

# The Argument from Species Overlap

by Jesse Ehnert

This thesis is submitted to the faculty of the  
Virginia Polytechnic Institute and State University  
in partial fulfillment of the requirements for the degree of

Master of Arts

in

Philosophy

Harlan B. Miller, Chair  
William FitzPatrick  
James C. Klagge

July 15, 2002  
Blacksburg, Virginia

Keywords: ethics, animals, nonhuman animals

Copyright 2002, Jesse Ehnert

# The Argument from Species Overlap

by Jesse Ehnert

## ABSTRACT

The ‘argument from species overlap’ (abbreviated ASO) claims that some human and nonhuman animals possess similar sets of morally relevant characteristics, and are therefore similarly morally significant. The argument stands as a general challenge to moral theories, because many theories hold that all humans possess greater moral significance than all nonhuman animals. In this thesis I discuss responses to the ASO, primarily those of Peter Carruthers, Tom Regan, Evelyn Pluhar, and Peter Singer. Carruthers denies the conclusion of the ASO, while the other three do not. I argue that the ASO is a sound argument, and that Carruthers’s attempts to counter it via his contractualist theory are unsuccessful. I next discuss the rights-based theories of Regan and Pluhar, which agree with the conclusion of the ASO but which, I believe, encounter significant theoretical difficulties. Finally, I address the ASO from a utilitarian perspective, first from Singer’s utilitarian formulation and then from a ‘welfare-utilitarian’ formulation. I answer a number of critical objections to welfare utilitarianism, and argue that the theory is most successful in facing the challenge of the ASO.

To my family.

## ACKNOWLEDGEMENTS

I am indebted to many people for helping make this work possible. I thank my thesis committee members for their support, advice, and patience. My adviser, Dr. Harlan B. Miller, deserves substantial credit for leading me to this project and for providing invaluable assistance as I made my first steps into the subject matter. Dr. William FitzPatrick and Dr. James C. Klagge each provided important insights into my work, revealing to me some of the more serious challenges to my arguments.

Many thanks go to the entire Virginia Tech philosophy department, the students, faculty, and administration. It has been an extraordinary pleasure to work with such a thoughtful, dedicated group of individuals.

I must also express my sincere appreciation to many of my friends and family members who were surprisingly willing to listen to and respond to my seemingly endless philosophical ramblings. Among these are my former classmates Bryan Baltzly, Seth Fairbanks, Zane Rogers, and Jason Rosencrantz, my brothers Brian and Terence, my mother Sherry and stepfather Al, and friends Jesse Fuchs, Elizabeth Owens, and Jeff Stern.

Finally, I would like to acknowledge Dr. Patrick Croskery, formerly of Virginia Tech, in whose class I discovered my love of moral philosophy.

## TABLE OF CONTENTS

<b>ACKNOWLEDGEMENTS .....</b>	<b>iv</b>
<b>CHAPTER 1: INTRODUCTION.....</b>	<b>1</b>
PRELIMINARIES .....	4
<b>CHAPTER 2: THE CONTRACTUALIST RESPONSE .....</b>	<b>7</b>
VARIOUS OBJECTIONS TO THE ASO.....	7
CARRUTHERS .....	11
CONTRACTUALISM AND ANIMALS.....	12
CONTRACTUALISM AND HUMANS.....	17
SUMMARY .....	23
<b>CHAPTER 3: THE RIGHTS RESPONSE .....</b>	<b>24</b>
REGAN.....	24
REGAN'S THEORETICAL METHOD .....	27
REGAN AND THE ASO.....	33
PLUHAR .....	35
PLUHAR'S THEORETICAL METHOD.....	35
PLUHAR AND THE ASO.....	37
<b>CHAPTER 4: THE UTILITARIAN RESPONSE.....</b>	<b>44</b>
UTILITARIANISM'S THEORETICAL FOUNDATION.....	44
UTILITARIANISM AND THE ASO.....	49
VARIETIES OF UTILITARIANISM .....	50
UTILITY.....	50
MAXIMIZATION.....	55
CRITICISMS OF UTILITARIANISM .....	56
STRICT DECISION MODEL.....	57
FAILURE TO PROVIDE ADEQUATE PROTECTION, PART ONE.....	59
FAILURE TO PROVIDE ADEQUATE PROTECTION, PART TWO.....	63
REPLACEABILITY.....	69
WELFARE UTILITARIANISM AND A FINAL RESPONSE TO THE ASO .....	77
CONCLUSION .....	79
<b>BIBLIOGRAPHY.....</b>	<b>81</b>
<b>VITA .....</b>	<b>82</b>

“I thought, ‘Oh my God, it's like eating my niece.’”

-- *Cameron Diaz on why she stopped snacking on bacon after she was told that pigs have the same mental capacity as a 3-year-old, in Esquire.*<sup>1</sup>

## CHAPTER 1: INTRODUCTION

A typical method for criticizing any moral theory is to test its adequacy in handling difficult, unusual, or extreme cases. After all, moral theories tend to agree about the basics, what might be termed ‘commonsense morality’. Typical elements of commonsense morality include presumptions against lying, harming, and killing, as well as presumptions in favor of their opposites: honesty, benevolence, and the prevention of death. A theory that contradicts too many of our common beliefs is a highly questionable theory, just as any scientific theory would be likewise suspect if it failed to account for what we saw in the world around us. Despite the fact that different theories of morality agree on many issues, they can and often do differ greatly about what they consider the ultimate grounding of moral value. These differences produce widely varying results when the theories are applied to cases for which they were not designed, cases that are often difficult, unusual, or extreme. If one theory more adequately handles those special cases, while simultaneously explaining commonsense morality, then we have reason to prefer that theory to the others: the others have been shown to be superficially successful, but not ultimately successful in all cases.

In the ongoing debate regarding the moral status of nonhuman animals (hereafter usually shortened to ‘animals’), philosophers have found themselves in exactly the sort of situation I described, one where differences in moral theories produce radically different positions on the issue. The rival theories all provide some justification for the (very high) moral status of most or all human beings, but arrive at a variety of conclusions when applied to the subject of animals. Proponents on one side of the debate attempt to prove that at least some animals deserve consideration equal to that all human beings. Others argue just the opposite, that animals have no moral worth whatsoever. Of course, there are plenty of positions that propose some middle ground, holding that animals have some moral value, but not to the extent that humans beings do. As long as each theory is

---

<sup>1</sup> Bill Zehme, “Cameron Diaz Loves You.” *Esquire* April 2002: 78. Quoted from Salon.com on the World Wide Web (<http://www.salon.com/people/col/reit/2002/03/12/nptues>).

generally acceptable, the debate ends here, at a stalemate. What is needed to move forward is a focusing issue for the debate, a narrower point within the debate that forces those on each side to scrutinize their own theories. Enter the ‘argument from species overlap’.

The argument from species overlap—or ASO, as I will abbreviate it—has played a major role in the debate because it provides a specific challenge for all moral theories.<sup>2</sup> As I will show, the ASO is a very effective tool for evaluating those generally acceptable rival theories, and revealing some to be, in fact, unacceptable. The following is the ASO in a very general form. This rendering of the ASO is based on the form used by Evelyn Pluhar in her book *Beyond Prejudice*.<sup>3</sup>

1. Individuals who possess similar sets of morally relevant characteristics are similarly inherently morally valuable.
2. Some nonhuman animals possess sets of morally relevant characteristics similar to those of some humans.
3. Therefore, some nonhuman animals and some humans are similarly inherently morally valuable.

The first premise of the ASO is nearly tautological: if two individuals share all characteristics upon which their moral value is founded, then they will have equivalent moral value. The contrapositive is even more obvious: if individuals do not share equivalent moral value, then the individuals cannot be identical in terms of morally relevant properties.

The second premise is more controversial. The claim is that whatever those properties are that determine the moral worth of an individual, there are some humans who possess those properties only to the extent that some nonhumans do. For example, many theories hold that the possession of a sufficiently rich mental life is the determining property. If so, then consider the fact that there are anencephalic humans born with almost no brain, but only a brain stem.<sup>4</sup> This human is utterly without consciousness, and therefore has a far simpler mental life than a great number of animals. There are many

---

<sup>2</sup> Although the argument is typically labeled the ‘argument from marginal cases’, I prefer this alternate wording, originally proposed by Harlan B. Miller (“A Terminological Proposal” in *SSEA Newsletter* 30 (March 2002)), as it is both more descriptive and less likely to be misinterpreted and cause offense.

<sup>3</sup> Evelyn Pluhar, *Beyond Prejudice: The Moral Significance of Human and Nonhuman Animals* (Durham and London: Duke University Press, 1995), 65. In my rendering of the argument I have chosen to omit Pluhar’s expression “maximally morally significant,” in order to reach a more general conclusion. The argument loses none of its force in the process.

<sup>4</sup> *Ibid.*, 8.

other humans who are far more fortunate, but who are nonetheless mentally undeveloped (such as a newborn infant) or defective. We can imagine a continuum on which human beings exist, ranging from those with no mental life whatsoever, on up to the epitome of a full human mental life. At many points along that continuum, we find not only humans, but nonhumans as well. Therefore, there are some humans who, in terms of possession of a mental life, are similar to some nonhumans.<sup>5</sup> If possession of a mental life constitutes a complete set of morally relevant characteristics, then the second premise is true. If indeed the second premise is true, and given the relatively unproblematic first premise, then the conclusion follows. Those humans and nonhumans who possess sets of morally relevant characteristics are similarly morally valuable.

If the ASO is a sound argument, then the proposition that all humans possess greater moral value than all animals is false. If the proposition is false, then any theory that produces this proposition is flawed. All other things being equal, we ought to reject those flawed theories and favor theories that successfully explain the conclusion of the ASO. On the other hand, if the ASO is shown unsound, then we should prefer those theories that explain the flaw in the ASO. It is in this manner that the argument serves the purpose I spelled out in my opening paragraph.

I will discuss four principal moral theories that have been applied to the ASO in this manner. They are: Peter Carruthers's contract theory, the rights-based theories of Tom Regan and Evelyn Pluhar, and the utilitarian theory of Peter Singer. I critique each candidate theory in turn. What I hope to show is that, while Singer's brand of utilitarianism runs into significant challenges, a certain sort of utilitarian theory, 'welfare utilitarianism', is ultimately the most successful at handling the demands of the ASO.

In this paper I will specifically focus on how each theory handles the morality of the *death and killing* of animals and humans. I choose this particular focus for a number of reasons. First, applying a theory to the subject of death is a clear and easy way of illustrating the relative moral worth of different individuals. Second, the morality of killing provides some of the most difficult cases for a utilitarian to handle. As I will be

---

<sup>5</sup> This notion of a continuum is an oversimplification, and does not capture the complex dimensions of mental abilities, but I think it nonetheless indicates something true about relative capacities between individuals.

defending a utilitarian theory, it serves my purpose to address the more difficult challenges to my own theory.

## **Preliminaries**

There are a number of facts that play a role in each of the four positions discussed in this paper. These facts are important because they underlie the discussion, but they are not at the focus of the debate. For that reason, I will enumerate them here and henceforth consider them presupposed in the debate.

First, there is the fact that many animals are sentient. It is a commonsense notion that ‘higher’ animals such as dogs and cats are conscious and can experience pleasure and pain. It is also likely that many birds, fish, reptiles, and amphibians also possess consciousness, though animals with simpler nervous systems may not be conscious to the same extent.

Peter Carruthers proposes an argument against animal consciousness in his *The Animals Issue*.<sup>6</sup> Conscious mental experiences, according to this argument, are those that can be the object of second-order beliefs. Animals that do have thoughts—mental experiences, desires, and beliefs—but do not possess second-order beliefs about those thoughts, are not conscious, and therefore their pleasure and suffering cannot matter morally. Carruthers argues that no animals, not even (nonhuman) primates, possess second-order beliefs. Therefore, the pleasure and suffering of any and all animals cannot matter morally.

Enough has been said in response to Carruthers’s position and similar arguments that I have nothing to add to that debate. One example is David DeGrazia’s *Taking Animals Seriously*, which explores animal minds in depth, and presents strong evidence in favor of animal consciousness not only in primates but most or all vertebrates. DeGrazia also responds directly to Carruthers’s argument.<sup>7</sup> Pluhar has a detailed objection to Carruthers as well.<sup>8</sup> Carruthers is aware that his argument is controversial, and is himself willing to put the debate aside and assume animals are conscious until

---

<sup>6</sup> Peter Carruthers, *The Animals Issue* (Cambridge: Cambridge University Press, 1992), chapter 8.

<sup>7</sup> David DeGrazia, *Taking Animals Seriously* (Cambridge, UK: Cambridge University Press, 1996), 112-115.

<sup>8</sup> Pluhar, 37-46.

proven otherwise: “Until something like a consensus emerges, amongst philosophers and psychologists concerning the nature of consciousness, and amongst ethologists over the cognitive power of animals, it may be wiser to continue to respond to animals as if their mental states were conscious ones.”<sup>9</sup>

A second presupposition is the fact that some humans are not ‘full persons’, rational agents possessing a rich mental life. Some humans are not agents at all: they have no preferences or goals whatsoever. Fetuses, at least up to some point of development, fall into this group, as do victims of comas or severe brain damage. The ASO, of course, depends upon the existence of these atypical humans to draw its conclusion about animals who similarly fail to qualify as full persons. The existence of atypical humans is not in itself a moral claim, but moral theories often propose the moral relevance of some set of mental properties. The fact that not all humans possess these properties is what gives the ASO its moral force.

A third presupposition is the fact that the set of moral agents and the set of moral patients are not necessarily identical. I am following Harlan B. Miller in my use of the terms ‘moral agent’ and ‘moral patient’: “To be a moral agent is to be an entity capable of actions that may appropriately be evaluated as right or wrong. To be a moral patient is to be an entity of such a sort that what is done to that entity by a moral agent is per se, subject to moral evaluation.”<sup>10</sup> It may be the case that all moral agents are moral patients, and vice-versa, but this case must be argued for; it is not the case by definition. (Miller, incidentally, holds that the two sets differ significantly: on his view, the set of moral patients includes human infants and children, who are not moral agents, while the set of moral agents includes abstract entities such as corporations, which are not moral patients.)<sup>11</sup> This distinction is important, because my use of ‘moral value’, ‘morally significant’, ‘the moral community’ and similar expressions regarding individuals throughout this essay indicate membership in the set of moral patients, and not necessarily the set of moral agents. My point here is not to depart from the sense of these

---

<sup>9</sup> Carruthers, 192-193.

