions, when combined, entail that Bill does not have the right to life from 3005-3008.

Now, to be sure, defenders of the view that only persons have a right to continued existence, where “person” is defined as “actual, continuing subject of experiences”, would not be disturbed by this result in cases where the whole brain (or even upper brain) is destroyed. They would say that the person who exists in 3010 is not identical with the person who existed in 3002, that no person at all exists in 3006, and that the human organism that existed in 3006—Bill—had no right to life. I, on the other hand, believe that these implications are good grounds for resisting, rather than adopting, such a view about persons.

(8) becomes all the more embarrassing when one realizes that the lapse in personhood is just as detrimental to Bill’s right to life whether it lasts for four years, four days, or four minutes. Imagine Bill goes in for minor surgery, and is given a general anesthetic that makes him completely unconscious during the surgery. Benn’s two substantive assumptions entail that Bill does not have the right to life during the time of this minor surgery.

There are at least three approaches for avoiding (8). The first is to claim that, even though Bill is not actually a person from 3005-3008, Bill still retains his right to life from 3005-3008, since Bill is a potential person during this time. According to this approach, an entity has the right to life at a given time as long as it is either an actual person at that time or a potential person at that time.

Unfortunately, this approach abandons (6). Indeed, this approach is tantamount to endorsing (2).

A second approach for avoiding (8) is to claim that, even though Bill is not actually a person from 3005-3008, Bill still retains his right to life from 3005-3008 because he already was
a person from 3000-3004. According to this approach, an entity has the right to life at a given time as long as it is either an actual person at that time or was an actual person at some previous time.

This second approach, unlike the first, is not tantamount to endorsing (2). Unfortunately, this second approach, just like the first, must abandon (6). The example of re-electing the potential president shows why. Al did not have the right to command the military from 3005-3008, even though Al already was a president from 3000-3004. If Benn’s two substantive assumptions are correct, there is no reason to think the situation is any different for Bill and the right to life than it is for Al and the right to command the military. Once (6) is abandoned, the alleged parallel between presidents and persons collapses, and the example of the potential president does not generate a presumptive case for rejecting (2).

A third approach for avoiding (8) is to claim that Bill does not cease being a person from 3005-3008, since the properties constitutive of being a person include both what may be called occurrent abilities and what may be called dispositional abilities. For example, if we focus on a definition of a person that emphasizes reason, then, according to this approach, something is a person at a given time as long as it has either the occurrent ability to reason at that time or the dispositional ability to reason at that time. The dispositional ability to reason just is the ability to acquire the occurrent ability to reason. According to this approach, even though Bill does not have the occurrent ability to reason from 3005-3008, he still has the dispositional ability to reason from 3005-3008. Therefore, Bill is still a person from 3005-3008.

This third approach, unlike the first one, does not require abandoning (6). However, this third approach has an interesting result: without explicitly endorsing (2), this third approach ends up generating a right to life for precisely the same entities as (2). This is because the talk about
dispositions is simply another way of talking about potential. Saying that Bill has, from 3005-3008, the dispositional ability to reason, is the same as saying that Bill has, from 3005-3008, the potential to reason.

Once it is recognized that this third approach depends upon potentiality in this way, it becomes clear that it is a mere notational variant on the first approach. The first approach combined two claims (focusing again on a reason-based definition of person):

1. X has the right to life only if X is a person or a potential person
2. X is a person =df. X has reason

The third approach takes these same two claims, and simply relocates the concept of potential (which it calls a disposition) from the first claim to the second claim:

1. X has the right to life only if X is a person
2. X is a person =df. X has reason or the potential to have reason

This third approach, then, is merely a slightly different way of fleshing out the claim this section began with at the outset:

(1) Potential Relevant to Moral Status: an entity’s potential is relevant to its moral status

While (2) is one way of fleshing out this claim, this third approach for avoiding (8) suggests another:

(9) Personhood* Generates a Right to Life: the fact that an entity is a person* (an entity with reason or the potential to have reason) is sufficient for it to possess a right to life.

(2) and (9) are extensionally equivalent: any entity that has a right to life according to one principle will also have a right to life according to the other principle.
Thus, these three ways of avoiding (8) while retaining a commitment to (5) involve either abandoning (6) or else packing potentiality into the concept of a person to begin with. If (6) is abandoned, the alleged parallel between persons and presidents collapses, and the example of the potential president does not generate a presumptive case for rejecting (2). If potentiality is packed into the concept of a person to begin with, then there will be no difference in the range of entities that have a right to life according to (1) and the range of entities that have a right to life according to (9).

There are three general lessons to be learned from this discussion of Benn’s example of the potential president, and these lessons apply whenever an account of the ethics of killing uses the term “person”. First, it is important to be clear on whether “person” is being used in a normative sense or in a descriptive sense. Second, it is important to be clear on whether “person” is to be taken as a phase sortal or a substance sortal. Third, whenever an account of the ethics of killing uses the term “person” in the descriptive sense as a phase sortal, it is important to be clear on whether X being a potential descriptive person is sufficient for X being a normative person. For example, if an account claims that X is a person =df. X has reason, does the account also claim that X being a potential person is sufficient for X having the right to life? If not, then one should expect the account of killing in question to be vulnerable to the problem of re-electing the potential president.

These three general lessons also can be extended to apply to accounts of the ethics of killing, like the present one, that do not use the term “person”. The first general lesson here is that it is important to be clear on whether any term or phrase is being used in a normative sense or in a descriptive sense. In the present essay, the phrase “having a set of higher-order typical human capacities” is being used in a descriptive sense. The second general lesson here is that
is important to be clear on whether any term or phrase is to be taken as a phase sortal or a
substance sortal. In the present essay, the phrase “having a set of first-order typical human
capacities” is to be taken as a phase sortal. Finally, the third general lesson here is that whenever
an account of the ethics of killing uses a term or phrase in the descriptive sense as a phase sortal,
it is important to be clear on whether X having the potential to fall under this term or phrase is
sufficient for X to have some moral status property or other now. If not, then one can be
confident that the account of killing in question will be vulnerable to the problem of the
temporary change. This is because the problem of re-electing the potential president is simply a
politically interesting version of the problem of the temporary change.
Chapter Two began the defense of the first step of the main argument of this essay, that if an entity is human, it has a set of typical human capacities. Chapter Three began the defense of the second step of the main argument, that if an entity has a set of typical human capacities, it has serious moral status. While the previous chapter built upon this account by explaining the connection between capacities and the Argument from Potential, this chapter explains the connection between capacities and the Argument from Marginal Cases (henceforth AMC).

1. The Argument from Marginal Cases and the Argument of This Chapter

The general strategy of the “Argument from Marginal Cases,” as its name suggests, is to argue from the moral status of certain “marginal cases” of human beings to the moral status of certain non-human animals. The term “marginal” here just means “nonparadigmatic”: marginal cases are nonparadigmatic human beings who seem to have (mental) capacities equivalent to the (mental) capacities of non-human animals. The AMC argues that, since this seeming equivalence of (mental) capacities is real, and since any plausible criterion of moral status must be spelled out in terms of (mental) capacities, consistency requires us to conclude that the marginal cases and the non-human animals have an equivalent moral status.1

---


2 Consequently, the AMC is occasionally called the “Argument for Moral Consistency” because it urges us to be consistent in our evaluation of the moral status of non-human animals and marginal cases. See Daniel Dombrowski, *Babies and Beasts: The Argument from Marginal Cases* (Chicago: University of Illinois Press, 1997), 24.
Although this is the general strategy of the AMC, particular versions of it vary. Each version has, at its core, the comparison of certain marginal cases and certain non-human animals. At least four things account for the differences between the versions.

First, the AMC comes in both critical and constructive versions. Tom Regan, who defends the claim that non-human animals possess rights, is an exponent of both versions. His formulation of the critical version runs as follows:

1. Given certain criteria of the possession of rights, some marginal humans and not just all animals will be excluded from the class of right-holders.
2. However, humans, including those who are marginal, do have rights and so belong in the class of right-holders.
3. Therefore, each and every one of the criteria of which (1) is true must be rejected as setting a requirement for the possession of rights.

Regan’s formulation of the constructive version of the AMC is this:

1. Humans, including those who are marginal, have rights and therefore belong in the class of right-holders.
2. However, given the most reasonable criterion of the possession of rights, one that enables us to include marginal humans in the class of right-holders, this same criterion will require us to include some (but not all) animals in this class.
3. Therefore, if we include these marginal humans in the class of right-holders, we must also include some animals in this class.

---

It is worth noting that premise 2 in the critical version and premise 1 in the constructive version are the same.

The second thing that accounts for the differences between the versions of the AMC is the fact that the AMC comes in both weak and strong versions. The weak version attempts to show that if the marginal cases have rights (or some moral status or other; more on this in a moment), then the non-human animals also have rights. The strong version adds to the weak version some justification for the claim that the marginal cases have rights. Both the critical and the constructive versions of the AMC listed above, as formulated, are weak versions, since each simply assumes (premise 2 in the critical version, premise 1 in the constructive version) that the marginal cases have rights.

This difference between the weak and strong versions allows critics of the AMC to turn the argument on its head. Such a critic can admit that the marginal cases and non-human animals have an equivalent moral status, yet deny that either the non-human animals or the marginal cases have rights. For example, if we begin with the weak version of the argument, which says that if the marginal cases have rights, then the non-human animals also have rights, and if we add the claim that the non-human animals do not have rights, it follows that the marginal cases do not have rights either. And some thinkers are willing to accept this. So then, although many use the weak version of the AMC to establish claims like “vegetarianism is morally obligatory,” there is nothing about the logic of the weak version that prevents others from using it to establish claims like “cannibalism is morally permissible.”

There is a third thing that accounts for the differences between the versions of the AMC, which has less to do with the argument’s form and more to do with its content. Although the AMC is often used by animal rights theorists to refute various criteria for the possession of
rights, its overall structure can be used even when the language of rights is not. This is because the comparison at the heart of the AMC is a useful heuristic device for testing any moral status concept. For example, a constructive version of the AMC formulated in terms of the strong moral presumption against killing might run as follows:

1. Humans, including those who are marginal, belong in the class of those whom there is a strong moral presumption against killing.