<sup>10</sup> Harlan B. Miller, “Introduction: ‘Platonists’ and ‘Aristotelians’,” in *Ethics and Animals*, ed. Harlan B. Miller and William H. Williams (Clifton, NJ: Humana Press, 1983), 12-13.

<sup>11</sup> *Ibid.*, 13.

expressions as they appear in Pluhar's discussion of the ASO; rather, I am clarifying the sense she must intend.

A fourth presupposition is that the commonsense idea that *at least some humans are more morally valuable than some animals* reflects something true about morality. Each of the four philosophers' theories arrives at this claim by one means or another. The real disagreement between them concerns which humans are more valuable than which animals, and to what extent. Their disagreement on these points can be understood in terms of their answer to the ASO, and is the essential issue of this thesis.

Finally, a brief note about how moral significance and consideration is understood. There are two categories I will discuss, *absolutism* and *gradualism*. An absolutist morality is one that divides the world into two groups, those individuals that matter morally, and those that do not. To matter morally is to matter maximally; there is no hierarchy of moral value, and no individual matters more than any other. A gradualist morality is one that begins with the absolutist division, but further divides the group of individuals that matter so that some individuals are 'higher up' the moral ladder than others. There are numerous varieties of gradualism, as there are many ways one can spell out the hierarchical scheme. For instance, one might consider the relative values of individuals' lives, insofar as their deaths are morally important, to lie along a gradualist spectrum, but view pleasure and suffering as counting equally across all individuals. Regan, for one, proposes just this view. Sophisticated theories can and often do contain both absolutist and gradualist elements. For instance, a gradualist theory can propose a 'top' of the hierarchy, or an imaginary line above which individuals possess maximal moral value, providing them with a special form of consideration not granted to individuals 'below' the line. As we shall find, a moral theory's absolutist or gradualist elements play an important role in the theory's handling of the ASO.

## CHAPTER 2: THE CONTRACTUALIST RESPONSE

### Various Objections to the ASO

Of the major positions on the ASO that I am considering, only the contractualist view argued by Carruthers flatly rejects its conclusion. However, there are many other philosophers who argue against the ASO. The reason I present Carruthers as the lone opponent of the ASO is primarily because the arguments of other opponents have already received a thorough response by Evelyn Pluhar,<sup>12</sup> and there is little that I can add. In addition, while Pluhar does defend the ASO against Carruthers's objections, I believe that his position remains important to consider. As part of my criticism of Carruthers's case, I will be making additional arguments that Pluhar didn't provide. Also, Carruthers makes interesting points regarding both Regan and Singer, several of which I incorporate into my discussion of their theories.

As a prelude to a discussion of Carruthers's position, it will be helpful to review the methods philosophers have used to oppose the ASO. Pluhar provides a comprehensive list of these objections in her book *Beyond Prejudice*. They can be classified into two categories. Objections in the first category attempt to reject the conclusion of the ASO—that some animals and some humans are similarly morally valuable—without clearly and directly opposing either premise. Typically, these objections can be dismissed by revealing an underlying misunderstanding of the ASO on the part of the objector.

For example, one objection to the ASO is that the conclusion could be seen as a justification for lowering the value we place on humans who are not full persons. Pluhar responds that, of course, that *is* one possible outcome. It is exactly that possibility that serves as a challenge to those who would deny consideration to animals.<sup>13</sup>

Others misunderstand the second premise of the argument, and claim that it does not accurately compare the abilities of atypical humans to the abilities of animals. Depending on which philosopher is making the objection, the comparison is found to be

---

<sup>12</sup> Pluhar, 67-107, 140-178.

<sup>13</sup> *Ibid.*, 72.

insulting either to atypical humans or to animals.<sup>14</sup> Still others misunderstand what is meant by the ‘moral value’ of an individual, and mistake it to mean moral agency.<sup>15</sup> Pluhar exposes each misunderstanding in turn, and in doing so does away with each of these objections.

Objections in the second category deny the second premise of the ASO: that some nonhuman animals possess sets of morally relevant characteristics similar to those of some humans. You will recall that these morally relevant characteristics are typically agreed to be one or more mental features, such as consciousness, a rich mental life, or perhaps rational agency. Now, for every such mental feature, there is some animal who possesses it to at least as great an extent as does some human. Therefore, in order to deny premise 2, the objector must add another item to the list of morally relevant features. This item ends up being species itself.

Defenders of the ASO often label their opponents ‘speciesists’, a term meant to carry the same kind of moral disapproval as ‘racists’ or ‘sexists’. Racists and sexists discriminate against certain individuals based on characteristics that are not morally relevant. Similarly, a speciesist is supposedly discriminating against individuals based on another morally irrelevant feature, one’s species. However, if species can be shown to be morally relevant, then discrimination based on species would be justified, and the speciesist label would lose its bite.

Most people would agree that, species aside, if an individual was capable of rational thought or had some sort of mental life at the level of normal humans, then that individual would be as morally valuable as humans. For this reason, a commonly accepted formulation of the speciesist position is that, to receive moral consideration equal to a normal human, one must either a) possess a certain set of mental properties, or b) be a member of a species characterized by those mental properties. This formulation clearly refutes the second premise of the ASO by drawing the conclusion that all humans possess a morally relevant characteristic not shared by animals, that characteristic being either of a) or b), above.

---

<sup>14</sup> Ibid., 77-85.

<sup>15</sup> Ibid., 74-77.

The burden of proof in this matter is on the opponent of the ASO; it must be shown that species membership is indeed morally relevant. However, as Pluhar reveals, no one has yet produced a satisfactory case. There have been extremely unsatisfactory attempts which I will not bother to go into here.<sup>16</sup> Of the better arguments, the common tactic is to show some important link between the atypical human and the mental properties that are characteristic of the human species.

One such attempt is to refer to the individual's *potential*.<sup>17</sup> Unlike nonhuman animals, humans who are not full persons at least had the potential to possess a rich mental life. According to this line of thought, because animals never even had the potential for such a mental life, they are not candidates for consideration the same way atypical humans are. Pluhar indicates the main problems with this move. First, having the potential for possessing something is logically distinct from actually possessing it. A potential employee receives no paychecks; a potential threat is not necessarily a threat. It is not enough to point to potential; what needs to be shown is that there is something morally relevant about the possession of the human's potential full personhood. Moreover, many humans have lost this potential. Therefore, it must also be shown that *past* possession of potential is morally relevant as well. Finally, some humans cannot even be said to have had potential. In extreme cases such as that of an anencephalic human, the individual at no point in time was a potential full person, and therefore gains no benefit from arguments from potential.

Arguments from potential are problematic, and in some cases utterly useless, when used to defend the moral significance of atypical humans. Another tactic, one that avoids mention of potential, instead refers strictly to a relationship between the individual and the normal functioning of members of the individual's species. I will refer to this as the *telos* argument. Pluhar finds philosophers with vastly different views on the moral status of animals—these include Joel Feinberg, Bernard Rollin, Alan Holland, and A. I. Melden—who make some version of *telos* argument.<sup>18</sup> One argument by Rollin, for

---

<sup>16</sup> Ibid., 140-146, 162-171. Arguments found in the second group of pages cited (appeals to kinship and opportunities for interaction) are actually considered the most plausible by Pluhar, but I disagree. To say the least, they are utterly unsatisfactory because they fail to confer a higher value on *all* humans than on any nonhumans, which is the goal of the speciesist.

<sup>17</sup> Ibid., 146-150.

<sup>18</sup> Ibid., 150-162.

example, says that we ought to respect the *telos* of an individual's species regardless of what species it is.<sup>19</sup> Rollin does not make a distinction between humans and other animals, but others do make the distinction, based on the fact that the human species, unlike the other species we know of, has full personhood as part of its *telos*. Because they view full personhood as the initial source of moral value, and because the human species uniquely (as far as we know) has full personhood as part of its *telos*, all humans deserve consideration above all animals.

There are as many different varieties of the *telos* argument as there are philosophers who utilize it, and Pluhar addresses each one on its own terms. While some of the minor objections she raises against specific arguments do not seem entirely justified,<sup>20</sup> her arguments successfully reveal the major flaws of all such attempts to distinguish humans from other animals. First, the *telos* argument claims that there is some sort of morally relevant relationship between individuals and the proper functioning of its species. But, as Rollins pointed out, this relationship, if it exists, exists both for humans and for nonhumans. When opponents of the ASO argue that the relationship is different for humans on account of their species, they are begging the question. They have inserted their conclusion—that species-membership is morally relevant—as a premise.<sup>21</sup>

Additionally, there is the question of what constitutes the 'norm' of a species.<sup>22</sup> If one determines the *telos* of a normal member of a species by taking a statistical sample of the members of a species, and finding the average condition, one sees some very bizarre results. For instance, Pluhar suggests a possible future world where few humans are full persons.<sup>23</sup> If the species were changed in this way, then non-full persons born at that point in time would not benefit from the *telos* argument. But this means that two individuals who are identical in all ways—even species membership—possess very different moral value, and merely because of the time at which they were born.

---

<sup>19</sup> Ibid., 154. The argument is summarized by Pluhar.

<sup>20</sup> See for example *ibid.*, 153: Pluhar argues that an individual with no moral value cannot suffer nonmoral evil. It is not entirely clear what she means by 'nonmoral evil', other than 'harm'. If this is what she means, I do not see why other individuals, such as plants or manmade machines, could not be the objects of nonmoral evil.

<sup>21</sup> *Ibid.*, 154-159.

<sup>22</sup> *Ibid.*, 159-161.

<sup>23</sup> *Ibid.*, 160.

Alternatively, one could determine the *telos* of a species to be represented by the most successful conditions (in terms of full personhood) found within the species, so that if at least one member of a species achieves full personhood, then full personhood is the norm. But this method runs into a similar problem: as soon as one member of a species is found to be a full person, all members of that species gain moral significance. If we try to avoid these problems by rejecting the notion that the norm of a species is based on empirical, statistical facts, then it is unclear how one would go about determining what the norm is. I see no recourse aside from some kind of metaphysical proof, and no such proof has, to my knowledge, been made.

We now turn to Carruthers who launches an impressive attack on the ASO, but whose arguments, like the others Pluhar addresses, ultimately fail.

### **Carruthers**

Few philosophers have opposed the ASO as directly and thoroughly as Peter Carruthers. In his book, *The Animals Issue*, he argues that indeed we ought to make a moral distinction between all humans and all animals, and that this distinction is non-arbitrary. However, unlike the attempts I have described above, Carruthers does not exactly argue that species membership is a morally relevant feature. He agrees that the second premise of the ASO is essentially true: both animals and some humans lack rational agency, which Carruthers argues is the fundamental source of moral value. But his moral principles are constructed in such a way that moral value can appear even where rational agency does not.

Empirical facts—facts about psychology in particular—play an important role in Carruthers's construction of principles regarding our treatment of non-full persons. While these facts lead him to grant a certain kind of moral value to animals, they lead him to find a much higher degree of moral value in atypical humans. In this way, Carruthers denies the conclusion of the ASO.

Carruthers begins by comparing his Rawlsian contractualist theory to Peter Singer's utilitarianism and Tom Regan's rights view. Arguing that the utilitarian and rights theories are too problematic to be acceptable, he goes on to show why contractualism does not provide any source of rights for animals. Animals, in his view,

possess no real moral value, but only extrinsic value. We may seem to have duties toward animals, but these duties are derived either from respect for pet owners and animal lovers, or from a duty to our own virtuous characters. On the other hand, all humans have equal and maximal inherent moral value. Carruthers argues that this moral distinction between species, while not a fundamental principle of contractualism, does indeed follow from the theory, and is non-arbitrary. He concludes his book with a social criticism regarding the contemporary animal rights movement. He says that the recent increase of concern for animals, both inside and outside of philosophy, is based on a “thoroughly misguided” morality.<sup>24</sup> Not only is the movement misguided, it is morally reprehensible.<sup>25</sup>

As I will argue, Carruthers does not successfully refute the ASO. His arguments granting even extrinsic value to animals fall short, and his attempts to bring all humans up to equal, maximal moral value are even more problematic.

## CONTRACTUALISM AND ANIMALS

Carruthers proposes a contractualist theory much like that of John Rawls, but extends the application of the theory beyond Rawls’s primary goal of determining societal structures and institutions.<sup>26</sup> According to Carruthers, the theory can be applied to the whole of morality. Carruthers’s contractualism is based on the notion of an imaginary contract between all rational agents.<sup>27</sup> The contract is hypothetical, a construction founded on an imaginary ideal: the theory generates a set of principles that we would agree to if we were *perfectly* rational. These moral principles are chosen in light of ‘broadly self-interested desires’, by which Carruthers means those desires that we would have regardless of the particular natures of our individual lives. Because these desires would primarily regard freedom, power and the like, a fundamental moral principle is respect for autonomy.<sup>28</sup> After all, given that the imaginary contractors do not know the particularities of their lives while agreeing to the moral principles, they would value their rational agency above all else. This fundamental principle would handle

---

<sup>24</sup> Carruthers, 196.

<sup>25</sup> *Ibid.*, 168-169.

<sup>26</sup> *Ibid.*, 37.

<sup>27</sup> *Ibid.*, 35-38.

<sup>28</sup> *Ibid.*, 40.

much if not all of our negative duties to one another. Many other principles might be agreed upon as well: for example, the perfectly rational agents would agree to principles of beneficence, in order that they could rely on aid when in need.

The contractualist excludes animals from consideration. The reason is clear, says Carruthers: animals are not rational agents.<sup>29</sup> Since only rational agents choose the moral principles, and since they do so out of broad *self*-interest, only rational agents are granted moral consideration. And since human beings are the only rational agents in the world so far as we know, only human beings receive moral consideration. An additional claim is that consideration is given to *all* human beings; this claim I will address later.

At this point, I will explain what kind of duties to animals come out of the contractualist theory. Since there can be no direct duties to animals—for this would entail that they deserve consideration—they can only be the object of indirect duties. Carruthers claims that there are three possible indirect duties. The first relates to property rights.<sup>30</sup> Since some animals have human (rational agent) owners, others would have the indirect duty not to damage the animal, just as they have a duty not to damage a person's car, or pet rock. This duty is quite weak, of course. This provides no protection to 'unowned' animals, and does not even protect the owned animals from their owner! (After all, I have the right to slash my own tires. Slashing my cat is no different insofar as this duty pertains.) Carruthers calls the second indirect duty a duty of 'legitimate public concern'.<sup>31</sup> Since there are rational agents in the world who sympathize with animals, there should be general rules protecting their (that is, the rational agents') interests. These rules would not be terribly strong; they would only be as strong as rules protecting similar human interests, such as those concerning public decency or the preservation of historical buildings. And, like those other rules, they would have little control over what is done in private, to one's own possessions. I can do what I like in the privacy of my own home, decent or not; the 'public concern' duty only demands that I not negatively affect those people who are more sensitive. So, in terms of the animals, they are protected in public, but not in private, by both the first and second types of indirect duties.

---

<sup>29</sup> Ibid., 98-99.

<sup>30</sup> Ibid., 105-106.

<sup>31</sup> Ibid., 106-107.

These two duties fail to do credit to our everyday moral sensibilities, a fact Carruthers admits.<sup>32</sup> After all, despite whatever pain we might feel for the pet owner when the pet is injured, it is not that the wrongness of injuring animals is based on whether or not the animals are owned. The difference between our typical moral evaluation of the kicker of a domesticated dog and that of the kicker of a wild dog is slight, if a difference is there at all. Moreover, there is little if any moral difference between a public display of such violence and a similar private indulgence. Therefore, Carruthers must rely on his third indirect duty, the duty ‘to develop and maintain a virtuous character,’ to do the bulk of the work.<sup>33</sup> In addition to the principles so far mentioned, principles directed at a virtuous character can be derived from contractualism. While the imaginary contractors are perfectly rational, human beings are not. Therefore, while a perfectly rational agent could accurately apply moral rules to every particular situation that she found herself in, it is a practical necessity for imperfectly rational agents that they cultivate virtuous dispositions. With the aid of a well-developed, virtuous character, an agent’s actions will likely approximate the principles of the perfectly rational agent. Because right actions will be more likely to occur, the imaginary contractors would agree to a duty to one’s character. All well and good; so, how does this apply to animals? Carruthers argues that to intentionally injure an animal, regardless of whether the animal is domesticated, and regardless of whether the act is done publicly or privately, is a demonstration of one’s cruelty and indifference. Such vices violate this third type of duty, and so are against the contractualist account of morality.

This is a very strong argument for the contractualist to use, because it not only escapes the problem of being too narrow, evident in the first two duty-types, but it also introduces some notion of giving animals *direct consideration*. It would seem that to do so would violate contractualism’s first principle about who counts. But it doesn’t. Strictly speaking, direct consideration is extended only to rational agents. However, the nature of virtue is such that the exercise of it brings about a sort of *secondary direct consideration*, which is extended both to rational agents and to others.<sup>34</sup> The virtue of beneficence, for example, is the disposition to feel sympathy for someone’s suffering and

---

<sup>32</sup> Ibid., 108-110.

<sup>33</sup> Ibid., 146-169.

<sup>34</sup> Ibid., 154.

to act to relieve that suffering. Since an animal is capable of suffering, even if not unjustly since that would require primary considerability, the virtue of beneficence would have us give direct consideration to the animal. It is not moral consideration in the primary sense, but it is what I will call virtuous, or secondary, consideration. Because contractualism demands virtuous characters for no reason other than the well-being of the rational agents, virtuous consideration is merely a secondary effect of real moral action.

Despite this argument's appeal, I do not think Carruthers achieves a significant amount of success from this third attempt. My first objection involves the source of these virtues. Again, the virtues related to duties of beneficence, non-maleficence, honesty, etc. ought to be cultivated in rational agents in order to approximate the requirements of duty when precise moral calculations are too difficult. Recall that the contractualist requirements of duty regard the direct consideration of rational agents and no others. My objection is this: could the contractualist not limit the scope of the virtuous dispositions to those who *truly* matter? It seems like a sufficiently beneficent character need only sympathize with one who is suffering not from pain, but from an obstacle to autonomy. The pain felt by a rational agent would be, of course, one example of such an obstacle, but pain *qua* pain is not the right kind of object of sympathy, given the enormous amount of things that feel pain and yet do not matter morally (under contractualism).

Consider this virtue, then: there should be, by the lights of contractualist theory, a virtue that approximates the principle of non-interference with the autonomy of others. Persons should be in the habit of promoting and facilitating the freedom of others. This disposition falls under the general heading of virtues of non-maleficence. Persons in the habit of controlling other persons—physically restraining them and so on—possess a vice contrary to the demands of contractualist morality. With that in mind, consider the scope of this virtue. Consider the variety of things in the world that rational agents physically control and restrain: the vehicles they operate, the tools they use, and, of course, the animals they keep. Insofar as these things have the capacity for motion (and independent motion in some cases), they are similar to the rational agents themselves. It does not follow, of course, that the presence of this shared property requires us to treat those things as we do rational agents for the sake of our virtuous character. If indeed we did have to exercise virtues beyond the scope of consideration, and virtues of non-

maleficence overflowed into our treatment of animals, one would think that the contractualist would forbid the very sorts of things that Carruthers considers permissible, such as the caging of animals. So, to return to the point about the virtue of beneficence: if we ought to be sympathetic to the pain of others, it ought to be enough that we are sympathetic to those others who actually matter.

There are two options for the contractualist. The first option, following the above reasoning, is to narrow the scope of virtuous activity to best approximate the real objects of duty. If the contractualist chooses this option, then the account of virtues would provide no consideration for animals. Animals would benefit only from the duties related to property rights and ‘public concern’, and as we saw, there is not much benefit there. The other option is to claim that such narrowing is impossible, that humans are psychologically unable to reliably distinguish between those who matter, and those who don’t. But this option is not plausible: certainly we can at least distinguish between humans and animals. On that point, Carruthers and I agree: in an argument regarding a separate issue—his defense of the inclusion of all humans into the moral community—he relies on the fact that it *is* easy for humans to psychologically separate their treatment of humans from their treatment of animals.<sup>35</sup> For example, the ability to make the distinction is what makes employment in factory farms and animal laboratories possible. In Carruthers’s words,

That someone can become desensitised to the suffering of an animal need not in any way mean that they have become similarly desensitised to the sufferings of human beings—the two things are, surely, psychologically separable.<sup>36</sup>

This being the case, Carruthers must admit that our sympathy for animals, if it is to be understood as a virtue, is a flawed virtue.

It should be clear by now that contractualism cannot even provide the secondary sort of direct consideration for animals that Carruthers wanted. The moral intuition regarding the wrongness of harming or killing an animal cannot be explained by appeal to the indirect duties of respect for property, ‘legitimate public concerns’, or even by appeal

---

<sup>35</sup> Ibid., 115.

<sup>36</sup> Ibid., 160.

to duties of character. If one wishes to subscribe to Carruthers's contractualism, one must bite the bullet and give up this intuition.

## CONTRACTUALISM AND HUMANS

Carruthers's theory fails to grant any significant moral consideration of animals, despite his best efforts. However, he believes that all humans possess equal, maximal moral value. If he is correct, then the conclusion of the ASO—that some nonhuman animals and some humans are similarly morally valuable—is false. However, he recognizes that bringing all humans into the moral community is a serious challenge.<sup>37</sup> As I have mentioned before, the first premise of the ASO is unproblematic, and the second premise, that some humans are in all morally relevant ways similar to animals, is not easy to deny. It is certainly not easy for Carruthers to deny, since his central morally relevant characteristic is rational agency. Infants and young children are not rational agents, and other humans have lost the capacity for rational agency altogether. Because contractualism requires moral agency for every member of the moral community, it appears that these groups of humans must be excluded, just as animals are.

Carruthers believes he can handle the cases of the very young more easily than other cases of non-full persons. Since the injuring of a child can harm the future rational agent, it appears that contractualism may provide some protection to them. We can assume that the imaginary contractors would agree to a principle of allowing the young to develop in healthy ways, as it serves the self-interest of the rational agent that child becomes. However, contractualism has a problem explaining what is wrong with injuring a child if the injury results in death. If the child dies, there is no future rational agent! Likewise, if the injury results in the child's inability to become rational, then no rational agent was harmed. It would seem that injuring a child is wrong if and only if there will exist a future rational agent who is thereby harmed. This result is, of course, unsatisfactory for Carruthers. Even if it made sense, somehow, to provide protection not only to actual but also to *possible* future agents—thereby explaining the immorality of killing children, etc.—this protection would likely be too drastic. Abortions and even contraception would be impermissible. Moreover, it seems that regular attempts to

---

<sup>37</sup> Ibid., 110.

produce offspring would become mandatory! Failures to make such an attempt would often prevent a future rational agent from ever existing. Clearly, the demands of this kind of moral principle are unacceptable by any reasonable standard. Therefore, Carruthers must look elsewhere to account for the moral value of non-full persons.

He relies on two arguments: a slippery slope argument and an argument from social stability. I will comment on each in turn. First, his slippery slope argument runs as follows. There is no magic line between rational agent and non-rational agent. While our society does in practice establish age-based restrictions, such as a twenty-one year drinking age, no one believes that a human becomes rational at any specific age. There is, as we all know, a large gray area. Strictly speaking, a contractualist would agree that the moral standing of a human improves between birth (where there is no rational agency) and full adulthood (assuming the adult has full rational agency). Despite the differences in the *real moral worth* of humans, based on their rational agency, there are practical issues that must be accounted for. Most people, he thinks, should not be trusted to handle a moral principle that distinguishes between rational and non-rational humans. They would tend to misuse the principle, and deny consideration not only to non-rational humans but also to rational humans who are considered ‘deviant’ for some (morally irrelevant) reason. The slippery slope lies between actual non-rational humans and humans who are fully rational but appear otherwise.

This argument is not without its problems. First, it is not clear how potential for misuse invalidates a moral principle. One might think that we have a moral duty to drive our cars below the speed limit, to protect our lives and the lives of others. However, there is frequent misuse of this principle. Ought we for that reason conclude that the principle is immoral? Of course not. But Carruthers would have a reply to the speed limit analogy. It is not that a principle treating different humans differently would be knowingly disobeyed, as in the case of the speed limit. Instead, people would unintentionally misuse it, due to a lack of intellectual ability. Carruthers believes that “most people are not very deeply theoretical,” and would not be able handle a principle that requires them to recognize rationality in others.<sup>38</sup> Since the rational contractors

---

<sup>38</sup> Ibid., 116.

would be aware of the lack of theoretical ability in these people, the revised principle of universal human considerability is necessary.

His reply might reasonably justify extending protection to non-rational people who are almost indistinguishable from rational agents. It would be disastrous to allow the killing of a senile or similarly incapacitated adult human, for instance, if indeed most people would be unable to determine whether such people were not rational. It is unclear, however, how this applies to the treatment of infants. Certainly no one has difficulty telling an infant from an adult. While there is certainly a gray area between infancy and adulthood in regard to the development of rational agency, no gray area exists between, say, a newborn and a six-month-old. Neither is a rational agent in the slightest. The slope does get slippery sometime after that, so, to avoid becoming slippery, we could adopt a principle protecting all humans except those below the age of six months. This principle should be perfectly acceptable to Carruthers.

He does not accept it, of course. He does not believe that people can handle the distinction between such obviously different people. The strange fact is that he *does* believe that people can make a distinction between humans and animals, and avoid a slippery slope between our treatment of the two groups. He writes, “someone who argues that since animals do not have rights, therefore babies have no rights, therefore there can be no moral objection to the extermination of Jews, Gypsies, gays, and other so-called ‘deviants’, is unlikely to be taken very seriously.”<sup>39</sup> His claim is that no slippery slope exists between animals and human infants. But remove the ‘animals’ bit from his sentence, and we still have a sentence that is unlikely to be taken seriously. Someone who argues that since babies (unlike all other humans) have no rights, therefore there can be no moral objection to the extermination of (adult) Jews, etc., is likewise making an obviously lousy argument. There appears to be nothing blocking my proposed re-revision of Carruthers’s principle, though it places infants in the same morally precarious situation as animals.

Finally, Carruthers has made a fundamental error in his argument that renders this entire line of thought unacceptable. According to contractualism, the only beings who matter morally are rational moral agents. It is simply contrary to the theory to grant

---

<sup>39</sup> Ibid., 115.

primary, direct consideration to any non-rational beings. I use ‘primary, direct consideration’ in the same manner as it was used in the earlier discussion of virtuous behavior; I contrast it with secondary, direct consideration, such as that granted to animals for the sake of virtue. Such consideration is required only for the sake of those who deserve primary consideration. Despite Carruthers’s attempt to present our reason for protecting atypical humans as different from our reason for protecting animals, the reasons are the same. Our treatment of any non-full person matters *only* insofar as it affects the treatment of rational agents. This is the moral fact of the matter, regardless of anything contingent facts may do to alter the final form of the moral principles. Do we really believe that humans who are not full persons deserve protection not for their own sake, but only because injury to rational agents might result? No, this does not even approach our moral intuitions. We protect the infant for the sake of the infant! To drive the point home, consider what Carruthers would have to say about a world where most people are not so terribly lacking in theoretical ability. If proper contractualist moral education could be provided, it would no longer be immoral to kill a child or conduct painful medical experiments on the mentally impaired. Perhaps Carruthers would bite the bullet here, and agree that such actions would in fact be permissible, given the different conditions, but it is a tough bullet to bite.

His second attempt to grant moral standing to all humans is, unfortunately, worse than the first. His ‘argument from social stability’ relies on the premise that many people would find it psychologically unbearable to live in a world where non-rational humans are not morally considerable.<sup>40</sup> Because people would be unable to accept a principle that denies full consideration to certain humans that they care about, such as their children, and because their inability to accept the principle would result in social instability, it is necessary to replace the principle with something more acceptable. A principle that grants full consideration to all humans is the best substitute; therefore, it is the moral principle to which the imaginary contractors would agree.

Carruthers entertains the objection that there are in fact socially stable communities where differential treatment of humans occurs, and that therefore his claim about human psychology is false. He responds by pointing out the differences between

---

<sup>40</sup> Ibid., 117.

those other cultures and our own. First, in some cultures, religion and tradition play a much stronger societal role than they do in ours, and allow discrimination between different humans while maintaining social stability. Without the powerful religious tradition, stability would be impossible. Second, some cultures are “teetering on the edge of survival,” and so allow practices such as infanticide only because it is necessary for survival.<sup>41</sup> To sum up Carruthers’s response: Our culture, unlike some other cultures, is such that the majority of us cannot psychologically handle any principle that denies consideration to all human beings. Therefore, we must grant consideration to all human beings.