2. However, given the most reasonable criterion of the basis of the strong moral presumption against killing, one that enables us to include marginal humans in the class of those whom there is a strong moral presumption against killing, this same criterion will require us to include some (but not all) animals in this class.

3. Therefore, if we include these marginal humans in the class of those whom there is a strong moral presumption against killing, we must also include some animals in this class.

The fourth thing that accounts for the differences between the versions of the AMC also has to do with its content. Different versions of the AMC can be generated depending on which marginal cases and which non-human animals are being compared. There is obviously great variety among non-human animals, and there are also many different types of marginal cases. A good illustration of this is found in those passages, quoted in Chapter One of this essay, from the beginning of Jeff McMahan’s book, The Ethics of Killing: Problems at the Margins of Life. McMahan claims that there are “four distinct categories into which we may sort most or all instances of killing for which there may be a reasonable justification,” and one of these

---

4 McMahan, vii-viii.
categories includes “cases in which the metaphysical or moral status of the individual killed is uncertain or controversial.” The way he begins his discussion of this category illustrates the heterogeneity of the class of marginal cases:

Among those beings whose nature arguably entails a moral status inferior to our own are animals, human embryos and fetuses, newborn infants, anencephalic infants, congenitally severely retarded human beings, human beings who have suffered severe brain damage or dementia, and human beings who have become irreversibly comatose.\(^5\)

Some of the marginal cases seem to have (mental) capacities that are not equivalent to, but lower than, many non-human animals. (Indeed, some of the marginal cases seem to have (mental) capacities that are the equivalent of vegetables.) The content of any particular version of the AMC will be a function of the specific sorts of beings compared.

The AMC is a hot topic in contemporary applied ethics. A recent book-length treatment of this argument goes so far as to say that the AMC is “an argument that has generated perhaps more light and heat than any other argument in moral philosophy over the last twenty years.”\(^6\)

Although it is possible to debate the moral status of marginal cases outside of the context of the AMC, the AMC is often lurking in the background of contemporary debates about marginal cases.

This chapter argues that the AMC can be answered by attending to an organism’s hierarchy of capacities. It will be argued that even the most extreme marginal cases among humans have a set of general, higher-order, typical human capacities. Since these higher-order

\(^5\) Ibid., vii.

\(^6\) Dombrowski, 3.
capacities are sufficient to generate a strong moral presumption against killing normal adult human persons, they are also sufficient to generate a strong moral presumption against killing the marginal cases. But since the non-human animals we are aware of do not have these higher-order capacities, it is an open question whether there is a strong moral presumption against killing any of them.

Two tasks are involved in explaining how capacities are relevant to the AMC. The first task is that of showing how it makes sense to think that the marginal cases possess higher-order capacities that generate a strong moral presumption against killing. But the second task is that of showing how one can accomplish this first task without being forced to admit that the non-human animals we are aware of possess such capacities.

The following two sections correspond to these two tasks. Section 2 explains why, although it may appear that some marginal cases do not have the relevant higher-order capacities that generate a strong moral presumption against killing, it turns out that all marginal cases really do have these relevant capacities. Section 3 replies to the objection that the present account faces a dilemma: either it is committed to making some non-human animals have the same higher-order capacities as humans, or else it is guilty of some morally objectionable form of “anthropocentrism” or “speciesism”.

2. Marginal Cases and Higher-Order Capacities

2.1 The Problem: Some Marginal Cases Appear Not To Have Any Higher-Order Capacity To Think

The class of marginal humans is large. Yet there are important differences within the class. It includes disabled human organisms such as those whose mental capacities are
compromised by disease, disorder, or accident. But it also includes undeveloped human organisms such as infants, fetuses, and embryos. Consequently, the Argument from Potential, discussed in Chapter Four, seems to be directly relevant to many of the marginal cases envisioned by the AMC: in particular, to normal, undeveloped human organisms. Since these marginal cases were already discussed in Chapter Four, the present chapter will focus upon disabled human organisms, including both those that are developed and those that are undeveloped.

Chapter Three argued that one of the advantages of grounding the strong moral presumption against killing in the higher-order capacities of individuals is that this strategy accommodates our reflective beliefs about cases involving temporary changes. For example, a human organism in a temporary coma still has the higher-order capacity to think, even though she does not have the immediate capacity to think. Since a strong moral presumption against killing is generated by possession of the higher-order capacity, there still exists a strong moral presumption against killing this organism during the time of her temporary change. By this route, any marginal case that is simply in a temporary change would still be such that there is a strong moral presumption against killing her.

Nevertheless, it might be thought that this strategy does not address other marginal cases. For example, there are human organisms in the late stages of dementia or Alzheimer’s disease, human organisms in “persistent” or even “permanent” vegetative states, and human organisms in “irreversible” comas. Such human organisms, it might be claimed, do not even have a higher-order capacity to think.

This sort of problem is discussed by a number of writers on the AMC. Recall how talk about higher-order capacities is a mere notational variant on talk about the potential to have an
immediate capacity. A number of writers on the AMC claim that, while grounding the moral status of marginal cases on their potential to have a certain capacity may indeed deal adequately with one sub-class of the marginal cases, it does not deal adequately with other sub-classes of the marginal cases. For example, Lawrence Becker writes that

There does not seem to be a morally relevant characteristic that distinguishes all humans from all other animals…The assertion that the difference lies in the potential to develop interests analogous to those of normal adult humans is also correctly dismissed. After all, it is easily shown that some humans—whom we nonetheless refuse to treat as animals—lack the relevant potential.  

Likewise, Daniel Dombrowski writes that “if the potential for a developed mental life is the criterion for moral respect, then there will still be marginal cases to lend support to the AMC.”

A few pages later, he re-emphasizes the point that no matter what criterion one picks—even if the criterion incorporates “potential”—there will still be marginal cases left over:

The most marginal of the marginal cases—those human beings who never have manifested any interests and never will do so, who apparently do not experience needs, affection, aversion, hopes, or fears—are in a situation quite different from that faced by a retarded human being or an infant.

Even Tom Regan ends up amending his critical and constructive arguments listed above, so that the marginal humans only include those “who are not irreversibly comatose.”

---


8 Dombrowski, 21.

9 Dombrowski, 26.

In order to make this problem as strong as possible, it is worth returning to the writings of C. D. Broad. As indicated earlier, Broad defends the view that “One of the peculiarities of minds in general, and of human minds in particular, is that they start with very few first-order powers, but rather with powers to acquire powers.”\textsuperscript{11} These powers to acquire powers are the same thing as what this essay has called higher-order capacities. Just as higher-order powers are powers to acquire powers, so too higher-order capacities are capacities to acquire capacities. But Broad immediately qualifies this idea as follows:

In this connexion it is important to distinguish between two cases, viz., the reversible and the irreversible. In the first case a power can be gained and lost and gained again repeatedly by appropriate changes in the external circumstances. A bit of iron can be magnetised, and demagnetised, and remagnetised, repeatedly. In the second case the substance has not the power to regain a certain power which it has lost, or to lose a certain power which it has gained. If you injure a man’s brain in certain ways, his mind will lose certain powers, and there is no known way of restoring these powers to his mind.\textsuperscript{12}

Since Broad uses a certain sort of brain-damaged human organism as an example of an individual that “has not the power to regain a certain power which it has lost,” and since higher-order powers just are higher-order capacities, then it seems Broad is committed to the idea that certain sorts of brain-damaged human organisms do not have even a higher-order capacity to think.

\textsuperscript{11} Broad, 267.

\textsuperscript{12} Ibid.
So, then, a higher-order capacity to think just is the potential to acquire the immediate capacity to think; but many writers on the AMC claim that some marginal cases do not have this potential. Likewise, a higher-order capacity to think just is the power to acquire the power to think; but even Broad claims that some brain-damaged humans do not have such higher-order powers. If Broad and many writers on the AMC are right, then there are marginal cases of human organisms that do not have even the higher-order capacity to think.

2.2. The Solution: Such Marginal Cases Have a Higher-Order Capacity To Think, But Need Technologically Sophisticated Therapy in Order To Realize It

This problem can be solved. If we attend carefully enough to the notion of a hierarchy of capacities, then every last one of the marginal cases—including even those Dombrowski calls “the most marginal of the marginal cases”—can be seen to possess the higher-order capacity to think. It is only the present state of technology that makes it difficult for us to recognize that such marginal cases possess the relevant higher-order capacity. But the technologies of the future—for example, neural therapies, neurosurgeries, reconstructive brain surgeries, and genetic therapies—will allow such marginal cases to overcome their cognitive deficiencies. This means that such marginal cases possessed the relevant higher-order capacity all along. They did not lack the capacity but the technology for realizing the capacity.

2.2.1. Marginal Cases Whose Marginality Is Due To Damage Have a Higher-Order Capacity To Think, But Need Brain Therapy of Various Kinds To Realize It

Recall Thomas Nagel’s example of an adult who suffers a life-changing accident that reduces them to having the mental life of a newborn infant. That human organism, while not
possessing the first-order capacity to think, nevertheless does possess a higher-order capacity to have this first-order capacity. This higher-order capacity, admittedly, would not be realized unless certain conditions were to be met—namely, the correction of the brain damage. But this should not cause us to be skeptical about the existence of this higher-order capacity. A human organism can have a capacity even if certain physical conditions block its realization.

The key for properly analyzing the problem Broad raises is to notice how he ends the last sentence of his discussion (quoted above) of the man with brain damage: “there is no known way of restoring these powers to his mind.” Just because there is no known way of restoring certain powers to a human organism (or, as Broad puts it, to the mind of a human organism), that does not mean there is no way at all of restoring those powers. For example, a thousand years ago the known ways of restoring the powers of thinking to a human organism were much fewer than the known ways of restoring such powers today. There are certain sorts of injuries to the brain that we can reverse with today’s technology, but that would have been irreversible a thousand years ago. If such an injury had occurred a thousand years ago, it would have been a mistake to claim that the injured organism did not have the power to regain the power to think. The organism did still have the power to regain the power to think. What was lacking was merely the technology to permit this higher-order power to be realized.