One way to counter Carruthers’s response is to deny the empirical claims he makes about other cultures. Pluhar, for one, has done so.<sup>42</sup> I choose to take another route, and examine how the claim, if true, should be understood. There are two possible ways to interpret the claim he is making about *our* culture, and I will address each one.

On the one hand, he might be claiming that our culture produces people who, unlike the people of other cultures, are psychologically unable to deny moral consideration to any humans. This is not a moral claim. However, a moral claim seems to follow from it: a contractualist should prefer a culture where we can make all morally relevant distinctions, and, so, not prefer ours. Carruthers, if he were making this claim, would have to admit that we *really ought* to deny non-rational humans primary, direct consideration, and that we are currently in an unfortunate situation where true morality is impossible. His imaginary contractors ought to construct principles that work toward ending this situation.

Carruthers attempts to correct this. He argues that we cannot modify our psychological states, and so the imaginary contractors would not demand that we do so. Instead, moral principles would conform to our psychology. This, he says, is what the argument from social stability is really saying. But consider the following implication of his argument. People in a strongly religious or traditional culture, who are psychologically able to kill or injure some kinds of non-rational humans, and who can make the proper distinctions between individuals so that no rational humans are harmed,

---

<sup>41</sup> Ibid., 119.

<sup>42</sup> Pluhar, 94-95.

are morally permitted to harm non-rational humans. The culture need not be, as Carruthers puts it, “teetering on the edge of survival.” So long as social stability is maintained, the contractualist does not judge harming non-rational humans to be morally wrong whatsoever. On this view, the real moral significance of non-rational humans simply varies from culture to culture. This fact stands in stark contrast to the contractualist view that all rational humans are *necessarily* maximally morally valuable, regardless of facts about their culture.

Carruthers should want to object with this line of thought. One would think that fundamental moral principles should not be based on contingent particularities within a certain culture. This smacks too much of cultural relativism. Of course, Carruthers cannot make this objection, because his moral principle protecting non-rational humans is undeniably based on cultural norms. Moreover, Carruthers does not seem entirely willing to admit that non-rational humans deserve no consideration except in cultures where rational humans happen to have a certain favorable disposition towards them. I expect most people would be similarly unwilling to make that concession. For this reason, it would be sensible to reject this interpretation of his argument and examine the other.

The other interpretation is that Carruthers believes that our cultural attitudes toward children and other non-rational agents are in fact morally correct, despite the conclusions of contractualist theory. If this is what he means, we are led to a bizarre result: contractualist morality, based on the imaginary agreement between perfectly rational contractors, would include principles that are somehow morally correct *prior to* the agreement! There would have to be some non-contractualist moral standard for this interpretation to be the case, and it is clear that Carruthers does not admit such a thing. This interpretation fails more quickly than the first.

In conclusion, the argument from social stability is a mess. At best, it implies that moral principles ought to be compromised for the sake of contingent cultural biases. Yet it makes more sense to think that someone is psychologically unable to fully meet the demands of morality than to think that principles of morality can so easily turn on cultural norms. Moreover, in the case of our treatment of non-rational humans, it seems clear that our commitment to their well-being (when we have it) stems not from some nonmoral inclination, but because we know that they do matter morally. The killing of a

small child is morally repugnant not because of its effect on social stability, but because the child *really matters*. Like Carruthers's slippery slope argument, his argument from social stability does not adequately explain the moral considerability of all humans.

## SUMMARY

Carruthers has attempted a refutation to the ASO. He agrees that the second premise is essentially true, because animals and some humans lack rational agency, the fundamental source of moral value. To rescue the non-rational humans, he locates secondary sources of value. Empirical facts—facts about psychology in particular—play an important role in the construction of principles regarding our treatment of non-full persons. While these facts lead the contractualist to grant some reasons to treat animals well, they support even more protection for humans. In this way, Carruthers denies the conclusion of the ASO.

A strength of the contractualist defense is that it aims to produce real moral value in all humans. Carruthers is not arguing for the mere practical consequence that we treat non-rational humans *as if* they had maximal moral value; that is tantamount to admitting that they *in fact* do not. Instead, he brings the empirical, psychological facts to bear on the imaginary set of rational bargainers at the foundation of contractualism. In this way, the empirical facts affect the moral principles, and not just the practical consequences.

As I have argued, however, it nonetheless does not seem possible for the contractualist to produce principles adequately protecting either animals or non-rational humans. Even where contractualism does grant a modicum of moral considerability to non-full persons, it does so for the wrong reasons. Compounding that problem, any protection that is given is in the form of principles precariously resting on contingent, empirical propositions that may one day not hold true. When such time comes, the contractualist would be led to deny moral consideration to whole multitudes of creatures (humans and otherwise) who were previously morally considerable.

In the end, the problem contractualism has in answering the ASO is its principal requirement of rational agency for membership in the moral community. As we shall see, other moral theories do not share this obstacle, and produce more acceptable responses to the ASO.

### CHAPTER 3: THE RIGHTS RESPONSE

Contractualism is just one ‘rationality-based’ theory, taking rational agency as the basis for inherent moral value. Kant’s moral theory famously does this as well. There is at least one reason why rational agency is so important for moral theorists: note that, regardless of the theory, rational agency is a necessary condition for morality to function *at all*. In order for us to act morally, to follow any moral principle, we must be capable of conceptualizing principles and acting deliberately. That is, all moral agents are rational agents. But the set of moral agents need not be identical to the set of moral patients. As we found with contractualism, making the two sets identical denies many humans and animals full moral standing, or any moral standing at all.

In this chapter, I examine the responses to the ASO developed by Tom Regan and Evelyn Pluhar. Regan and Pluhar hold similar, rights-based moral theories. As I will show, Regan proposes a plausible direction for an answer to the ASO, but runs into several problems. Pluhar attempts to recover Regan’s theory, and produces a more acceptable answer to the ASO. She is somewhat successful. I will argue, however, that the shortcomings of both Regan’s and Pluhar’s theories warrant a second look at the theory considered by both to be the ‘runner-up’ theory: utilitarianism.

#### Regan

Tom Regan has played a large role in popularizing the challenge the ASO presents. In his book, *The Case for Animal Rights*, he applies the ASO to Kantian morality, and shows that Kant’s withholding of moral value to all but rational agents excludes not only animals, but also many humans from the moral community.<sup>43</sup> Non-rational humans may receive indirect consideration, but as we’ve seen already with Carruthers’s contractualism, indirect consideration is insufficient. He rejects Kant’s theory, and seeks out one that does not suffer from the same problems.

The theory he seeks will be something other than rationality-based. Reviewing other possible morally relevant characteristics, he considers and rejects the alternative of a being’s *simply being alive*. While the equal inclusion of all living things in the moral

---

<sup>43</sup> Tom Regan, *The Case For Animal Rights* (Berkeley and Los Angeles: University of California Press, 1983), 174-185.

community would avoid contradiction with the ASO, there would be many other results that are counterintuitive to say the least.<sup>44</sup> So Regan turns to his notion of a ‘subject-of-a-life’ as the source of inherent value.<sup>45</sup> Not as narrow as the rational agency requirement of Kant and Carruthers, Regan intends the term to capture most humans and other mammals aged one year or more, and probably others.<sup>46</sup> A sufficient condition for being a subject-of-a-life under Regan’s definition is possession of the following features:

1. Beliefs and desires;
2. Perception, memory, and a sense of the future, including one’s own future;
3. Emotions, and the ability to feel pleasure and pain;
4. Preference and welfare interests;
5. The ability to initiate action in pursuit of one’s goals;
6. Psychophysical identity over time; and
7. A welfare of one’s own.<sup>47</sup>

To summarize the criteria, one might say that a subject-of-a-life has an *experiential welfare*. Regan seems to require possession of all of the above features, though he is not entirely strict. For instance, he intends to count non-rational humans among his subjects-of-a-life, despite the fact that some cannot initiate actions. In any event, Regan holds that all subjects-of-a-life have equal inherent value. He does not make an argument for this claim, aside from pointing out that a theory based on his claim fares better than rationality-based theories when it comes to handling issues surrounding the ASO.<sup>48</sup> He also does not give an answer to whether or not individuals possessing some but not all of the required features have any moral value, or how that would be determined.<sup>49</sup> These omissions have negative implications for the theory, as I will show later.

The notion that being a subject-of-a-life is a morally relevant characteristic is not limited to any one specific theory. Many moral theories use something similar to Regan’s list of features as the basis for inherent value. I will refer to any such theory as a ‘sentience-based’ moral theory (bearing in mind that some theories in this category demand something more than bare sentience for moral significance) to be contrasted to ‘rationality-based’ theories such as those of Kant and Carruthers. Proponents of any

---

<sup>44</sup> Ibid., 241-243.

<sup>45</sup> Ibid., 243-248.

<sup>46</sup> Ibid., 73-81.

<sup>47</sup> Ibid., 243.

<sup>48</sup> Ibid., 247.

<sup>49</sup> Ibid., 264.

sentience-based theory are going to be in a better position to handle the ASO than are those who argue for a rationality-based theory, because they are not forced to contrive devices to handle non-rational beings.

However, different sentience-based theories handle the conclusion of the ASO in different ways. The conclusion, recall, only says that some animals and some humans are *similarly* inherently morally valuable. It does not specify to what extent any individual is morally valuable; that issue is left undetermined by the argument.<sup>50</sup> A full response to the ASO is not tantamount to mere agreement with its conclusion; it must also make the further determination of extent. For this reason, Regan examines various sentience-based theories, including several forms of utilitarianism.<sup>51</sup> In the end, for reasons I will discuss in more detail in the next chapter, Regan rejects all forms of utilitarianism. As an alternative, he proposes a theory that is Kantian in appearance, taking on the language of rights and with attention paid to respect for autonomous action and treating others as ends-in-themselves. However, it is a sentience-based theory, not a rationality-based theory.

Regan's defense of his theory is largely of a negative sort: he develops his principles out of a rejection of the most plausible available theoretical alternatives, most notably utilitarianism. He believes that the best utilitarian theory will ultimately fail to respect the inherent value of any individuals. For this reason, he postulates a 'respect principle': *treat those individuals who have inherent value in ways that respect their inherent value*. Spelled out, this principle consists of two main ingredients: a Kantian notion of an end-in-itself, and, following from that, a rejection of utilitarianism.<sup>52</sup> Specifically, Regan claims that we violate the respect principle when we make utilitarian calculations about morally significant beings.<sup>53</sup> So, just as his focus on the subject-of-a-life is the upshot of rejecting rationality-based theories, so his moral principles are built (in part) upon the rejection of utilitarianism. As we shall shortly see, this method of theory construction leaves something to be desired.

---

<sup>50</sup> Incidentally, Pluhar utilizes a second formulation of the ASO to conclude that non-rational humans and animals have inherent value equal to rational agents. In this paper, I do not focus on this formulation as the ASO, though it appears in my discussion of Pluhar's response to the ASO.

<sup>51</sup> Regan, 140-143, 200-235, 250-258.

<sup>52</sup> *Ibid.*, 248-250.

<sup>53</sup> *Ibid.*, 248-249.

## REGAN'S THEORETICAL METHOD

Having established the underlying principle of respect, Regan follows the following process. First, he derives secondary principles of action from the respect principle. Then, he examines the implications of those principles. From there he develops additional principles, also based on the respect principle, to handle any unsatisfactory implications.

The first principle he derives from the respect principle is a principle against harming, what he terms the "harm principle."<sup>54</sup> Out of respect for others, we must regard their welfare as important, because having an experiential welfare is what makes subjects-of-a-life morally valuable in the first place. Therefore, we have a prima facie duty not to harm others, as this would detract from their welfare. It is at this point that Regan first introduces the notion of rights into his theory; he defines rights as valid claims with corresponding duties.<sup>55</sup> The duty against harming, then, corresponds to a right not to be harmed.

From the harm principle Regan adds two corollaries. First there is his so-called "miniride principle," which states that, given a situation where we must either harm one group of beings or similarly harm another group, we ought to choose to harm the smaller group. The second principle is called the "worse-off principle," and states that, given a situation where we must either harm one group of beings or harm to a greater degree a smaller group, we ought to harm the larger group. This latter principle follows from the rejection of utilitarian calculations found in the respect principle. I will have more to say on this later on.

Next, he spells out some of the implications of these secondary principles. The principles are not absolute, he argues, because they may come into conflict with other principles, which themselves can similarly be derived from the respect principle.<sup>56</sup> Situations that exemplify such a conflict include situations such as acting in self-defense against an aggressor, punishing guilty moral agents, and allowing for non-obligatory, supererogatory actions. In these cases, argues Regan, we must balance the harm principle against other principles that come into play in each specific case.

---

<sup>54</sup> Ibid., 262-263.

<sup>55</sup> Ibid., 271-273.

<sup>56</sup> Ibid., 286-287.

If his rights theory succeeds where other theories, notably utilitarianism, do not, then all that is left to do is discover what answer his theory gives to the ASO. However, I do not believe it has succeeded. Primarily, it fails in regards to the theoretical method I have just described. It also has some bizarre implications regarding the ASO, which I will also discuss.

Problem #1: Weak foundation.

Regan claims that his proposed fundamental principle, the respect principle, “illuminates and unifies” many of our well thought out moral beliefs.<sup>57</sup> It is important to note here that the beliefs he mentions in association with this claim are actually nothing more than a rejection of a utilitarian morality. Are we to accept Regan’s theory on the basis that we reject someone else’s? It would appear so. It is as if the failure to make a positive case for utilitarianism has resulted in the establishment of the alternate theory. Aside from being the wrong way to go about constructing one’s moral theory, it underestimates the theoretical possibilities within moral philosophy. Surely more than one option is available to the opponent of utilitarianism.

Naturally, Carruthers—whose contractualist theory we examined in the previous chapter—is in just the sort of position that would lead him to attack Regan on this front. And Carruthers is very instructive on this point; he indicates a major flaw in Regan’s theory. The flaw is its lack of what Carruthers refers to as a ‘governing conception’. A governing conception is the explanation a moral theory provides as to what, indeed, morality *is*. It explains the following things: 1) the nature of morality, what moral ideas are about; 2) how we come to have moral knowledge; and 3) the basis of moral motivation.<sup>58</sup> After all, to develop a moral theory, one must presuppose that morality is a valid subject of rational inquiry, and something about which we can have knowledge (points 1 and 2). And it is similarly clear that moral principles, when we come to know them, have some sort of motivating force compelling us (with varying success) to act morally (point 3). If a theory cannot make sense of these most basic moral data, then the theory is groundless.

---

<sup>57</sup> Ibid., 259.

<sup>58</sup> Carruthers, 23.

Pluhar has a related argument against Regan, in her *Beyond Prejudice*.<sup>59</sup> Regan's theory stems from the respect principle, which is itself generated by his notion that all beings with an experiential welfare, all subjects-of-a-life, possess inherent value. This latter notion is controversial for two reasons. First, we need not agree with Regan's judgment that experiential welfare is a morally relevant characteristic. While we may very well run into difficulty with the ASO if we commit to a rationality-based theory, that does not automatically entail that we should move to a sentience-based one. That move is not by default; rather, it must be argued for. Second, we need not agree with Regan's specific notion of how we are to treat inherent value merely because it rejects utilitarianism. As I stated before, we can reject Regan's notion of inherent value even if we reject utilitarianism. Pluhar, who does indeed reject utilitarianism, follows her criticism of Regan's theory with an attempt to bolster the rights theory with additional arguments.

#### Problem #2: Vague decision model

A second weakness of Regan's theory is that it lacks a clear method for handling practical conflicts that result from the application of the respect principle and its derivatives. As I mentioned earlier, the duties Regan derives from the respect principle, e.g., the harm principle, are not absolute. There are situations when it is permissible to harm a moral patient, as long as we find that other principles, similarly derived from the respect principle, act more strongly upon us. Unfortunately, we are never told how to determine which principles act more strongly, and when. Instead, it appears that Regan falls back on moral intuitions as the basis for resolving conflicts between principles. While these intuitions play some role in the development of a theory, they cannot act as our sole mechanism for moral decision-making, lest we end up with a theory that mirrors whatever we happen to believe about individual moral cases.

For example, notice Regan's handling of cases of self-defense.<sup>60</sup> According to the harm principle, an innocent victim of an aggressor would not be permitted to harm the aggressor. He calls this notion, that it would be immoral to harm someone even in self-

---

<sup>59</sup> Pluhar, 231-240.

<sup>60</sup> Regan, 287-290.

defense, the ‘pacifist principle’, and argues that it runs sharply against our moral senses. In order to avoid the principle, he resorts to a consequentialist consideration: he claims that when defending against an aggressor and causing harm results in the prevention of a greater amount of future harms, the harm we cause is justified. He adds that, aside from the harm principle and the consequentialist consideration, there is also a ‘proportionality principle’ at work: we should inflict harm on the aggressor that is proportional to the harm being prevented.

Consider also how Regan allows for supererogation.<sup>61</sup> He presents an example where a racecar driver is seriously injured. Without medical assistance, he will die. There are a number of medical personnel in the area, but they all have patients who will suffer some serious harm (paralysis, or a lost limb) if not immediately treated. Regan’s theory—and specifically the worse-off principle—produces an obligation for the patients to forego treatment in order to save the driver. But that sort of choice is one that most of us would consider to be above and beyond the call of duty. Regan agrees, and shows that the situation actually involves a conflict of principles. The worse-off principle is the first one. The second principle involves situations such as that of the racecar driver, who voluntarily choose to enter into a high-risk activity. In these cases, Regan argues, the person knowingly waived the full protection of principles like the worse-off principle, and so we ought to treat the person accordingly.<sup>62</sup>

As a third and final example, consider how Regan explains the rightness of punishing the guilty.<sup>63</sup> It is a widely held notion, of course, that a person guilty of a crime ought to be punished. However, the punishment—whether a fine, imprisonment, death, or otherwise—counts as a harm, and so the harm principle tells us that punishment is wrong. Regan notes this, but offers no real answer. He mentions that we ought to respect the rights of the criminal, and that the proportionality principle would hold here as much as in the case of self-defense. But this does not tell us why we should punish in the first place. He finally offers the suggestion that his rights theory is “sympathetic to” the notion of punishment, but he fails to explain why.

---

<sup>61</sup> *Ibid.*, 320-322.

<sup>62</sup> *Ibid.*, 322.

<sup>63</sup> *Ibid.*, 290-291.

In each of these examples, Regan offers a supplementary principle to account for cases where the harm principle fails to provide an acceptable answer. And he tries, with varying success, to show how those supplementary principles are derived from the respect principle. Unfortunately, he fails to adequately explain how these principles interact. For instance, the consequentialist answer to the self-defense example shows that, at times, the consequences of one's actions take precedence over the strict duty not to harm. But when do consequences take precedence? In that one case, at least. But not always. At one point Regan declares that "side effects" of an action are irrelevant to the action's rightness or wrongness.<sup>64</sup> It is unclear which consequences are understood as side effects, but in any event, some consequences are trumped, morally, by the harm principle. But nowhere are we told how and when one principle overrides another.

A similar point can be made in regards to the supererogation principle. When someone enters into a high-risk activity, how are we to determine the extent to which we must continue to obey the harm principle? For instance, what would happen if we construed the racecar example so that, rather than a racecar driver, the near-death victim was a pit crew member working out on the racetrack? Would this still qualify as a high-risk activity? It is unclear how we are to weigh the risk accepted by the victim against the worse-off principle. Additionally, I should point out the strangeness of the example: the idea is that the other injury victims, not the medical personnel, are making the supererogatory choice to forego treatment in order to save the driver. But this setup is a little hard to believe; one would think the medical staff would normally be the ones making treatment decisions. But if it were up to the medical staff, an important factor in the supererogatory nature of the choice, i.e., sacrificing one's own good for that of another, is removed from the dilemma. What would Regan say about the medical staff's moral obligation? Would it be similarly supererogatory for them to assist the racecar driver, even though they are not making a sacrifice in doing so? Regan never provides us with the decision mechanism to address this.

The problem I am describing, the lack of any clear decision model in Regan's rights theory, is closely related to the initial problem I mentioned, the theory's weak foundation. In both cases, the cause of the problem is Regan's heavy reliance on moral

---

<sup>64</sup> Ibid., 312-315

intuitions. Certainly our intuitions matter—indeed, my own paper evaluates how well the various theories can handle commonsense morality—but a theory must rely on more than intuitions in its explanations, lest it become a mere enumeration of prevailing attitudes. The very basis of Regan’s theory, the respect principle, was created out of only two things: 1) a rejection of utilitarianism, largely on intuitive grounds, and 2) the assertion that all subjects-of-a-life have inherent value, a moral intuition. With nothing else to direct the theory, Regan simply throws out various moral problems and invents new principles as needed to account for that particular problem. And while some principles do seem to be derived from the respect principle, others (e.g., the proportionality principle) seem made out of whole cloth. In fact, most of the secondary principles contain elements that are introduced without explanation; the miniride and worse-off principles, for example, refer to harming *innocents*, specifically. Regan never explains where the notion of innocents comes from, or how it was derived from the respect principle.

Even worse, the vagueness of Regan’s intuitive foundation occasionally leads him to remain absolutely silent on an issue. Take for instance his answer to the morality of abortion.<sup>65</sup> When Regan speaks of subjects-of-a-life, he is careful to note that he is referring to mammals aged one year or more. Therefore, fetuses as well as the newly born are not assumed to be subjects-of-a-life (though they may be). But Regan makes it clear that being a subject-of-a-life is a sufficient, but not necessary, condition of possessing inherent value. Therefore, the inherent value of fetuses and the newly born is left unresolved. While that fact alone is not particularly problematic, its implications are. It turns out that anything that is not a subject-of-a-life is, possibly, inherently valuable. The lack of a necessary condition for inherent value makes for a mysterious moral theory. If, indeed, some non-subjects-of-a-life, whether newborn humans, fish, trees, lampshades, or anything else, turn out to be inherently valuable, how would we discover that fact? Would a new moral intuition become suddenly available to us? Regan gives us no clue.

---

<sup>65</sup> Ibid., 319-320.

## REGAN AND THE ASO

At best, Regan's theory is a partial theory. At worst, he has only systematized a list of moral intuitions. Nonetheless, we can apply it to the ASO and see what sort of an answer it provides. Immediately, it is clear that Regan is not going to have the difficulty Carruthers has providing at least an initially acceptable answer regarding the inherent value of both non-rational humans and animals. Of course, not all humans and animals are covered by Regan's theory (non-subjects-of-a-life are, again, an open question), but I will not linger on that point. What is more interesting is how Regan explains the differential treatment that is due to different individuals.

Like Carruthers's contractualism, Regan's rights theory is absolutist, rather than gradualist. You either have inherent value or you don't; no one has more inherent value than another. Therefore, for Regan, animals (at least, mammals of one year or more) deserve consideration equal to a normal adult human. In this respect, the theory is a clear departure from contractualism. Many find Regan's conclusion too radical, as most people believe that a human being has greater inherent value than some, if not all, other mammals. If it turns out that non-mammalian animals also qualify as subjects-of-a-life, as could very well be the case with some birds and reptiles, then his conclusion would appear all the more extreme. While this point alone is certainly not enough to lead one to reject the theory, adding this point to the underlying theoretical problems surely could be.

Interestingly, Regan does temper his theory's principles in a way that at first glance seems to approach a more commonsense moral perspective. While he maintains that all subjects-of-a-life are equally inherently valuable, he does not hold that all have an equal right to life. Recall that all have an equal right not to be *harmed*—due to the harm principle. And the worse-off principle requires that, given a choice where some harm must be inflicted on one group of beings or another, we should choose the lesser harm, regardless of the number of beings in each group. What Regan suggests is this: *a death constitutes a loss, or harm, that varies in magnitude depending on which individual dies.* Regarding a lifeboat thought-experiment where we must sacrifice one of four normal adult humans and a dog, he writes, "no reasonable person would deny that the death of any of the four humans would be a greater *prima facie* loss, and thus a greater *prima facie*

harm, than would be true in the case of the dog.”<sup>66</sup> To obey the worse-off principle, then, means that some lives are indeed worth more than others.

So, while the rights theory does offer a great deal of protection to the lives of both humans and animals, to the extent that in some respects all moral patients are considered equal, it also differentiates between different sorts of lives in a way that makes it acceptable to most people. However, there are glaring problems with this answer to the ASO that must be considered.

First, there is an arbitrariness about the harm principle and the manner in which it is derived from the respect principle. Regan argues that in order to show respect to a morally valuable individual, one acknowledges a *prima facie* duty not to harm the individual. Yet, it also follows that one would have a similar *prima facie* duty not to *kill* the individual. Regan gives us no reason why we ought to handle the two duties differently. He describes killing as a category of harming, but there is no reason to do so except that he derived the harm principle first. If he had begun by deriving from the respect principle the principle against killing (or, a principle allowing others to live), he could have then derived the harm principle (a principle allowing others to live *well*) from that. This sort of arbitrariness is the result of the theory’s vague decision model, as described earlier.

Second, Regan’s notion that the harm of death varies among individuals, and his subsequent conclusion via the worse-off principle about how to handle lifeboat dilemmas, is less acceptable than it appears at first glance. Concerning the choice between a normal dog and a normal human, Regan himself admits that his theory requires us to sacrifice *any number of dogs* if it would save a single normal human. After all, the worse-off principle tells us that the numbers do not matter. Because the death of each dog is less of a harm than the death of the normal human, the worse-off principle tells us to save the human, and thereby choose the lesser harm. Perhaps many people would consider this a reasonable conclusion. However, consider the fact that Regan is no speciesist: he accepts the conclusion of the ASO, and is therefore willing to swap out the normal dogs for any other similar beings, including human non-full persons. The startling claim that the lives of a million non-rational humans should be allowed to die in order to save a single

---

<sup>66</sup> Ibid., 324.

normal human is a far cry from commonsense morality, yet it clearly follows from Regan's arguments.

All told, the primary difficulty with Regan's answer to the ASO is the lack of theoretical grounding and the consequent failure to provide a clear decision model. Even if he had managed to avoid his highly questionable answer to the lifeboat dilemma, he would still lack the tools necessary to make a persuasive case. Evelyn Pluhar, another defender of a rights-based moral theory, argues similarly against Regan, but attempts to present a similar theory while avoiding Regan's pitfalls. We turn to her theory now.

### **Pluhar**

The most serious flaw with Regan's case, according to Pluhar, is that it relies too heavily on moral intuitions.<sup>67</sup> Recall that it was this flaw that Peter Carruthers also exploited in his attack on Regan. Pluhar also finds fault with Regan's applications of his theory, specifically with how he applies his worse-off principle to the question of killing humans and animals. She presents an argument on that subject similar to mine, though it is not identical. I shall say more about that later, when I discuss the application of her theory to the ASO. First, let us consider her attempt to derive a more successful foundation for a rights-based theory.

### **PLUHAR'S THEORETICAL METHOD**

Pluhar avoids the pitfalls of Regan's theory by appealing to our rationality rather than our intuitions. She borrows an argument from Alan Gewirth, from his books *Reason and Morality* and *Human Rights*. The argument, which I will call the 'argument from purposive agency', proceeds from the point of view of a rational, purposive moral agent, and shows that consistency demands that the agent must acknowledge the rights of others.<sup>68</sup>

1. I am a rational purposive agent with goals that matter to me, that I hold as (non-morally) good, and that I want to achieve.
2. In order to accomplish any of my goals, necessarily I must have freedom and well-being, which I therefore must hold as necessary (but again, nonmoral) goods.

---

<sup>67</sup> Pluhar, 236-240.

<sup>68</sup> *Ibid.*, 241-4.

3. All other agents ought not to remove or interfere with my freedom and well-being; that is, I have *rights* to freedom and well-being.
4. Grounding my rights claim is the fact that I am a purposive agent, that I have goals that matter to me and that I want to achieve.
5. If being a purposive agent is a sufficient ground for having these rights, then all purposive agents have these rights.
6. Therefore, all purposive agents have rights to freedom and well-being.

Beginning with the reflective agent's conception of himself *as an agent*, the argument takes purely prudential premises and produces a moral conclusion. This move can initially appear illicit, and has been criticized as such.<sup>69</sup> However, I believe the argument is quite persuasive if read correctly.

What Pluhar and Gewirth are *not* saying is that rational purposive agents must attribute rights to themselves, and therefore all purposive agents have rights. This third-person way of spelling it out fails to capture the nature of the argument, because the argument is necessarily from the point of view of a rational, purposive agent. The first two premises are *my* conception of *myself* as a purposive agent, with certain necessary goods that *I* require. I therefore insist upon and approve of the non-interference, on the part of others, with my possession of these goods. My insistence and approval is equivalent to my claim that I have rights (premise 3). At this stage in the argument, I am not making a moral claim; that is, I do not need to be able to justify my rights-claim to others. All that matters is that *I* accept it. In premise 4 I recognize that my acceptance of my rights-claim stems from the basic fact of my being a purposive agent. If I were not a purposive agent, and so did not have goals and necessary goods, then I would not have a reason to attribute rights to myself. Therefore, (via premise 5, a formal truth,) and still within the first-person perspective, I must claim rights for all purposive agents. I accept the conclusion, a moral claim about all purposive agents, because of my necessary rights-claim for myself.