But notice what this entails. A thousand years from now the known ways of restoring the powers of thinking to a human organism will be much greater than the known ways of restoring such powers today. There are certain sorts of injuries to the brain that we cannot reverse with today’s technology, but will be able to reverse a thousand years from now. If such an injury occurs today, it is a mistake to claim that the injured organism does not have the power to regain
the power to think. The organism does still have the power to regain the power to think. What is
lacking is merely the technology to permit this higher-order power to be realized.

Hence there is the logical space to claim, in the case of the sorts of brain injuries Broad is
concerned with, that the loss of certain powers is only “irreversible” in the sense that we do not
currently possess the technology to reverse it. But this sort of irreversibility is not very
impressive from a metaphysical point of view. It tells us more about the state of our society’s
medicine than it does about the state of our patient’s mind.

The situation does not change even when an accident or disease destroys the entire upper
brain of a human organism, wiping out all first-order capacities for experiences. Nor does it
change when the brain damage involves the destruction of the whole brain. Stem-cell therapies
of the future, we can imagine without too much difficulty, will enable the organism to grow back
the relevant brain parts, in a way similar to the way a starfish that loses one of its leg parts is able
to grow back those relevant leg parts.

2.2.2. Marginal Cases Whose Marginality Is Due To an Inherited Condition Have a Higher-
Order Capacity To Think, But Need Genetic Therapy To Realize It

It might be thought that this way of dealing with marginal cases is acceptable in cases
involving brain damage, but not acceptable in cases involving inherited genetic conditions that
prevent the capacity to think from emerging. But this thought is mistaken because the gap
between brain damage and genetic condition is a gap that can be bridged.

The argument bridging the gap between brain damage and genetic condition can be
summarized in this paragraph and spelled out in the paragraphs that follow. One can imagine
special cases that collapse the distinction between brain damage and a change in genetic
condition. If the way of dealing with marginal cases outlined above is acceptable in cases involving brain damage, then it is also acceptable in these special cases. But if the way of dealing with marginal cases outlined above is acceptable in these special cases, then it is also acceptable for cases involving inherited genetic conditions. Therefore, if the way of dealing with marginal cases outlined above is acceptable in cases involving brain damage, then it is acceptable for cases involving inherited genetic conditions.

Begin by imagining a special case that collapses the distinction between brain damage and a change in a human organism’s genetic condition. A clear sort of case is where the brain damage is brought about by an accident that changes certain genes that code for specific brain functions. For example, imagine that a scientist works with a machine that produces massive amounts of different kinds of radiation. Let us call two of these kinds of radiation A-rays and B-rays. A-rays produce a genetic mutation in every cell of the human body such that the cells are genetically equivalent to the cells of a human organism with a genetically based brain handicap. B-rays reverse the effect of A-rays. One day the scientist inadvertently steps in front of the machine while it is emitting A-rays, and all of her cells promptly mutate.

If the way of dealing with marginal cases outlined above is acceptable in cases involving brain damage, then it is also acceptable in these special cases. After all, there does not seem to be a metaphysically relevant difference between the brain damage in such a special case and the brain damage in a more traditional case—for example, where the brain damage is the result of an injury to the brain from an automobile accident. There are certain differences in the details of how the brain damage came about and about how the brain damage can be repaired. But the special cases and the traditional cases are similar in that technological assistance is needed to repair the damage. In the traditional cases, it takes some sophisticated technology involving
brain surgery to fix the injury caused by the automobile accident; but in this special case, it just takes a flip of the switch to turn on the B-rays and fix the genetic condition caused by the A-rays. Therefore, after the administration of the A-rays and before the administration of the B-rays, it makes sense to say that the science worker still has the higher-order capacity to think. And this is so, in spite of the fact that after the administration of the A-rays and before the administration of the B-rays, our science worker does have a genetic condition that prevents the capacity to think from emerging. The basic insight of this analysis is this: the mere fact that a human organism possesses a genetic condition that prevents the capacity to think from emerging does not threaten the claim that the human organism possesses a higher-order capacity to think.

But there is no reason to think that this basic insight is limited to cases where the genetic condition is changed or caused by an accident. Imagine a second woman who is the same age as the science worker, but who has never worked in the science lab. This second woman has precisely the same genetic condition that A-rays can cause and that B-rays can fix. However, in the case of this second woman, the genetic condition is not a mutation caused by A-rays but is an inherited condition she has always possessed. Surely, if the science worker has the higher-order capacity to think, so does this second woman. After all, for both women, it just takes a flip of the switch to turn on the B-rays and fix the genetic condition. In short, since there is not a metaphysically relevant difference between a genetic condition that a human organism gets from an accident and a genetic condition that a human organism gets from its parents, it follows that there is not a metaphysically relevant difference between a genetic condition that a human organism gets from a change and a genetic condition that a human organism inherits. Whatever is true about the higher-order capacities in cases involving a changed genetic condition will also be true in cases involving an inherited genetic condition.
What emerges from the last three paragraphs is this: if the way of dealing with marginal cases outlined above is acceptable in cases involving brain damage, then it is acceptable for cases involving inherited genetic conditions. When a human organism has an inherited genetic condition that prevents the capacity to think from emerging, this does not automatically mean that the organism does not possess a higher-order capacity to think. Rather, it simply signals the fact that our present technologies have not advanced to the point where the human organism’s higher-order capacity to think can be realized. The genetic therapies of the future—whether they consist of B-rays or something else—will allow such organisms to realize their higher-order capacity to think.

This solution allows for a thing’s passive capacities to count as members of the set of that thing’s higher-order capacities. There are two reasons for allowing a thing’s passive capacities to count in this way. First, a thing’s passive capacities are just as real as its active capacities. Although their conditions for realization are different, the passive capacities are ontologically on all fours with the active capacities. Second, there is no reason for thinking this basic parity between passive capacities and active capacities changes when one moves to a different level in a hierarchy of capacities. Just as an active first-order capacity is ontologically on all fours with a passive first-order capacity, so too an active second-order capacity is ontologically on all fours with a passive second-order capacity. This pattern continues at each level in a hierarchy of capacities. The result is that, even if a brain-damaged (or genetically handicapped) human organism does not have the active higher-order capacity to think, this says nothing about whether she has the passive higher-order capacity to think. If she has a passive higher-order capacity to think, then she has a higher-order capacity to think.
If the argument of this sub-section is correct, then the higher-order capacities of an organism at a time should not be identified with that organism’s genetic code at that time. This is consistent with the arguments already given in the previous chapter on the Argument From Potential. For example, all of the somatic cells in a given human body have the same genetic code, but surely they do not all have the same higher-order capacities. Likewise, all of the somatic cells in the body of a given human adult have the same genetic code as the original one-celled zygote from which that adult organism began, but surely the original zygote does not have the same higher-order capacities as each of these adult somatic cells.

### 2.3. An Objection: This Solution Is Too Generous in Ascribing a Capacity To an Individual

One objection to this solution is that it is too generous in ascribing a capacity to an organism. In particular, this solution is willing to ascribe a capacity to an organism even when a great deal of external assistance is needed for that alleged capacity to be realized. According to this objection, it is an abuse of language to ascribe a higher-order capacity to think to a brain-damaged (or genetically handicapped) human organism, merely on account of the possibility of future technologies that could correct this brain damage (genetic handicap).

This objection is premature, and for two reasons. First, as Elizabeth Prior has shown, we should be careful not to let our ordinary-language usage of dispositional terms be decisive in situations where that ordinary-language usage conflicts with more precise scientific usage of dispositional terms. Two related points from Prior are especially relevant here. The first point concerns the relationship between a disposition and its background conditions; the second point concerns the relationship between a disposition and its “initiating cause”.
First, consider the relationship between a disposition and its background conditions. Prior uses the disposition of water-solubility to argue that we are caught between two conflicting intuitions whenever we attempt to decide in what case a disposition may be truly ascribed to an item:  

(1) “an intuition toward linking a dispositional predicate very firmly to a particular set of conditions”

(2) an intuition “to treat dispositional predicates as incomplete predicates”

If we follow (1), “we will ascribe the disposition of water-solubility to salt because salt dissolves in water at stp [standard temperature and pressure], but we will say that phosphorous is not water-soluble because it does not dissolve in water at stp.” If we follow (2), “we will say salt is water-soluble at stp, phosphorous is not water-soluble at stp, but phosphorous is water-soluble at high temperatures and pressures.” Prior then gives two reasons for preferring intuition (2) to intuition (1). Negatively, the only reason we have for preferring (1) is that “we tend to connect an ordinary-language dispositional concept more closely to one particular set of conditions than to any other,” and even this tendency is open to counterexamples. Positively, we should prefer (2) because

Those possessing the most precise dispositional concepts—the scientists—use dispositional predicates in this way. When a scientist talks about solubility it is always solubility in a particular solvent at a particular temperature and pressure. If he fails to indicate explicitly a solvent, temperature and pressure, it is assumed that he is talking about solubility in 100 ml of water at stp.

All quotations in this paragraph refer to Prior, Dispositions, 7.
So, then, the first relevant point from Prior amounts to this: although ordinary language may point slightly against attributing a disposition to an item, on account of nonstandard background conditions, nevertheless, scientists still use disposition predicates as incomplete predicates, predicates whose completion consists precisely in the filling in of details like background conditions.

The second relevant point from Prior is this: nonstandard initiating causes are to be treated in exactly the same way as nonstandard background conditions are treated. One of her examples here is the disposition of fragility: “Suppose we have a piece of steel…[that] will not shatter if struck with a blow of medium force…[but] will shatter if struck a blow of immense force…The question…is whether this [second] shattering is a manifestation of fragility.” She argues that the same negative and positive reasons for preferring intuition (2) to intuition (1) apply here: the only point in favor of (1) is ordinary language, but (2) has scientific practice in its favor.