If the argument from purposive agency works, it is easy to see why the conclusion is moral, despite the nonmoral nature of the premises. Since all rational, purposive agents must accept the conclusion from their own point of view, it is imperative upon each of them to recognize the rights of each of the other purposive agents.

---

<sup>69</sup> See for example R.M Hare's criticism in his "Do Agents Have to be Moralists?" in *Gewirth's Ethical Rationalism*, ed. Edward Regis, Jr. (Chicago and London: University of Chicago Press, 1984), chapter 4.

I find the overall structure of the argument highly persuasive, and have only one major concern with the content, which regards the nature of rights. As Pluhar describes the argument, my claiming rights to freedom and well-being (in premise 3) seems to be the logical correlative to the expression, “all other agents ought not to remove or interfere with my freedom and well-being.” It is likely that Pluhar wants the notion of rights to contain more substance than this in order to locate rights (rather than correlative duties) at the center of her theory. It also bears mentioning that Gewirth has argued specifically for the primacy of rights before duties.<sup>70</sup> Unfortunately, Pluhar never explains exactly how we are to understand this introduction of rights. This problem, I believe, is central to the difficulties Pluhar ultimately faces regarding the ASO.

Nonetheless, Pluhar’s use of the argument from purposive agency introduces a critical element that is not present in Regan’s case: a theoretical foundation. As we saw, the lack of a sufficient foundation and decision model was the primary flaw in Regan’s answer to the ASO. In order to assess Pluhar’s solution, recall Carruthers’s ‘governing conception’ requirement for an acceptable moral theory. A governing conception explains 1) the nature of morality, what moral ideas are about; 2) how we come to have moral knowledge; and 3) the basis of moral motivation.<sup>71</sup> I believe Pluhar’s theory answers those requirements in the following manner:

- 1) Morality is a set of principles respecting each individual’s rights to freedom and well-being.
- 2) We come to know moral truths by recognizing our concept of ourselves as purposive agents, and rationally constructing principles consistent with that concept.
- 3) There is a basic human need to justify our actions; justification requires consistency, and consistency demands moral behavior.

Theoretical foundation firmly in place, we can now assess the way Pluhar handles the ASO.

## PLUHAR AND THE ASO

The argument from purposive agency demonstrates that purposive agency itself is the primary morally relevant property. Purposive agency, the possession of goals and the

---

<sup>70</sup> Alan Gewirth, *Human Rights* (Chicago and London: The University of Chicago Press, 1982), 14-15.

<sup>71</sup> Carruthers, 23.

ability to act, is found not only in full persons, but also in many non-rational humans and animals. Accordingly, Pluhar accepts the second premise of the ASO, that some animals possess sets of morally relevant characteristics similar to those of some humans. She does not rule out the possibility of circumstances that require us to favor one individual over another, but, circumstances aside, it is the fact that the individual has *goals that matter to it* that generates moral value.

Pluhar refers to two versions of the ASO: a ‘categorical’ version and a ‘hypothetical’ one.<sup>72</sup> The latter version is very similar to the one I have employed in this paper. The categorical version contains an additional premise (3), and a stronger conclusion:

1. Individuals who possess similar sets of morally relevant characteristics are similarly inherently morally valuable.
2. Some nonhuman animals possess sets of morally relevant characteristics similar to those of some humans.
3. Non-rational humans are as inherently morally valuable as normal adult humans.
4. Therefore, nonhuman animals possessing sets of morally relevant characteristic similar to some humans are as inherently morally valuable as normal adult humans.<sup>73</sup>

One would think that Pluhar introduces this version of the ASO in order to prove its soundness, yet it is unclear that she does so. The additional premise (3) is ambiguous. Is it the claim that all non-rational humans are as inherently valuable as normal adults, or merely that some are? Pluhar does not tell us. If the claim is the former, then her theory does not cohere with this argument: some humans—fetuses in an early stage of development, the irrevocably comatose, and other extreme cases—are not purposive agents. Pluhar admits that these humans are not granted moral status by her theory.<sup>74</sup> However, if the claim (and the wording in the conclusion) regards only *some* non-rational humans, then her theory does support the argument.

Like Regan’s rights theory, Pluhar’s explains the wrongness of cruelty to animals and non-rational humans more successfully than Carruthers’s contractualism. Carruthers must rely on arguments showing that the cruelty is wrong without it being wrong *to the*

---

<sup>72</sup> Pluhar, 63-66.

<sup>73</sup> *Ibid.*, 64.

<sup>74</sup> *Ibid.*, 259.

*animal or non-rational human*, because he holds rational agency as the morally relevant characteristic. Regan's and Pluhar's theories do not face this challenge. And Pluhar improves upon Regan by including a moral foundation. It is now time to examine how her theory leads to specific applications. As we shall see, she differs with Regan on certain specific issues. While I agree with some of her criticisms of Regan, I believe that significant difficulties also arise under her theory.

Regan's worse-off principle, remember, states that given a situation where we must either harm one group of beings or harm to a greater degree a smaller group, we ought to harm the larger group. He holds the additional belief that animals are not as harmed by death as humans are, so that, given a lifeboat case where either an animal or human must be sacrificed, he believes we should prefer the death of the animal. I noted earlier that Regan's position also requires that we sacrifice a million dogs, or a million non-rational humans, to save a single full person. I also pointed out that Regan's position is contingent upon his view of the right not to be killed as derived from the right not to be harmed, which seems arbitrary and just as easily could have been the reverse.

Pluhar does not lodge these criticisms against Regan, but she, too, disagrees with him on this point. The death of a dog and the death of a full person, she argues, are equivalent harms.<sup>75</sup> Supporting this contention is the argument from purposive agency, which locates moral value in the individual's possession of goals. Both individuals, in death, lose the opportunity to pursue goals *at all*. Therefore, their losses are equal, regardless of the nature of their particular goals. Pluhar notes that Regan should have arrived at this conclusion as well, given his claim that all subjects-of-a-life are equally inherently valuable, despite the richness of their experiences.<sup>76</sup> I am not sure her arguments on this point are altogether successful—more on that shortly.

The worse-off principle served two purposes for Regan. First, it generated an anti-utilitarian decision model that better protects individuals' rights. In combination with the claim that animals are harmed less by death than humans, his principle also introduced a significant element of gradualism into an otherwise absolutist morality. Though Pluhar most likely would stand by the worse-off principle for the first reason

---

<sup>75</sup> *Ibid.*, 289-292.

<sup>76</sup> *Ibid.*, 289.

(though she never explicitly says this<sup>77</sup>), she does not believe it serves the second, gradualist purpose. She does arrive at some positions that have a gradualist tone, but via other routes that maintain the absolutism at her theory's core. Focusing on dilemmas where one or another individual must die, Pluhar locates three factors that can make the death of one of the individuals morally preferable. These three factors are: distress, the relative complexity of simpler beings, and nearness to natural death.

First, it is possible that one individual would be more harmed by death than another because one would experience a greater level of distress in the process.<sup>78</sup> Since Pluhar holds that death itself harms each individual equally, any prior distress for one of them would tilt the scales. The death of the individual who would suffer less in the process is morally preferable.

Second, the death of an extremely simple being may be morally preferable to that of a more complex one. Faced with the need to kill an animal (e.g., for food, to survive when edible plants are unavailable), Pluhar believes it is morally preferable to kill a clam than a fish, and killing either of them is preferable to killing a chicken.<sup>79</sup> This is not based on the comparable intelligence of each animal, but rather on each individual's ability to care about its goals, since purposive agency is the morally relevant property. Now, there are two ways one can read Pluhar here, and I will comment on each reading.

On the first reading, Pluhar is claiming that a fish may care about its goals, but that, if it does, it cares *less* than the chicken cares. This reading produces a situation where some individuals are more morally valuable than others, based on their level of concern with themselves. Pluhar seems agreeable to this possibility: "[W]e should not kill without good reason. How good the reason must be depends, of course, on how morally significant the 'killee' is."<sup>80</sup> However, it seems to me that this line of thinking runs afoul when we consider choices between killing two similar individuals, say, two normal adult humans. Should we try to ascertain which human cares more about her

---

<sup>77</sup> Her wording of the worse-off principle concerns only choices involving the harm of one individual or another; she does not specifically address situations where different numbers of individuals are involved. It nonetheless seems reasonable for her to agree with Regan on this matter.

<sup>78</sup> Pluhar, 292-293.

<sup>79</sup> Ibid., 259.

<sup>80</sup> Ibid., 258-259.

well-being? I imagine Pluhar would strongly oppose this result, especially because of her firm conviction that all purposive agents have an equal right to life.

On the second reading, she attaches moral value to the *probability* that an individual is a purposive agent. This reading is, I think, the right one. It is clearly supported by Pluhar's claim that "beings *more likely* to be consciously purposive than others, even if they are not clear cases, should...be spared if we have the option" (emphasis hers).<sup>81</sup> However, her notion of "how morally significant the 'killee' is" does not seem appropriate on this reading; perhaps she means to say, "how *likely* to be morally significant." But how are we to handle the relative probabilities? Consider the choice between killing a clam or killing a fish. The fish is far more likely to be a purposive agent, but one clam might not provide a full meal. If the choice is between one fish and several clams, is it still preferable to kill the clams? Pluhar does not tell us. It seems wrong to kill too many possible moral patients in order to spare a single, more likely moral patient: the harm done, if indeed the simpler individuals turn out to be moral patients, is tremendous.

The third factor able to tip the scales in favor of one life over another is the age, or nearness to death, of the individuals: "An octogenarian who has had a lifetime to formulate and fulfill goals is harmed less by death than a teenager, even if the teen has developed far fewer interests at that point."<sup>82</sup> The same goes for the terminally ill, and others who are close to death. The harm is unavoidable for those individuals, and this fact mitigates the wrongness of killing.

Pluhar does not fully explain this point, and leaves some worries unanswered. She is granting additional moral value to the projected length of one's life, a move seemingly in conflict with her earlier declaration that all lives are of equal worth, regardless of their content. Does this mean that we should attach moral value only to quantity of experiences, and not quality? One would think she would attach moral value to neither quantity nor quality of experiences; after all, she said that death deprives every individual equally, that each individual loses everything. It does make intuitive sense that an individual who will die tomorrow is not suffering a much greater harm by dying

---

<sup>81</sup> Ibid., 259.

<sup>82</sup> Ibid., 292.

today, but Pluhar does not tell us what grounds this judgement. If the ground is indeed the remaining quantity of experience, then we can apply this line of thought to many more lifeboat dilemmas than just those containing octogenarians and the very near-death. Between an average teenager and an average thirty-year-old, the former's life is more valuable. Between an average thirty-year-old and any age dog (as dogs do not normally live more than twenty years), the thirty-year-old's life is more valuable. And a young Galapagos tortoise (assuming it counts as a purposive agent, a possibility Pluhar leaves open), who may live two hundred years, has a more valuable life than any human! I doubt Pluhar would accept any of these judgements of relative moral value, but she does not give us a reason not to arrive at them.

Of the three factors mitigating the harm of death, two are at least questionable. Moreover, Pluhar's acceptance of Regan's worse-off principle implies that we should apply Regan's worse-off principle to cases where two *groups* of individuals are involved. Recall Regan's argument that the numbers do not matter; it was this aspect of Regan's principle that led to his conclusion that a million non-full persons ought to be sacrificed to spare a single full person. Therefore, we ought to be able to apply the million-to-one result to Pluhar's three factors. A million painless deaths are preferable to a single painful death; a million simpler individuals should be sacrificed instead of a more complex individual; a million octogenarians or a million dogs should be sacrificed rather than a young, healthy human. These results are not so counterintuitive as to be fatal to Pluhar's theory, but it should be noted that she has not ultimately overcome the sort of problems we encountered examining Regan's theory.

Pluhar's moral foundation, on the other hand, is her significant contribution to the rights theorist's answer to the ASO. The argument from purposive agency provides the right kind of grounding for an acceptable theory, and effectively counters the most serious criticisms lodged against Regan. There remain some lingering doubts about how Pluhar arrives at some of her secondary principles,<sup>83</sup> but we can nonetheless credit her

---

<sup>83</sup> Pluhar never explains how we might arrive at the worse-off principle via the conclusion of the argument from purposive agency. She also proposes the existence of acquired rights and duties, without any specific argument in their favor (Pluhar, 283 and elsewhere). Other principles, such as her 'liberty principle' (which I do not address in this paper) similarly lack arguments in their favor (ibid., 298-300). Moreover, some applications of her theory, such as her line on abortion (ibid., 253) are in conflict with these same

with a theoretically compelling answer to the ASO. And compared to the answer based on Carruthers's rationality-based contractualist theory, Pluhar's answer is undeniably more acceptable.

---

principles. Most of these principles and applications are developed strictly out of intuitive notions about moral rights, and are therefore vulnerable to the same criticisms as those made against Regan's theory.

## **CHAPTER 4: THE UTILITARIAN RESPONSE**

In the preceding chapters I introduced three non-utilitarian theories and evaluated their respective answers to the ASO. Contractualism denies the conclusion of the ASO, but fails to account for a wide range of moral beliefs regarding non-full persons. The contractualist must conclude that all non-full persons, whether animal or human, are hopelessly excluded from real membership in the moral community, as none of the attempts to grant them secondary consideration are ultimately successful. Tom Regan's rights theory proves more successful in handling non-full persons, but lacks a theoretical foundation. The arbitrary nature of Regan's moral principles leads to his failure to answer the ASO adequately. Finally, Evelyn Pluhar's rights theory combines the best aspects of the first two theories: a solid theoretical foundation like that of Peter Carruthers's contractualism, and the persuasive theoretical application that Regan had sought.

It is remarkable that each of these philosophers, before defending their own respective theories, spends considerably more effort arguing against utilitarianism than any other rival theory. In the writings of all three, utilitarianism appears to play the role of 'runner-up' theory. In this chapter, I describe the theory of utilitarianism in its various formulations, and produce an initial response to the ASO. In doing so, I discuss the theory's initial appeal, the source of utilitarianism's 'runner-up' status. Next, I note the major criticisms against the theory, and answer each charge in turn. Much of my discussion refers to Peter Singer's utilitarian formulation, which is largely successful against its opponents, but I also attempt to revise the theory in order to handle the most difficult challenges.