These two points from Prior are relevant to the objection under consideration because the technological assistance needed to restore a brain-damaged human organism’s powers of thinking is simply a nonstandard initiating cause (and the same thing goes for a genetically handicapped human organism). Since an item can still have a disposition when nonstandard initiating causes are required to manifest that disposition, and since dispositions and capacities are the same thing, it follows that an organism can still have a capacity when nonstandard initiating causes are required to realize that capacity. And of course, there is no reason to think that this situation is any different for a higher-order capacity than it is for a lower-order capacity. Prior’s analysis thus provides an excellent first reason for thinking the objection under consideration is premature.
The second reason this objection is premature is going to take a few paragraphs to summarize. To begin with, even the activities of a normal, healthy human adult require external assistance of some kind. This is true, for example, for every act of her thinking: if certain external conditions (e.g. temperature and pressure) were not within a certain range, she would not be able to think the way she does. The same could be said if certain entities are not present in her immediate environment. If you take away her supply of oxygen, she will soon be unable to think. But this does not mean that she does not have the capacity to think. It merely means that she requires external assistance of some sort if she is going to realize the capacity to think.

But the assistance of oxygen is not importantly different from the assistance of technology, as the following two cases make clear. First, imagine two normal adult human organisms, A and B, who differ only in their spatial location—everything else about them, molecule for molecule, is the same. A is on the surface of the earth and is enjoying thinking about philosophy, but B is a thousand miles above the surface of earth and is enjoying very little right now. Do A and B have the same capacities or not? B will soon be unable to think because of the lack of oxygen in her spatial location. But surely the fact that B needs this external assistance does not threaten the claim that B has the capacity to think. The existence of B’s capacity to think does not depend on B’s location in space. Of course, the realization of this capacity may indeed depend on B’s location in space. But the existence of this capacity does not.

This truth about location in space is paralleled by a similar truth about location in time. Imagine now two brain-damaged human organisms, A and B, who differ only in their temporal location—everything else about them, molecule for molecule, is the same. A is, like you, living during a time when technology allows her sort of brain damage to be repaired; B is, unlike you,
living a thousand years ago, during a time when the technology which would allow for her sort of brain damage to be repaired has not yet been invented. Do A and B have the same capacities—in particular the same higher-order capacities—or not? B will not be able to think because of the lack of technology in her temporal location. But surely the fact that B needs this external assistance does not threaten the claim that B has the higher-order capacity to think. The existence of B’s higher-order capacity to think does not depend on B’s location in time. Of course, the realization of this higher-order capacity may indeed depend on B’s location in time. But the existence of this higher-order capacity does not. And a similar analysis would apply to inherited conditions.

It might seem, at first glance, that this second reason for thinking the present objection mistaken is in tension with the first reason for thinking the present objection mistaken. For while this second reason claims that a capacity (disposition) can exist even when the conditions required for its actualization (manifestation) are not present, the first reason claims that a capacity (disposition) is an incomplete predicate which is not definable apart from the conditions required for its actualization (manifestation). The way of resolving this tension, it seems, is the idea that a capacity (disposition) can exist as long as there are some conditions that will allow for its realization (manifestation). This idea preserves the intuition that capacities exist even in the absence of the conditions required for their realization, while still capturing Prior’s insight that dispositions are incomplete predicates.

Interestingly, however, Prior objects to this idea, because she is concerned it will compromise the usefulness of our dispositional terms. As she puts the point,

Dispositional predicates are useful because they divide up the world into those items which possess a particular disposition D and those items which do
not...[but] would lose this utility if our criterion for ascribing disposition D to an item were simply that that item would manifest that disposition under some set of conditions.\footnote{Prior, 6.}

I believe Prior’s concern here is misplaced. Although dispositional predicates formulated using specific conditions may be more useful than dispositional predicates formulated using nonspecific conditions, this does not mean that the latter are useless. Indeed, the reasons Prior gives for thinking the former useful are also reasons for thinking the latter useful. She claims if we use dispositional predicates as incomplete predicates,

These predicates will retain their utility. For in most cases the predicate ‘has disposition D (under conditions C)’ may be truly ascribed of some objects but not of others. Thus the predicate ‘has disposition D (under conditions C)’ helps us to divide up the world.\footnote{Ibid., 8.}

What Prior seems not to notice is that the very same thing can be said about dispositional predicates when the parenthetical qualifier ‘under conditions C’ is replaced with ‘under some set of conditions.’ In most cases the predicate ‘has disposition D (under some set of conditions)’ may be truly ascribed of some objects but not of others. Take solubility. Although Prior recognizes that “virtually any solid will dissolve under some set of conditions,”\footnote{Ibid., 6.} this still does not mean the dispositional predicate ‘soluble’ would be useless when defined as ‘has disposition to dissolve (under some set of conditions)’. For even on this definition, not everything in the world is soluble. For example, gases, liquids, fundamental particles, numbers, sets, and mental...
states are not soluble on this definition. Consequently, the predicate ‘has disposition to dissolve (under some set of conditions)’ helps us to divide up the world.

In summary, then, the objection under consideration in this sub-section claims that it is an abuse of language to ascribe a higher-order capacity to think to a brain-damaged (or genetically handicapped) human organism, merely on account of the possibility of future technologies that could correct this brain damage (genetic handicap). But there are two reasons why this objection is premature: first, following Prior, a higher-order capacity to think can still exist even though non-standard initiating causes (like technological assistance) are needed before that capacity is realized; second, the assistance provided by such technology is not importantly different than the assistance provided by standard sorts of assistance, such as breathable oxygen.

It is time to sum up the discussion of this section. Attending to the notion of a hierarchy of capacities is relevant to the AMC because it explains why there can still be a strong moral presumption against killing one sub-class of marginal cases—for example, sleeping humans, anesthetized humans, and temporarily comatose humans. However, it might be objected that attending to the notion of a hierarchy of capacities is not relevant to another sub-class of marginal cases—for example, the irreversibly comatose, those in a persistent vegetative state, or those who are congenitally mentally deficient because of a genetic condition. Nevertheless, by attending more carefully to the notion of a hierarchy of capacities, all human organisms can be seen to have the morally decisive higher-order capacities in question, even if not all human organisms realize those higher-order capacities.
3. Higher-Order Capacities and the Argument from Marginal Cases

3.1. The Main Dilemma

It is one thing to explain why there is a strong moral presumption against killing marginal cases. It is another thing to confront the AMC. For at this point in the argument, it seems like the above solution faces the following dilemma: either it admits that many non-human animals that we know of possess the higher-order capacities that marginal cases possess, or it does not admit this. If the above solution does admit this equivalence of capacities, then it must also admit that there is a strong moral presumption against killing these non-human animals. If the above solution does not admit this equivalence of capacities, then it is guilty of some morally objectionable form of “anthropocentrism” or “speciesism”.

The first horn of the dilemma runs as follows. The above solution seems to be committed to the idea that many non-human animals that we know of possess the higher-order capacities that marginal cases possess. After all, just as the technology of the future might enable us to induce genetic changes in a mentally deficient human organism so as to allow her to think like a normal adult human organism, so too the technology of the future might enable us to induce genetic changes in a non-human animal (e.g. a chimpanzee) so as to allow the non-human animal to think like a normal adult human organism. If the above solution is indeed committed to this idea, then there will be a strong moral presumption against killing these non-human animals. For the possession of the relevant set of higher-order typical human capacities (remember, I have focused on the capacity to think merely because it is a convenient illustration) is sufficient to generate the strong moral presumption against killing, whether or not the individual possessing this set is member of the human species.
The second horn of the dilemma runs as follows. If the above solution does not accept the idea that many non-human animals possess the higher-order capacities that marginal cases possess, then it is guilty of some morally objectionable form of “anthropocentrism” or “speciesism”. This is because it attempts to draw an arbitrary metaphysical or moral line between the humans and non-humans, or between our species and other species. Such line-drawing presumably commits one to the same sort of metaphysical and moral arbitrariness that other such line-drawing—such as sexism and racism—commits one to.

If the above solution is indeed forced into this dilemma, this would re-emphasize the resilience of the AMC, since the possession of a certain set of higher-order capacities turns out to be just the sort of criteria that plugs in nicely to Tom Regan’s constructive argument mentioned above. Although an appeal to higher-order capacities does have the benefit of allowing us to say what many want to say about marginal cases, it does this only at the cost of admitting that there is a strong moral presumption against killing certain non-human animals that we are aware of.

3.2. The First Horn of the Dilemma

3.2.1. Tooley’s Cat, Boonin’s Spider, and McMahan’s Dog

Before explaining why the above solution is not committed to the first horn of the dilemma, it is important to begin by examining the writings of three philosophers who have, in their own ways, formulated arguments which best express the thrust of this first horn: Michael Tooley, David Boonin, and Jeff McMahan.

---

17 Tooley, 192.

18 Boonin, 24-25.

19 McMahan, 302-329.
First, Michael Tooley constructs a thought experiment in which a kitten gets an injection that makes the kitten capable of developing thought patterns just like normal adult human thought patterns. Although this thought experiment is part of a longer complex argument against the moral relevance of potentiality, it is the thought experiment itself (and not the longer complex argument) that is relevant to the first horn of the dilemma:

Suppose that at some time in the future a chemical is discovered that, when injected into the brain of a kitten, causes it to develop into a cat possessing a brain of the sort possessed by normal adult human beings. Such cats will be able to think, to use language, to make decisions, to envisage a future for themselves, and so on—since they will have all of the psychological capacities possessed by adult humans.  

The relationship between Tooley’s thought experiment and the first horn of the dilemma is this. The above solution says that if a future technology could transform an organism so that the organism possesses the immediate capacity to think, it follows that the organism had the higher-order capacity to think to begin with. Tooley’s thought experiment simply fills in the details by making the technology an injection and the organism a kitten. Thus the above solution would seem to imply that, in Tooley’s thought experiment, the kitten had the higher-order capacity to think to begin with. (And since the possession of this higher-order capacity to think is sufficient to generate the strong moral presumption against killing, there would be a strong moral presumption against killing the kitten, even before it received its injection.)