### **Utilitarianism's Theoretical Foundation**

Utilitarianism is a moral theory that has as its core principle, 'maximize utility', the term 'utility' understood to mean nonmoral goods, such as pleasure, happiness, desire-satisfaction, or well-being. Underlying a moral agent's acceptance of the core principle is the agent's recognition of the fact that her own interests are no more important, from an objective perspective, than the interests of anyone else. The agent who does not recognize this fact is not thinking ethically. Peter Singer argues along this

line. He claims that, to think ethically, we must “go beyond the ‘I’ and ‘you’ to the universal law, the universalisable judgment, the standpoint of the impartial spectator or ideal observer, or whatever we choose to call it.”<sup>84</sup>

Intuitively, Singer’s claim makes sense. But as we have seen, appeals to intuitions are often not persuasive. I believe a more solid case can be made for a utilitarian starting point, if we reexamine Pluhar’s argument from purposive agency. While it may seem ironic if not implausible to use her theory’s foundation in order to ground utilitarianism, I believe it is the correct line to take. Let us begin by presenting her argument, again:

1. I am a rational purposive agent with goals that matter to me, that I hold as (non-morally) good, and that I want to achieve.
2. In order to accomplish any of my goals, necessarily I must have freedom and well-being, which I therefore must hold as necessary (but again, nonmoral) goods.
3. All other agents ought not to remove or interfere with my freedom and well-being; that is, I have *rights* to freedom and well-being.
4. Grounding my rights claim is the fact that I am a purposive agent, that I have goals that matter to me and that I want to achieve.
5. If being a purposive agent is a sufficient ground for having these rights, then all purposive agents have these rights.
6. Therefore, all purposive agents have rights to freedom and well-being.<sup>85</sup>

I believe the following revision to the argument is both more persuasive, and can point the way toward a utilitarian morality:

1. I am a rational purposive being with goals that matter to me, which are therefore (nonmorally) good.
2. All agents ought not to interfere with the achieving of my goals; rather, they ought to promote them.
3. Grounding my ought-claim is the fact that I am a purposive being, that I have goals that matter to me and that I want to achieve.
4. If being a purposive being is a sufficient ground for these ought-claims, then those ought-claims apply to all purposive beings.
5. Therefore, as a rational being I ought not to interfere with the goals of all purposive beings, but rather promote them.

The most significant item that I have added to the argument is the notion of the *promotion* of goals in addition to non-interference with them. I do not believe this addition runs counter to Pluhar’s version of the argument, but builds upon it in an

---

<sup>84</sup> Peter Singer, *Practical Ethics*, 2<sup>nd</sup> ed. (Cambridge, UK: Cambridge University Press, 1993), 12.

<sup>85</sup> Pluhar, 241-244.

uncontroversial way.<sup>86</sup> I have also changed ‘agent’ to ‘being’, except where the notion of agency was critical (I cannot make an ought-claim to non-agents). This is a reasonable modification. Pluhar means nothing more by ‘agent’ than a being with goals it wants to fulfill.<sup>87</sup> But if being a purposive agent merely entails the possession of goals, then the expression ‘purposive agent’ is redundant. More important is what I have removed: the notion of rights.<sup>88</sup> Although it might be argued that the introduction of rights in Pluhar’s third premise necessarily follows from preceding premises, Pluhar offers no such argument. She refers to rights as the “logical correlative” of ought-claims,<sup>89</sup> but clearly rights are meant to add something more than the ought claim; if not, my version of the argument, which includes ought-claims, subtracts nothing from her version. Note that Pluhar uses her version of the argument to generate rights-claims that are intended to be more demanding than the ought-claims available to the utilitarian. But, again, it is unclear where the strong notion of rights has its source in her argument.

Perhaps my version loses the notion of rights by arriving at a conclusion regarding the treatment of the goals of individuals, rather than the treatment of individuals themselves. Purposive beings, not their goals, are the proper objects of rights-claims. However, I believe this counts as an advantage of my version of the argument. Both versions have as their premise the claim that my goals are what I hold as good; Pluhar’s addition of freedom and well-being as necessary goods seems unnecessary. If my desire for freedom and well-being arises out of my desire to achieve my goals, why should the former desire become the focus of the argument? This is left unclear. But since both our versions introduce an initial ought-claim that regards not myself, but either my goals or the combination of my freedom and well-being, it seems that my version is more straightforward by maintaining the focus on those goals, rather

---

<sup>86</sup> As Pluhar specifically spells out the argument, agents “ought *at least* to refrain from” interfering with the freedom and well-being of purposive agents (emphasis mine). This suggests that agents perhaps should promote these things as well.

<sup>87</sup> Pluhar, 248-249.

<sup>88</sup> I have also removed a phrase from Premise 1. It is enough that the goals matter to me. The fact that I want to achieve them is incidental, and may not even be the case, as I draw out in my formulation of utilitarianism. The fact that some things matter despite my not wanting them seems to be neglected by Pluhar’s theory.

<sup>89</sup> Pluhar, 242.

than shift the focus to myself. For this reason, I do not see any obvious objection to my formulation.

Indeed, I believe my revision of the argument from purposive agency can serve as a persuasive but generic ethical foundation. It can serve as the basis not only for utilitarianism, but also for many other (though not all)<sup>90</sup> ethical theories. All I have meant to accomplish in my revision is a more generalized conclusion, one that underpins ethical reasoning without arriving at any particular moral theory. In doing so, I arrive at Singer's claim about the universal nature of ethics—going beyond the 'I' and 'you'—without reliance on intuitions.

If we accept Singer's claim, the next step is to determine a specific moral principle or set of moral principles. An initially appealing principle turns out to be, in fact, a utilitarian principle. The conclusion of my version of the argument from purposive agency does produce such a principle, if we equate the non-interference with, and promotion of, the goals of purposive beings to the maximization of utility (nonmoral goods). This equivalence is perfectly reasonable given the first premise of the argument, that goals are what count as nonmoral goods. Depending on how we spell out the goals of purposive beings, we can arrive at a variety of utilitarian principles. Later, I will present and discuss these varieties.

Singer arrives at the same principle, though he talks of 'interests' instead of goals.<sup>91</sup> The difference at this stage is unimportant; later on I will make distinctions in order to refine the theory. What is important is that, at this point, utilitarianism is revealed to be what Singer calls a 'minimal' theory: when we reason from the pre-ethical to the ethical, "we very swiftly arrive at an initially utilitarian position."<sup>92</sup> That we do so is largely agreed upon. Carruthers, Regan, and Pluhar all recognize the initial plausibility of utilitarianism; it is a primary reason why utilitarianism remains their runner-up theory.

"Utilitarianism has the reputation of being a theory with considerable appeal," writes Singer. He continues,

Some are attracted to it merely by its simplicity, but there is more to its appeal than simplicity, as is shown by the fact that those who defend

---

<sup>90</sup> For instance, the premise that purposive agency grounds ought-claims would not be acceptable to a contractualist such as Carruthers, who grounds ought-claims in rational agency.

<sup>91</sup> Singer, 13-14.

<sup>92</sup> *Ibid.*, 14.

pluralistic ethical theories almost always include some kind of utilitarian principle among the things they value or regard as duties. While it is common for writers in ethics to deny that utilitarian considerations are the only valid moral considerations, it is quite rare for them to deny utilitarian considerations any place at all in their moral systems.<sup>93</sup>

Here we find another factor behind utilitarianism's runner-up status. Regan, for instance, admits that consequences do matter morally, though his theory is on the whole non-consequentialist.<sup>94</sup> Pluhar also allows consequences to play a role, as evidenced in my discussion of her theory in the previous chapter; the notion that being close to death reduces an individual's claim against being killed, for instance, introduces a consequentialist attitude. And none of Regan, Pluhar, or Carruthers would deny that utilitarian considerations are sometimes appropriate: causing only a few individuals to suffer a specific harm is undeniably preferable to causing many to suffer the same harm, just as causing a single individual to suffer a lesser harm is preferable to inflicting a greater harm on (either the same or another) individual.

Carruthers praises utilitarianism for its 'governing conception'. In the previous chapter I explained Carruthers's use of that term, and proposed a governing conception for Pluhar's rights theory. The following is utilitarianism's governing conception, according to Carruthers:<sup>95</sup>

- 1) The nature of morality: Morality is a set of principles aimed at maximizing utility.
- 2) Moral epistemology: We come to know moral truths by empirically determining how best to maximize utility.
- 3) The basis of moral motivation: There is a basic human feeling of sympathy for others with whom we have contact. The faculty of reason motivates us to broaden our sympathy into consideration beyond our immediate surroundings, to universal consideration.

Carruthers thus determines utilitarianism to have a theoretical foundation equally persuasive as his own contractualism or—given my analysis in the foregoing chapter—Pluhar's rights theory.

---

<sup>93</sup> Peter Singer, "A Utilitarian Population Principle," in *Ethics and Population*, ed. Michael D. Bayles (Cambridge, Massachusetts: Schenkman Publishing Company, 1976), 85.

<sup>94</sup> Regan, 310.

<sup>95</sup> Carruthers, 26-27.

Utilitarianism's theoretical foundation is compelling. The reason behind its rejection by opponents is therefore not the foundation, but the alleged failure of its application. We turn now to its application.

### **Utilitarianism and the ASO**

It should be noted that I have not yet addressed a specific utilitarian theory, but only a generalized version, without a specific definition of utility or an explanation of how it is to be maximized. Nonetheless, we can use the generalized version to produce a partial answer to the ASO. For whether utility can be measured in terms of pleasure, happiness, desire-satisfaction, or well-being, the utilitarian must accept the second premise of the ASO: Some nonhuman animals possess sets of morally relevant characteristics similar to those of some humans. Some nonhumans are capable of feeling pleasure and pain, of having desires, and of having a well-being to the same degree as some non-rational humans or even full persons. Therefore, the utilitarian must accept the conclusion of the ASO: some nonhuman animals and some humans are similarly morally considerable.

Now comes the question: and how morally significant is that? We saw that contractualism was led to the uncomfortable answer: not at all. The rights theories delivered more compelling responses, and in absolute terms: all purposive agents (or subjects-of-a-life) are equally morally considerable. The utilitarian response is typically gradualist: the moral significance of an individual is contingent upon the extent to which that individual contributes to overall utility.<sup>96</sup> If there are three individuals, two possessing five units of utility (regardless of how that might be defined) and the remaining one possessing ten utility-units, then the moral significance of the third individual is equivalent to that of the other two.

This is an admittedly simplistic utilitarian response, but it concurs to some extent with contemporary popular moral convictions about the relative moral status of humans and animals. As long as animals are considered to have a lesser capacity for possessing nonmoral goods, we can view them as less important, morally, than humans. However,

---

<sup>96</sup> It might be more appropriate to say that moral value is contingent upon the extent to which the individual *can* contribute to overall utility. At this point, however, I am merely establishing a simplistic utilitarian starting point.

the ASO demands that any humans relevantly similar to those animals are similarly less valuable. It is on this point that the utilitarian view is sharply opposed to the popular view. This situation is a natural result of the ASO, which demands similar consideration of similar beings, despite the contemporary view of a large divide between all humans and all animals. There are two directions, each unpopular, that the defender of the ASO can take: either hold that some animals are much more morally valuable than commonly thought, or hold that some humans are much less morally valuable than commonly thought. Both paths have been defended; R. G. Frey is famous for taking the latter route.<sup>97</sup> I, like Singer, take the first route. Of course, both routes may be taken in combination, and, as we will later see, Singer's and my views may be perceived to do so.

Now we shall move beyond the simplistic, generalized utilitarian answer to the ASO, and toward the specific utilitarian formulation I want to propose.

### **Varieties of Utilitarianism**

The utilitarian's fundamental principle is 'maximize utility.' There are therefore two questions that one must answer in order to arrive at a specific utilitarian theory: 1) What is 'utility'? and 2) How does one 'maximize' it?

### **UTILITY**

What is utility? Classic utilitarians such as Jeremy Bentham and John Stuart Mill refer to happiness or pleasure as utility. Mill calls the utilitarian principle the "greatest happiness principle," and describes happiness as "pleasure and the absence of pain."<sup>98</sup> But happiness and pleasure are not merely to be understood as physical pleasure; Mill stresses the importance of intellectual pleasures as well, noting that they are generally considered preferable to ephemeral physical pleasures.<sup>99</sup> Nonetheless, the classic formulation has come to be understood as *hedonistic utilitarianism*, due to its focus on pleasure and pain.

---

<sup>97</sup> See his line on vivisection in his *Rights, Killing, and Suffering* (Oxford: Basil Blackwell, 1983), 115-116.

<sup>98</sup> John Stuart Mill, *Utilitarianism*, Oskar Piest, ed. (New York: Macmillan Publishing Company, 1957), 10.

<sup>99</sup> *Ibid.*, 11-15.

Peter Singer's formulation is a departure from hedonistic utilitarianism. His formulation, called *preference utilitarianism*, refers to interests rather than pleasures: actions resulting in the satisfaction of interests are right actions, and actions resulting in frustrated interests are morally wrong.<sup>100</sup> However, preference utilitarianism is not drastically dissimilar from classic, hedonistic utilitarianism. After all, our preferences and our happiness/pleasure tend to go hand-in-hand. It is unusual to hold a preference for something the obtaining of which does not bring us pleasure, or at least decrease pain. Singer is aware of the similarity between the two formulations, and suggests that they may even be identical.<sup>101</sup>

There are a whole host of standard objections to utilitarianism that are applicable to either utilitarian formulation. One such objection targets the idea that the good is what we happen to prefer, or what makes us happy. The philosopher Robert Nozick captures this objection in his discussion of an 'experience machine'.<sup>102</sup> This imaginary machine completely removes you from reality, but produces in your mind the experiences of doing the things you love, and fills you with a sense of extreme happiness and well-being. If the resulting happiness would outweigh the happiness of living in the real world, one would think that a utilitarian would recommend that you plug yourself into the machine. (We are assuming that the necessities of staying alive in the real world—food, shelter, etc.—are accounted for.) This conclusion, however, tends to fly in the face of what is understood to be morally good activity.

This objection works best against a truly hedonistic notion of utility, as it presupposes that what we desire is merely a *sense* of well-being, and not actual well-being. If being happy, being satisfied emotionally, is the true end of our desires, then this objection carries significant weight. However, preference utilitarianism does not lead us to this view. What we really prefer (or most of us, anyway) is our *actual* well-being, and therefore the experience machine would not satisfy our preferences. Of course, the machine would *seem* to satisfy them, but then all we conclude is that plugging into the machine *seems* to be morally right.