---

20 Tooley, 191.
A more explicit formulation of the first horn of the dilemma is found in David Boonin’s discussion of what he calls “the species essence argument.” Boonin summarizes a version of this argument taken from Stephen Schwartz:

1. A person is “a being who has the basic inherent capacity for thinking in the broadest sense regardless of how developed or blocked it is.”

2. “…it is an essential property of every living member of the species homo sapiens that it has the capacity to function as a person…”

3. “the capacity to function as a person confers on one a right to life…”

Therefore,

4. “being a member of homo sapiens does ensure that one has a right to life.”

The relevant part of Boonin’s criticism of this argument is his objection to the second premise:

The claim that every member of homo sapiens has the capacity to function as a person is false. There can, for example, be human fetuses with such severe deformities that they will never develop a brain capable of sustaining thought, or even any brain at all. These are human beings who have not even the capacity for functioning as a person and so are not persons on Schwarz’ definition of the term.

Boonin then considers, and quickly rejects, a possible reply that is very similar to the above solution:

---

21 Boonin, 23-24. There is also an additional part of the argument that says, “since every human fetus is a member of homo sapiens, it follows that every human fetus has a right to life” (24). But this additional part of the argument is not important for focusing on marginal cases.

22 Ibid., 24.
One could, I suppose, characterize such a fetus as a person whose capacity for thought simply happens to be “blocked” by a contingent fact about its head. But then it is difficult to see why we should not also call the spider crawling up my window a person. If he were able to develop a big enough brain, he too would be able to function as a person, so he is simply a person whose capacity is blocked by the fact that he will never have a large enough brain.\textsuperscript{23}

The relationship between Boonin’s discussion and the first horn of the dilemma is this. The above solution claims that an organism can still possess a higher-order capacity to think even if certain physical conditions prevent (or ‘block’) that capacity from being realized. Boonin invites us to consider how the only conditions that prevent (or ‘block’) a spider from having the immediate capacity to think are certain physical conditions. Thus the above solution would seem to imply that a spider has the higher-order capacity to think. (And since the possession of this higher-order capacity to think is sufficient to generate the strong moral presumption against killing, there would be a strong moral presumption against killing the spider.)\textsuperscript{24}

Finally, Jeff McMahan constructs a thought experiment in which a dog gets genetic therapy that makes the dog capable of developing thought patterns just like the thought patterns of normal human adults. McMahan uses this thought experiment to make a fairly explicit statement of the first horn of the dilemma. Since McMahan’s thought experiment includes a technological element (like Tooley’s thought experiment) and is explicitly used to make the

\begin{footnotes}
\item[23] Ibid., 24.
\item[24] Boonin also objects to the second premise because of “human beings who permanently lose their capacity for functioning as a person, such as those whose higher brain regions are irreparably destroyed” (24). Presumably, any reply to this objection along the lines of the above solution would be rejected for the same reasons.
\end{footnotes}
point of the first horn of the dilemma (like Boonin’s discussion), the remainder of this subsection will be spent giving McMahan’s thought-experiment a careful exposition.

There are two basic ideas behind McMahan’s thought experiment. (1) If a human being with a genetically determined cerebral deficit can nevertheless count as having the “intrinsic” potential to develop the cognitive characteristics of a mature human being, in virtue of the fact that the genetic therapy of the future can enable such individuals to overcome these genetically determined cerebral deficits, then a dog can likewise count as having the “intrinsic” potential to develop the cognitive characteristics of a mature human being. (2) If having this “intrinsic” potential is what determines something’s moral status, then there is no difference in the moral status of the dog and the moral status of the impaired human.

To appreciate this thought experiment, it is important to locate it in the flow of McMahan’s longer discussion of “whether the [human] fetus’s potential can plausibly be regarded as a basis for respect” or more generally “as a basis for moral status.”

McMahan believes that if the fetus’s potential to become a person is to be a basis for moral status, then this potential must be grounded in the intrinsic properties of the fetus. McMahan then argues that the fetus’s potential to become a person is not grounded in the intrinsic properties of the fetus.

He begins by distinguishing three cases. First, The Normal Fetus is “a developed fetus that is in every way normal and healthy”. Second, The Fetus with a Chemical Deficit is a fetus whose brain is developing normally except that it is deficient in a certain chemical (e.g., a neurotransmitter) without which the fetus will never be a person, because without this chemical,

---

25 McMahan, 309.

26 McMahan uses the term “person” as what some logicians call a phase sortal, like the term “adolescent”: an entity can fail to be a person at one time, become a person, and then fail to be a person again. And the grounding he has in mind seems to be epistemic: “there must be something about the fetus now that justifies the claim that it is a potential person” (309, emphasis mine).
the fetus, even when grown up, will have cognitive capacities that do not surpass the cognitive capacities of a chimpanzee. Third, *The Fetus with Cerebral Deficits* is a fetus whose brain is developing abnormally in that if it continues on its developmental path, the fetus will never be a person. McMahan then asks a pivotal question: “On what basis might it be claimed that these latter two fetuses are potential persons?”

McMahan rejects the answer that “these fetuses are both the sort of entity that normally becomes a person,” since

To point out that these two fetuses are entities of a kind whose normal members tend to become persons is not to show that they have the potential to become persons. It is only to note that normal members of the kind have that potential. But these two members of the kind are not normal members. And their abnormality is precisely that they lack something that is necessary for them to become persons.

McMahan considers a second answer: “In order for X to have the potential to become a Y, it must be possible for X to become a Y.” McMahan is willing to admit that this answer allows for The Fetus with a Chemical Deficit to count as a potential person. After all, he says, we admit that a seed is a potential plant even when it is not given the water it needs to grow into one. But McMahan is not willing to admit that this answer allows The Fetus with Cerebral Deficits to count as a potential person, and he presents two parallel thought experiments to show why.

His first thought-experiment can be summarized as an argument:

---

27 Ibid., 310.

28 Ibid.

29 Ibid., 311.
(1) A child born without eyes a thousand years ago did not have the potential for sight.
(2) A child born without eyes in a world in which eye transplants are routinely performed would have the potential for sight.
(3) But there is no intrinsic difference between these children. Therefore,
(4) The difference between the potentials of these children cannot be a matter of their intrinsic properties.

The second thought experiment has the same form as the first:

(1*) A fetus with cerebral deficits in a world, such as ours, in which cerebral augmentation is not possible, does not have the potential to become a person.
(2*) A fetus with cerebral deficits in a world in which cerebral augmentation through genetic therapy is possible would have the potential to become a person.
(3*) But there is no intrinsic difference between the fetus with cerebral deficits in our world and the fetus in the world in which cerebral augmentation is possible. Therefore,
(4*) The difference between the potentials of these fetuses cannot be a matter of their intrinsic properties.

McMahan characterizes the genetic therapy, in premise (2*), as:
A form of genetic therapy that, if administered to the fetus with cerebral deficits, would cause it [to] grow the cerebral tissues necessary for normal cognition, and…the growth of these tissues would be identity-preserving.  

McMahan thinks the upshot of these thought experiments is this. If The Fetus with Cerebral Deficits has the potential to become a person, then this potential is not grounded in the fetus’s intrinsic properties. Therefore, in the case of The Fetus with Cerebral Deficits, the fetus’s potential to become a person cannot be a basis for the fetus’ moral status.

Before going on to explain the rest of McMahan’s argument, it is important to pause and notice something. If the analysis of higher-order capacities presented above is correct, then each of these thought experiments begins with a faulty first premise. Surely, the child without eyes does have the potential for sight even in worlds where this potential never gets realized. A similar remark applies to the fetus with the cerebral deficit. Thus, in the second thought-experiment, McMahan should have argued as follows:

(2*) A fetus with cerebral deficits in a world in which cerebral augmentation through genetic therapy is possible would have the potential to become a person.

(3*) But there is no intrinsic difference between the fetus with cerebral deficits in our world and the fetus in the world in which cerebral augmentation is possible.

Therefore,

~(1*) A fetus with cerebral deficits in a world, such as ours, in which cerebral augmentation is not possible, does have the potential to become a person.

30 Ibid.
This objection against McMahan’s argument relies upon the idea that an entity’s potentials depend only upon its intrinsic properties plus the laws of nature, and upon the idea that the worlds in (2*) and (3*) are not just any two possible worlds, but are alternative histories of (or alternative times in) the actual world. McMahan’s original argument, on the other hand, seems to assume that an entity’s potentials depend upon what is technologically possible at a given time in a given world.

But McMahan has an answer to this type of objection. For his very next move evaluates what he calls a “more radical” view of potential that would block his argument from succeeding. He characterizes this “more radical” view as follows:

As long as it is physically possible for the fetus with cerebral deficits to develop the cognitive capacities that are constitutive of personhood in a way that is identity-preserving, that fetus counts as a potential person…what the potential essentially consists in is an intrinsic receptivity to an identity-preserving transformation into a person. This is a fact about the fetus itself: that it is the sort of thing that can in principle be transformed into a person while continuing to exist.\textsuperscript{31}

The thought experiment involving the dog now enters the argument as an objection to this “more radical” view of potential:

If it is physically possible, through some as-yet-undiscovered form of genetic therapy, to augment a defective fetus’s brain in a way that will enhance its future cognitive capacities, it is surely physically possible to achieve the same result in an animal—for example, a dog. If, therefore, we claim that a fetus with cerebral

\textsuperscript{31} Ibid.
deficits is a potential person on the ground that it is physically possible for its brain to develop in ways that would be identity-preserving and would overcome or repair the deficits, we must concede that a dog is a potential person for the same reason. And if we claim that the fetus’s potential to become a person is a basis for moral status (because it is grounded in a suitably intrinsic receptivity to transformation), we must concede that a dog has an equivalent status, other things being equal. Since, however, no one would (or should) accept that dogs are potential persons with a moral status appropriate to their nature as such, we must abandon the broad conception of potential that implies that they are.  

McMahan’s line of reasoning can be reconstructed as a reductio:

(1) X is a potential person at time t if and only if it is physically possible for X at time t to undergo an identity-preserving transformation into a person

[Definition of “potential person”]

(2) It is physically possible for a living thing that is a dog at time t, but not a person at time t, to undergo an identity-preserving transformation into a living thing that is a person at a later time t*. [Assumption]

Therefore,

(3) A living thing that is a dog at time t, but not a person at time t, is a potential person at time t. [From (1) and (2)]

(4) If X is a potential person at time t, then X has moral status at time t.

[Assumption]

Therefore,

32 Ibid., 312.
(5) A living thing that is a dog at time t, but not a person at time t, has moral status at time t. [From (3) and (4)]

(6) But a living thing that is a dog at time t, but not a person at time t, does not have moral status at time t. [Assumption]

Therefore,

(7) Statement (4) is false. [From (1), (2), (5), and (6)]

McMahan concludes that, no matter how the genetic therapy is described, there is always going to be the non-human animal as a counterexample:

There is, I believe, no basis for claiming that the fetus with cerebral deficits has the potential to become a person that does not also imply that a dog has that potential...we should abandon the ambition to include fetuses with cerebral deficits within the category of potential persons.\(^{33}\)

The relationship between McMahan’s discussion and the first horn of the dilemma is this. McMahan’s reductio can be easily recast using the concepts of higher-order capacities and serious moral status:

(1a) X has a higher-order capacity to think if and only if it is physically possible for X at time t to undergo an identity-preserving transformation into an entity with the immediate capacity to think [Definition of “higher-order capacity to think”]

(2a) It is physically possible for a living thing that is a dog at time t, but not an entity with the immediate capacity to think at time t, to undergo an identity-

\(^{33}\) Ibid.
preserving transformation into a living thing that is an entity with the immediate capacity to think at a later time \( t^* \) [Assumption]

Therefore,

(3a) A living thing that is a dog at time \( t \), but not an entity with the immediate capacity to think at time \( t \), is an entity with a higher-order capacity to think at time \( t \). [From (1a) and (2a)]

(4a) If \( X \) is an entity with a higher-order capacity to think at time \( t \), then \( X \) has serious moral status at time \( t \). [Assumption]

Therefore,

(5a) A living thing that is a dog at time \( t \), but not an entity with the immediate capacity to think at time \( t \), has serious moral status at time \( t \). [From (3a) and (4a)]

(6a) But a living thing that is a dog at time \( t \), but not an entity with the immediate capacity to think at time \( t \), does not have moral status at time \( t \).

[Assumption]

Therefore,

(7a) Statement (4a) is false. [From (1a), (2a), (5a), and (6a)]

In summary, then, the first horn of the dilemma claims that if certain marginal humans are allowed to count as possessing the higher-order capacity to think, in virtue of what the technology of the future can enable them to do, then certain non-human animals we are aware of must also be allowed to possess the higher-order capacity to think.
3.2.2. It Is Not Possible To Perform These Transformations on a Non-Human Animal.

The first horn of the dilemma relies upon the assumption that it is possible to perform these transformations on a non-human animal: more precisely, the assumption is that the non-human organisms we are aware of could be given the immediate capacity to think like a normal human adult while still continuing to exist. This assumption is present in the three authors just examined. For example, Tooley’s claim that the injection “causes it [the kitten] to develop into a cat possessing a brain of the sort possessed by normal adult human beings”\(^34\) relies upon the assumption that the living thing that is a kitten before the injection is the same living thing as the living thing after the injection. Likewise, Boonin’s claim that “If he [the spider crawling up my window] were able to develop a big enough brain, he too would be able to function as a person…”\(^35\) relies upon the assumption that the living thing that is a spider before the increase in brain size is the same living thing as the living thing after the increase in brain size. Finally, McMahan assumes that the living thing that is the dog before the genetic therapy is the same living thing as the living thing after the genetic therapy: “If, therefore, we claim that a fetus with cerebral deficits is a potential person on the ground that it is physically possible for its brain to develop in ways that would be identity-preserving and would overcome or repair the deficits, we must concede that a dog is a potential person for the same reason.”\(^36\)

My strategy for replying to the first horn of the dilemma is to deny this assumption. This strategy would claim that it is not possible to perform these kinds of transformations on a non-human animal: more precisely, it is simply not true that the non-human animals we are aware of

\(^{34}\) Tooley, 191.

\(^{35}\) Boonin, 24.

\(^{36}\) McMahan, 312.
could be given the immediate capacity to think like a normal human adult while still continuing to exist.

This envisaged strategy could be fleshed out in one of two ways. First, it could be making a claim about the limits of technology. It may well be the case that, as a matter of fact, the technology of the future will never be able to (for example) take a dog, produce certain genetic changes in it, and end up with an organism that thinks just like a normal human. This is because the sorts of changes envisioned would by physically impossible, due to the way the bodies of organisms work. If the cells in a dog brain reject or attack the human tissues that are injected into that dog brain, the dog will not be able to develop the immediate capacity to think like a human thinks. I am inclined to think that this first way of fleshing out the strategy is more promising than McMahan is willing to allow. But I really do not know enough about the biology involved in such tissue transplants to say for sure.

The second way of fleshing out this strategy is to claim that, even if the tissue injected into the dog was not rejected, still, as a matter of metaphysics, the genetic changes done to the dog would not transform the dog in an identity-preserving way. I would like to explore in a bit more detail the two basic ideas that would need to be accepted to sustain this way of fleshing out this strategy:

(1) All organisms are members of natural kinds, which means, among other things, that certain sorts of changes to a given organism will be identity-preserving changes and other sorts of changes to that organism will be identity-undercutting changes.

(2) The changes required to give the ability to think to any of the non-human animals that we are aware of would be identity-undercutting changes.
Although these ideas are related, they are also importantly distinct. One could accept the first but not the second, and vice-versa.

There are different ways of defending (1). One way is to rely on the claim that biological species are natural kinds. Since all organisms are members of biological species, it would directly follow that all organisms are members of natural kinds. Unfortunately, this way of defending (1) runs into what is sometimes called the ‘species problem.’ The proper analysis of the concept of a species is a long-standing debate among biologists and philosophers of biology, and this debate shows no signs of being solved in the foreseeable future. There are at least a dozen different rival accounts of what it means for \( x \) to be a member of the same species as \( y \).\(^{37}\) While some accounts emphasize the phenotypic similarities between \( x \) and \( y \), others emphasize their genotypic similarities, or reproductive potential, or geographical proximity, or propinquity of descent, or some other common feature and/or relationship. Because of this intractable disagreement, defending (1) using the concept of a species would require being very clear about which concept of a species is in view. It would also require justifying this particular concept of a species rather than the other concepts on offer.

Fortunately, there is an easier way of defending (1) that avoids the concept of a species entirely. This is to rely on the claim that there is some unique way of constructing a biological taxonomy such that the most basic categories of this taxonomy are natural kinds. Since all organisms fall into the categories of this biological taxonomy, it would directly follow that all organisms are members of natural kinds.

Many philosophers think that the members of a natural kind possess something intrinsic in common with each other. Elliot Sober, for example, after stating that “a standard philosophical view about natural kinds” is essentialism, which “holds that each natural kind can be defined in terms of properties that are possessed by all and only the members of that kind,”\(^{38}\) goes on to explain that an essentialist definition of gold “must cite a property that is intrinsic to gold things; the cited property [in this case, atomic number 79] does not require that any relations obtain among gold things.” Likewise, T. E. Wilkerson claims that one of the conditions that must be met by any interesting account of natural kinds is that “members of natural kinds have real essences, intrinsic properties that make them members of the relevant kind, and without which they could not be members of the relevant kind.”\(^{39}\)

If humans and non-human animals are members of different natural kinds, then it follows that humans do share something special in common with each other that they do not share with non-human animals: namely, a real essence. McMahan’s argument is looking for an intrinsic difference between dogs and humans; the idea that humans and dogs are natural kinds presents an intrinsic property that a human being possesses and a dog does not. Human beings share a real essence; dogs share a real essence; these real essences are different from each other, but each real essence is intrinsic to the respective organisms that possess it.

However, in order for a doctrine of natural kinds to adequately answer the first horn of the dilemma, it needs to be careful not to identify an organism’s real essence with its genome. For consider again the fetus with cerebral deficits. If her set of potentials or dispositional properties is identified with her real essence, and if her real essence is identified with her


genome, then it follows that she is not a potential person. After all, her genetic material was the problem to begin with, since it seemed to undercut the idea that she is a potential person.

The contemporary approach to natural kinds, as represented by Kripke and Putnam, usually attempts to identify a thing’s real essence with some structural feature of that thing. For example, if something looks like gold but does not have atomic number 79, then it is not gold; likewise, if something is functionally just like water but does not consist of H$_2$O, then it is not water. When this understanding of natural kinds is introduced to do work in constructing a biological taxonomy, it will end up emphasizing a genetic real essence. For example, if something looks like a duck, walks like a duck, and quacks like a duck, this still does not guarantee it is a duck. It needs to possess the genetic real essence (whatever precisely it is) that all real ducks have. Likewise, if “human” is a natural kind like “gold” or “water”, then unless something possesses the genetic real essence (whatever precisely it is) that all real humans have, that thing will not be a human.

This is a problem that even an advocate of biological natural kinds must confront. For example, T. E. Wilkerson argues that natural kinds exist in biology, but that species are not natural kinds. Wilkerson considers an objection to a Kripke-Putnam approach, namely, the objection that “species are not uniquely determined by genetic constitution” or, in other words, “genetic real essences of natural kinds do not exist.”\footnote{T. E. Wilkerson, “Species, Essences, and the Names of Natural Kinds,” \textit{Philosophical Quarterly} 43 (1993): 7.} This objection to a Kripke-Putnam approach can point to various examples in real-life biology: there may be a good deal of genetic variation between the parts of the same individual (in plants) or between the members of the
same biological kind; conversely, there may be a good deal of genetic similarity between closely related species.\footnote{Ibid., 7-8.}

One reply to this objection might be to claim that: “the genetic feature we are looking for is a structural feature of the genetic material—for example, the number of chromosomes peculiar to each species.”\footnote{Ibid., 8.} But Wilkerson rejects this reply, since not all humans have 23 pairs of chromosomes (for example, those with Down’s syndrome have an extra chromosome) and many plants have multiple sets of chromosomes (a feature called polyploidy). Wilkerson’s conclusion is that “the more we attempt to isolate the genetic features that determine biological species, the more hopeless the task becomes.”\footnote{Ibid.} The problem, as he restates it, is this:

If natural kinds are determined by real essences, and if species are good examples of natural kinds, then we appear to have produced a contradiction, since species are not determined by real essences.\footnote{Ibid., 10.}

An obvious solution to this problem is to claim that species are determined by real essences, but that these real essences are not genetic. Wilkerson’s own solution, however, is to keep the real essences genetic and to simply increase the number of natural kinds in biology. Here is how he summarizes it:

There are natural kinds. Each natural kind is determined by a real essence, a property or set of properties necessary and sufficient for membership of the kind in question. The real essence in turn grounds the causal powers of individual
members of the kind. Biological natural kinds are determined by genetic real essences which are causally responsible for the behaviour of individual members of the kind. But, since there is considerable interspecific genetic similarity and intraspecific genetic variation, there are far more biological natural kinds than species.\textsuperscript{45}

Whatever the merits of Wilkerson’s proposal may be for systems of biological taxonomy, his proposal is not helpful for explaining how a human with a genetic handicap can be a member of the same natural kind as you and I. For if the real essence that grounds membership in biological natural kinds is equated with the genome, then there would be as many natural kinds as there are genetically diverse individuals. But this proliferation of natural kinds is tantamount to giving up on the doctrine of natural kinds in biology. Imagine a “biological periodical table”, if you will, that is designed to parallel the periodic table of the elements. Just as gold has its own box with its own atomic number, so each genetically distinct individual would have its own box with its own (let us say) genetic number. If a population of genetically identical clones existed, then of course many different instances of the genetic real essence would exist. But everyone else would be all by himself or herself in the biological periodical table. Consequently, if one wants to explain why the human organism with a cerebral deficit and the dog are members of different natural kinds, Wilkerson’s proposal works just fine. But if one wants to explain why the human organism with a cerebral deficit and the normal human organism are members of the same natural kind, the real essence they share must go deeper than, and be different from, a purely genetic constitution. Otherwise, the human with a genetically determined cerebral deficit will be a member of a different natural kind from the rest of us.

\textsuperscript{45} Ibid., 16.
There are at least three reasons for believing that an organism’s real essence is deeper than, and different from, its genetic constitution. First, the following thought experiment gives us a reason for not identifying an essence with a genotype. Suppose you have a one-celled organism. On Monday, you put it into a state of suspended animation and extract its DNA. For the next three days, you tinker with its DNA in a separate part of the lab. On Friday, you reinsert the DNA back into the organism and thaw out the organism. The organism exists even when its DNA has been removed. It continues to have its essence even when its genotype has been removed. Therefore its essence is not its genotype. Second, the possibility of an organism surviving a small genetic mutation suggests that an organism’s genetic code should not be equated with its real essence. Third, the fact that it is possible to have more than one genome within the same organism also suggests that an organism’s genetic code should not be equated with that organism’s real essence.

The essence of an organism is partly characterized, but not fully exhausted, by making reference to phenotypes and genotypes. The essence is phenotypic in the sense that the essence is characterized by the range of phenotypes an organism can exhibit while still remaining the same organism. But the essence is not strictly phenotypic, because the characterization referencing the phenotypes is subjunctive: the important thing is not what phenotypes the organism actually has, at this very moment, but rather what phenotypes it would have if it were put in this or that set of circumstances. Consequently, an organism can still have its essence before any of its possible phenotypes have been expressed, and the capacity to realize a given phenotype is more central to its essence than the actual expression of that phenotype. The essence is genotypic in the sense that a genetic structure of an organism is often part of the
physical basis for an active higher-order capacity. But the essence is not *strictly* genotypic, because it is not to be identified with the genetic structure of an organism.

Idea (2), the second basic idea behind this strategy, was that the changes required to give the ability to think to any of the non-human animals that we are aware of would be *identity-undercutting* changes. To grasp what this means, consider first the concepts of incapacities and essential incapacities. An incapacity is simply an inability to do something: for example, I have an incapacity for speaking Chinese, an incapacity for remembering your experiences, and an incapacity for omnipresence, omniscience, and omnipotence. An essential incapacity is an essential inability to do something, an inability that one cannot lose while remaining who one is: for example, I have an essential incapacity for remembering your experiences, since if I were to lose this incapacity, I would become someone else: namely, you. Likewise, I have an essential incapacity for omnipresence, omniscience, and omnipotence, since if I were to lose this incapacity, I would become someone else: namely, God. However, my incapacity for speaking Chinese is not an essential incapacity, since I can lose it while remaining who I am.

Philosophers have sometimes claimed that a property can be essential to an object in one of two ways. First, a property is *kind-essential* if its being had by an individual is needed for that individual to belong to a particular kind. Second, a property is *individually-essential* if the individual that has it could not have existed without having it. Some philosophers have thought that, if a given property is kind-essential to an individual, then that property is also individually-essential to that individual. Others prefer to make room for the thought that a given property might be kind-essential to an individual without being individually-essential to that individual.

Idea (2) should be read as the claim that each of the particular non-human animals we are aware of has an individually-essential incapacity to obtain the immediate capacity to think. As
before, however, let me emphasize that I focus on the capacity to think just for the sake of convenience. What matters is that each of the non-human animals we are aware of has an individually-essential incapacity to obtain the whole set of typical human immediate capacities, one of which is thinking.

Another way of grasping the idea behind (2) is to invoke the concept of a modal boundary. There are innumerable ways that an entity can be modified, but an entity’s modal boundary is the metaphysical line beyond which that entity cannot go. For example, the caterpillars we are aware of would cross their modal boundary if they changed into puppies. But the caterpillars we are aware of do not cross their modal boundary merely by changing into butterflies. On the other hand, if we discovered a group of organisms that looked like caterpillars, but that changed into puppies, we would not say that these organisms were caterpillars that had crossed their modal boundaries. We would say that they were not caterpillars at all: perhaps we would call them “scatterpillars”. Even the character from Greek mythology named Proteus had his modal boundaries: even though he could take on the typical capacities of a donkey, and then take on the typical capacities of a human, and so on, still, he could not become omnipresent, omniscient, and omnipotent.

One of the best reasons for believing in basic idea (2) comes from attempting to construct a temporary change argument in which a human temporarily changes into a non-human animal. It seems that there are certain sorts of apparent temporary changes that are not temporary changes at all, but are rather instances of one individual ceasing to exist and another individual coming to exist. Recall the example above where a scientist accidentally steps in front of a machine while it is emitting A-rays. Imagine that the administration of A-rays, instead of transforming this scientist into an individual with a genetic disability, instead transforms the
scientist into a dog. I believe that the scientist before the administration of A-rays is *not* the same organism as the dog after the A-rays. The apparent transformation of the scientist was actually not a case of transformation at all, but one of annihilation and creation: the original scientist was annihilated (or disembodied, for those who believe in the possibility of disembodied existence), and a brand new organism was created. And this is *so even if* the machine is able to emit B-rays that “transform” the dog into a human organism (who we may call “the resultant scientist”) a few minutes later. The B-rays would not actually transform the dog, but would annihilate (or disembody) it and create the resultant scientist in its place. The resultant scientist would not be the original scientist.

The hypothetical example of “scatterpillars” tells us something important. If we came across a particular non-human animal that appears to have been transformed, in an identity-preserving way, so that it now seems to have the typical human capacities, such as thinking, we have two options for describing this. On the one hand, we could say that this apparently identity-preserving transformation was actually identity-undercutting, and that the original animal ceased to exist (or at least ceased to exist *right there*, if one wishes to make room for disembodied non-human animal souls) at the moment a new organism began to exist. On the other hand, we could say that this apparently identity-preserving transformation was indeed genuinely identity-preserving, and that we were mistaken in our original thought that this particular non-human animal possessed an individually-essential incapacity to obtain the typical human capacities. The individual that changed was a “scatterpillar” after all, and not a caterpillar.

Some might be reluctant to embrace (2) for reasons that have already been dealt with in earlier chapters. For example, the claim that the technology is being applied to a living thing at every moment during the process of transformation is not a decisive reason for thinking that the
original thing and the resultant thing are the same. Denying the numerical identity of the original dog and the resultant organism with the immediate capacity to think is simply the third step in a natural progression, the first two steps of which have already been seen to be acceptable. The first step in this progression (from Chapter Two and the present chapter) dealt with re-arranging an island of raw materials so that it came to constitute a human organism. The mere facts that the original individual was spatiotemporally continuous with and mereologically indistinguishable from the resultant individual were not sufficient for making the original individual numerically identical to the resultant individual. The second step in the progression (from Chapter Four) dealt with changing a part of a human organism so that it came to constitute a new human organism. Here, as in the first step, the original individual was spatiotemporally continuous with and mereologically indistinguishable from the resultant individual. But in addition, the original individual and the resultant individual were both alive and were genetically identical to one another (i.e. they had the same genetic code). Yet these facts were not sufficient for making the original individual numerically identical to the resultant individual. Whether one is a materialist or a dualist of one stripe or another, the lesson learned from these first two steps is this: whatever it is that provides the locus of identity through time, it is not the elements of spatiotemporal continuity, mereological indiscernability, or genetic indiscernability. Nor is it the combination of these elements under conditions where both the original individual and the resultant individual are alive. Therefore, one cannot rely upon these elements when one claims that a dog before the application of technology is the same individual as the resultant individual after the application of technology.

There are at least three objections to this strategy. First, it might be objected that any doctrine of natural kinds in biology is inconsistent with the theory of evolution. In reply, this
objection is mistaken for a reason Elliot Sober gives in his evaluation of different evolutionary arguments against the idea that *species* are natural kinds. Even though the above discussion avoided the species concept, and even though Sober himself does not think species are natural kinds, relying upon him here is instructive because the point he makes is relevant to evolutionary arguments against essentialism in biology. One such argument runs as follows: (1) natural kinds are immutable; (2) species evolve; therefore (3) species are not natural kinds. Sober replies that, just as an atom smasher can transform lead into gold without undermining the idea that the chemical elements have immutable essences, “the fact that a population belonging to one species can give rise to a population belonging to another species does not refute essentialism about species.”

And just a page later, he says:

> In general, essentialism is a doctrine that is compatible with certain sorts of vagueness. The essentialist holds that the essence of gold is its atomic number. Essentialism would not be thrown into doubt if there were stages in the process of transmuting lead into gold in which it is indeterminate whether the sample undergoing the process belongs to one element or to the other. I suspect that no scientific concept is *absolutely* precise; that is, for every concept, a situation can be described in which the concept’s application is indeterminate. Essentialism can tolerate imprecisions of this sort.

Sober’s point about the transmutation of elements is relevant to the mutation of organisms. It seems that essentialism about chemical elements does not stand or fall depending on one’s theory about the origin and historical evolution of these elements. For example, if it were to be

---

46 Sober, 146-7.

47 Ibid., 148.
discovered that all of our present elements originally emerged, very gradually, from a sort of
primordial stuff, this would not throw into doubt the theory of natural kinds about gold. But then
why should it be any different for biological organisms? It would seem that essentialism about
biological organisms does not stand or fall depending on one’s theory about the origin and
historical evolution of these organisms. For example, if it were to be discovered that all of our
present organisms originally emerged, very gradually, from a sort of primordial soup, this should
not throw into doubt the theory of natural kinds about humans.

A second sort of objection is that this strategy is just very difficult to believe, because it
does seem possible to imagine a non-human organism being transformed, in an identity-
-preserving way, in such a way that it comes to think just like a human being. Such
transformations pop up in all the time in various kinds of literature around the world. Consider
the Biblical story of Balaam’s donkey:

Balaam got up in the morning, saddled his donkey and went with the princes of
Moab. But God was very angry when he went, and the angel of the LORD stood
in the road to oppose him. Balaam was riding on his donkey, and his two servants
were with him. When the donkey saw the angel of the LORD standing in the road
with a drawn sword in his hand, she turned off the road into a field. Balaam beat
her to get her back on the road. Then the angel of the LORD stood in a narrow
path between two vineyards, with walls on both sides. When the donkey saw the
angel of the LORD, she pressed close to the wall, crushing Balaam’s foot against
it. So he beat her again. Then the angel of the LORD moved on ahead and stood
in a narrow place where there was no room left. When the donkey saw the angel
of the LORD, she lay down under Balaam, and he was angry and beat her with his
staff. Then the LORD opened the donkey’s mouth, and she said to Balaam, “What have I done to you to make you beat me these three times?” Balaam answered the donkey, “You have made a fool out of me! If I had a sword in my hand, I would kill you right now.” The donkey said to Balaam, “Am I not your own donkey, which you have always ridden, to this day? Have I been in the habit of doing this to you?” “No,” he said.”

Whether or not the event described in this story was historical, it is surely metaphysically possible. And yet this event seems to embody precisely the sort of identity-preserving transformation (of the donkey) that the above view of natural kinds says is not metaphysically possible. The organism before the divine intervention and the organism after the divine intervention are the same organism: “Am I not your own donkey, which you have always ridden to this day?”

One way to reply to this objection is to say that, although the events in these stories are intelligible, the proper interpretation of these events need not admit that the individuals in these stories retained their identity through time. Of course, it may have seemed to Balaam, and his donkey, that the organism after the divine activity was the same individual as the donkey before the divine activity. After all, the resultant organism had at least apparent memories of being beaten, and supposedly it had similar mental states to the mental states of the original donkey. But these sorts of considerations—apparent memories, similar mental states—are notoriously insufficient for genuine identity through time. Perhaps it is preferable to say that Balaam and his donkey were mistaken.

---

But another way to reply to this objection is to say that, in this particular case, the donkey did not have an individually-essential incapacity to think. Saying this does not commit one to the view that all donkeys have a passive higher-order capacity to think. It only commits one to the view that Balaam’s donkey had a passive higher-order capacity to think. And if it turned out that Balaam’s donkey had a set of higher-order typical human capacities, then, in the case of Balaam’s donkey, we would have to admit that it had serious moral status. When confronted with the case of Balaam’s donkey, we employ the same move that we employ when confronted with a scatterpillar. In both cases, we simply admit that the individual substance was not the sort of thing we had assumed it was at the beginning, and that it did not have the modal boundary that we had originally assumed it had.

A third objection to this strategy agrees with basic idea (1) but denies basic idea (2). There are various alternatives for individuating natural substances. For example, perhaps the relevant substantial kind that a spider is a member of is not spider but rather living animal. On this view, as long as a living animal that is a spider remains a living animal, then that living animal will still continue to exist no matter what sorts of changes it undergoes: the living animal that is a spider could receive a huge brain, and learn philosophy, and indeed could even become a dog or a cat or a human. Or perhaps the relevant substantial kind that a spider is a member of is not living animal, but living organism. On this view, a living organism that is a spider could become an oak tree or an amoeba while still continuing to exist.

The best reply to this objection is that choosing from among these alternatives for the relevant substantial kind—living animal, living organism, and so on—is fundamentally and inevitably a matter of testing our modal intuitions with thought experiments of various kinds. Each alternative has its interesting results, or philosophical bullets to bite, and the task of
choosing between them is one of deciding which bullets are the least worst to bite. I believe that the alternative I have suggested is at least as good as, and indeed, better than, the ones suggested by this objection. Does it really seem plausible to claim that the same living organism that is now a spider could one day become an oak tree, an amoeba, a human, and a dog?

3.3. The Second Horn of the Dilemma

The second horn of the dilemma claims that the above solution is “anthropocentric” or “speciesist” in some morally objectionable way. But this objection is mistaken, and several of the reasons it is mistaken have already been touched upon in one way or another in the discussion of the first horn of the dilemma. For example, the strategy discussed above deliberately avoided the concept of a biological species when defending basic idea (1). The discussion of real essences and genetic codes was neutral between humans and non-human animals. Likewise, when defending basic idea (2), the idea of a temporary change was revised to take account of the fact that humans, too, can only undergo certain sorts of changes and retain their identity through time. Certain kinds of identity-undercutting changes can happen to living human organisms.

There are at least four more reasons why the above solution is not anthropocentric or speciesist in any morally objectionable sense. First, the above solution does not say that there is not a strong moral presumption against killing the non-human animals we are aware of. The present account leaves this question completely open, and is fully compatible with the claim that there is, in fact, a strong moral presumption against killing the non-human animals we are aware of. Indeed, the above solution is fully compatible with the claim that the strong moral presumption against killing the non-human animals we are aware of is generated by the higher-
order capacities of these animals. Perhaps the set of typical dog capacities generates a strong moral presumption against killing dogs. Nothing this essay has said precludes this possibility.

Second, the above solution is willing to admit that, after certain identity-undercutting changes have taken place in a non-human organism, the resulting organism now possesses the set of higher-order capacities which, in humans, is what grounds the strong moral presumption against killing. There is a strong moral presumption against killing these resultant organisms, whether or not there was a strong moral presumption against killing the original organisms. For example, if the strong moral presumption against killing is generated by the possession of a higher-order capacity to think, then there is a strong moral presumption against killing the organism that results from altering Tooley’s kitten and the organism that results from altering McMahan’s dog and the organism that results from altering Boonin’s spider.

Third, if there are non-human animals with exactly the same higher-order capacities as humans—for example, if we discovered a race of “shumans” on Mars whose members possess the same set of higher-order capacities as you or I, even though they had ZNA instead of DNA in their cells—then there would be a strong moral presumption against killing these non-human animals, even though they were not part of our species. This essay has argued that whatever possesses the relevant set of higher-order capacities, whether it is part of our species or not, is such that there is a strong moral presumption against killing it.

Fourth, if there is, in fact, a strong moral presumption against killing some of the non-human animals we are aware of because of the set of higher-order capacities they possess, the above solution would be useful in defending the moral status of the marginal cases of such non-human animals. For just as marginal humans still have the distinctively human set of higher-order capacities, so too the marginal chimpanzees still have the distinctively chimpanzee set of
higher-order capacities. If it turns out that there is a strong moral presumption against killing normal adult chimpanzees because of the higher-order capacities they have, then there will also be a strong moral presumption against killing marginal chimpanzees—those suffering from temporary changes, brain damage, mental retardation, and so on.
CONCLUSION

The main argument of this essay had three steps:

1. If an entity is human, it has a set of typical human capacities.

2. If an entity has a set of typical human capacities, it has serious moral status.

Therefore,

3. If an entity is human, it has serious moral status.

This argument, I have claimed, is sound as long as typical human capacities are understood to apply to (a) both “active” capacities and “passive” capacities, (b) both “lower-order” capacities and “higher-order” capacities, and (c) only “general” capacities that can be shared by distinct individuals, rather than “specific” capacities that necessarily apply to only one individual (such as my capacity to remember my very own experiences).

I have argued, along the way, for the view that “persons” like you and I are human organisms, although I defined “human” in such a way that it leaves open the possibility that a human organism might survive its biological death. Finally, I have claimed that the two main problem areas that threaten to undermine each of the two premises in the main argument can be dealt with in a satisfactory way. The first main problem area concerned normal human organisms at the beginning stages of life, and the second concerned abnormal human organisms—so-called “marginal cases”—such as human organisms that are disabled, diseased, or genetically deficient in some way.