---

<sup>100</sup> Singer, *Practical Ethics*, 2<sup>nd</sup> ed., 94.

<sup>101</sup> *Ibid.*, 14.

<sup>102</sup> Robert Nozick, *Anarchy, State, and Utopia* (New York: Basic Books, Inc., 1974), 42-45.

A stronger objection, directed at preference utilitarianism, is that we do not always prefer what is good. I may in fact prefer to use the experience machine: maybe I am a lazy person who would rather feel good than work to achieve real goals. Worse yet, a person might prefer to do bad things: someone might enjoy lying, stealing, or killing. Should these preferences be counted towards overall utility?

One response to this objection is to bite the bullet, and agree that people who prefer to use the experience machine should do so, all other things being equal. And people who prefer to harm or kill others ought to do so as well, all things being equal. Of course, in the latter instance all things will not be equal, since the harm inflicted (assuming the victim is not a masochist) will make the action morally wrong. But insofar as any action satisfies a preference, it is a right action.

Another response to the objection is to refer to ‘rational preferences’ rather than our actual, often irrational preferences. Spelling out the notion of rational preferences is no easy task, but many utilitarians work in that direction in order to avoid the counterintuitive consequences of a simpler notion of preferences. Generally speaking, a rational preference is a preference developed with perfect knowledge about one’s well-being and long-term goals. If it turns out that the preference of using the experience machine, or of harming others, is irrational, then those preferences do not count toward real utility. We see here a way for the utilitarian to deny that all preferences are good, regardless.

The view of utility I propose is similar to the rational preferences model, and even a model based on interests, as long as ‘interests’ are spelled out a certain way. My utilitarian formulation is ‘welfare utilitarianism’, and equates utility with actual well-being, rather than preferences or perceived well-being. I am not the first to put forth this formulation. Robert E. Goodin is among those who support this view of utilitarianism, and he notes that the seeds for welfare utilitarianism can be found in Mill.<sup>103</sup> The welfare-utilitarian considers the moral patient’s well-being, what will benefit the life of the moral patient, rather than what the individual happens to desire. Of course, the individual’s preferences may largely coincide with what is truly in the individual’s

---

<sup>103</sup> Robert E. Goodin, “Utility and the Good,” in *A Companion to Ethics*, ed. Peter Singer (Oxford: Blackwell Publishers Ltd., 1993), 243.

interest. We often develop long-term plans under the impression that the success of the plan will result in the kind of life that is good for us. However, in cases where a preference is in conflict with what is in our interests, we ought not satisfy that preference.

I have used the term 'interests' to mean something different than Singer's notion of interests, which appears to be synonymous with preferences. When I mention an individual's interests, or something being in an individual's interest, I am talking about the individual's *welfare*. The same can be said for the goals of an individual. As a welfare-utilitarian, I will be indicating those goals the attainment of which contributes to well-being. Recall my version of the argument from purposive agency: the initial claim is that I have goals that matter to me. Welfare utilitarianism points to a specific understanding of this claim and the subsequent claims in the argument. For something to 'matter to me,' it must promote my well-being. Therefore, what I hold true is that I have goals that promote my well-being, and that the attainment of those goals is for that reason good.

There are objections to welfare utilitarianism, and one objection indicates a disadvantage it has compared to the classic or preference formulation. Whereas Singer can determine and rank preferences simply by asking the individual possessing them, the welfare-utilitarian cannot determine well-being by asking. Bear in mind, all of these formulations encounter difficulty when utility is compared between two different individuals. Even Pluhar's and Regan's rights theories suffer from this problem when their harm principle is applied. But the welfare-utilitarian's ability to compare two states of well-being, even within a single individual, appears more limited than the preference-utilitarian.

Goodin argues that the welfare-utilitarian is actually in an advantageous position compared to other utilitarians, especially when comparing utility between two individuals. He writes,

[T]he problem [of interpersonal utility comparisons] is a problem only for hedonistic or preference utilitarians. They are the ones asking us to get inside someone else's head. Welfare-utilitarians, by abstracting from people's actual preferences, definitely are not. We can know what is in people's interests, in this most general sense, without knowing what in particular is inside their heads. Furthermore, at some suitably general level at least, one person's list of necessary basic resources reads much like anyone else's. Whereas preferences, pleasures

and pains are highly idiosyncratic, welfare interests are highly standardized. All that goes a very long way toward helping to solve the problem of making interpersonal utility comparisons.<sup>104</sup>

Something similar can be said for comparing different states of well-being in a single individual. Subjective harms such as pains can be measured in terms of how they detract from the individual's welfare overall, rather than trying to determine just how much something hurts. This kind of measurement is what one makes when taking a child to the doctor: the child may experience the pain of a needle, but the child's welfare is served in doing so, regardless of the magnitude of pain involved. (Of course, one could invent an odd case where the pain was so intense that the child suffered long-term physical or psychological damage as a result. But these consequences can be added to the utility calculation without concern for exactly how the pain *feels*.) The welfare-utilitarian therefore has a real advantage over many other utilitarians, and this is especially true when we apply it to the ASO, since we are then dealing with individuals who very well may not be able to communicate to anyone their specific preferences, or pleasures and pains.

Singer's preference utilitarianism leads him to consider morally relevant any beings with preferences. This set of beings is virtually identical to the set of conscious beings. (It is hard to fathom a conscious individual with no preferences—though one might exist.) The welfare-utilitarian faces a more difficult challenge marking out the morally relevant characteristic; it simply will not do to draw the line at individuals with a welfare. After all, all living things have a welfare. For that matter, so do corporations, works of art, kitchen appliances, and friendships. All of these things can flourish (or at least work properly), be damaged, or be destroyed. An additional argument is required to narrow the focus of morality to a reasonable set of individuals.

Consider the theoretical foundation of utilitarianism: I begin by considering the value of my own well-being (or happiness, or preferences) and move to a universal standpoint, from which I conclude that all relevantly similar individuals also deserve consideration. In the first premise of the argument from purposive agency I necessarily have a well-being, so that is one necessary characteristic. The fact that my well-being 'matters to me' is not a subjective claim under the welfare-utilitarian view, as I have just

---

<sup>104</sup> Ibid., 246.

noted. But there is another element in that premise as well: the 'I' itself. In recognizing the value of my well-being, I am presupposing my own consciousness. It is my awareness, combined with my having a well-being, that leads to my insistence that my well-being be promoted. Without consciousness, I lose any reason to make the ought-claim. Therefore, consciousness as well as possession of a welfare are required for moral considerability. The foregoing argument narrows the focus of welfare utilitarianism to the set of conscious, purposive beings. Incidentally, this set of beings successfully includes those hypothetical beings who have a welfare but no preferences, beings who are excluded by preference utilitarianism.

## MAXIMIZATION

Given a concept of utility, how is it maximized? First and foremost, to maximize utility is to generate the largest amount of nonmoral goods among all moral patients. What counts as nonmoral goods depends on one's concept of utility; the welfare-utilitarian holds that the set of nonmoral goods consist of the fulfilling of welfare-goals. For example, all other things being equal, a world with no one going hungry is preferable to a world with hunger.

If we imagine a fixed population of moral patients, there is an obvious general answer as to how utility is maximized: consider the total welfare of all moral patients, and determine what change in total welfare would result from each of our choices of action. The action that brings about the highest total utility is the morally correct action. However, fixed populations are only commonly found in the short run, and many moral decisions must consider lives coming into and out of existence. For this reason, calculating utility becomes more complex, as does calculating how to maximize it.

One formulation of maximizing utility is the 'total view'. On this view, utility is totaled regardless of the number of moral patients contributing to that total. The total view produces some readily acceptable results: for instance, killing a moral patient is bad to the extent that doing so decreases total utility. The converse result, that increasing utility by bringing more lives into the world is morally right, is more controversial. Later on I will address specific criticisms that attack this aspect of the total view.

I believe the total view correctly describes how utility ought to be maximized, but utilitarians dissatisfied with the total view's implications have proposed the 'average view'. According to this view, utility is measured as an average among all moral patients, regardless of the total number of individuals. While bringing one more individual into the world may increase total utility, the average utility may decrease as a result of overpopulation or some other factor. In this case, the average view tells us not to introduce more individuals into the world. This result lends credibility to the average view.

However, there is another aspect of the average view that renders it unacceptable. Given a world with average utility  $U$ , the average view must hold as morally wrong the introduction of a new being with utility below  $U$ , even if that being has a life worth living.<sup>105</sup> The rightness of bringing another being into the world therefore depends on the current well-being of the world's inhabitants, a counterintuitive situation to say the least. The fact that the dependency is such that individuals in a world filled with miserable beings are morally right to add another miserable being, as long as the new being is at least slightly less miserable, whereas individuals in the best-off world are prohibited from adding another very well-off being who is slightly less well-off than average, reveals the absurdity of the average view. We should therefore prefer the total view, which faces its own criticisms but does not generate undeniably unsatisfactory principles.

Under both the total and average views, the distribution of goods is irrelevant. All that matters is the *aggregate* of nonmoral goods, regardless of whether it is totaled or averaged. Therefore, if we can generate the largest aggregate of nonmoral goods with an unequal distribution—even if only one individual benefits while all others suffer losses—then that distribution is morally preferable. This fact about utilitarianism has been the target of many of the theory's opponents, as I will discuss in the next section.

### **Criticisms of Utilitarianism**

In this section I review major criticisms against both utilitarianism and the utilitarian's answer to the ASO.

---

<sup>105</sup> Derek Parfit makes this point in his *Reasons and Persons* (Oxford: Clarendon Press, 1984), chapter 18. We are assuming that the well-being of others does not increase as a result of the introduction of the new being to the point where they counterbalance the utility of the new being.

## STRICT DECISION MODEL

Carruthers raises a general, theoretical objection to utilitarianism.<sup>106</sup> Unlike his contractualism, under which moral agents agree upon a set of moral principles, utilitarianism produces one and only one duty: to maximize utility. Actions are therefore divided into two categories: obligatory and prohibited. According to Carruthers, there is no category of permissible actions. Moreover, there is no possibility of supererogatory action; our duty is to maximize utility, and we cannot rise above the call of that duty. This simplistic nature of duties fails to give us what we need from a moral theory, and so utilitarianism must be rejected.

This objection is a reaction to the utilitarian's focus on states of affairs: from any given point in time there are a set of possible future states of affairs, and the utilitarian demands that we aim for the best state of affairs. The contractualist, on the other hand, holds that "the lives or interests of individuals cannot be interfered with merely to subserve greater overall utility," and leaves open "a substantial domain within which they remain free to get on with their own lives, and to develop their own concerns and interests, without direction from morality."<sup>107</sup> Because Carruthers (as well as Regan and Pluhar) concentrate on negative rights—the right not to be harmed or killed, the right to be free to pursue one's goals—utilitarianism appears to face a unique problem here.

One utilitarian response is hinted at (but rejected) by Carruthers: part of maximizing utility includes allowing individuals to be free to pursue their own interests and desires.<sup>108</sup> As a welfare-utilitarian, my focus on the well-being of individuals requires the acknowledgement of the kind of things we are. Without a biological and psychological understanding of humans and animals, it is impossible to determine what constitutes their welfare. Psychologically, it is clear that humans who are moral agents cannot be made to operate as utility-calculators, to "run around doing good all the time," (as Carruthers puts it,) without detracting significantly from their welfare.<sup>109</sup> The

---

<sup>106</sup> Carruthers, 32-34.

<sup>107</sup> *Ibid.*, 40.

<sup>108</sup> *Ibid.*, 33-34.

<sup>109</sup> *Ibid.*, 32.

utilitarian adds this consideration into the moral calculus, and concludes that some amount of freedom of action is absolutely necessary.

Carruthers finds this answer lacking. He replies that what the utilitarian is doing is making it obligatory for moral agents to do what they want.<sup>110</sup> This is a strange criticism. He seems to think that an obligation for him to do as he pleases—without concerning himself with morality—is a coherent obligation. What the utilitarian proposes is that some amount of our lives must be devoted to self-interest. After all, I am often in the best position to know how to improve my own welfare. In addition to, and often prior to moral considerations, I must look after my own health, shelter, livelihood, and other necessities of life. The particular manner in which I secure these is very much up to me; there is no obvious best way for any individual to live his life. That I look after my own interests never feels like a moral obligation, despite the fact that I recognize my own well-being as contributing to overall utility. Utilitarianism therefore does not enforce free time at certain intervals upon moral agents; rather, it ‘obligates’ us to freely choose whatever path we like to flourish as individuals. Real obligations only arise after we are living in a manner that allows time for moral thinking.

Moreover, utility-calculations can hardly generate the sort of precision that would always indicate one particular action as the clear means to maximize utility. Of course there are clear cases. We certainly know that killing a moral patient (with a life worth living) detracts from utility, and that saving them from being killed, when no similar harm would result, prevents a loss of utility. But there are an infinite number of less clear cases: the career path one takes, contributing to one benevolent charity instead of another, going to one social gathering instead of another—these are all choices where we have no way of gauging with precision the relative utility. It may well be that one choice *is* preferable to all alternatives, but until we have perfect knowledge, there is substantial moral freedom.

Something similar can be said about supererogatory action. If we do not know with certainty the comparative utility resulting from two choices, but one choice detracts significantly from one’s own welfare, then the less selfish choice can reasonably be considered supererogatory. In fact, choices seen as supererogatory often have this

---

<sup>110</sup> Ibid., 33.

feature. If someone sacrifices her own life to save the life of another, the resulting utility may be no higher than if she allowed the victim to die. But the decision to put someone else's welfare before one's own is both morally good (utility-maximizing) and beyond the call of duty (because utility-maximizing alternatives are available). Furthermore, the choice reflects a virtue in the person making the sacrifice—selflessness—a virtue any utilitarian would seek to promote in moral agents. Here we see how utilitarianism provides substantial room for freedom of action and supererogation, without turning either into a strict duty.

On this issue we can even turn the tables on the anti-utilitarians who bring up this charge. To the contractualist, we might ask why the imaginary bargainers would not agree to principles aiming at the best possible world. If maximizing utility is nothing less than bringing about the best possible world (and that's really the whole point), then why would we adopt principles that do less? The notion that we ought to respect rights, even when many are worse off as a result, is a common view but one that I think needs serious reexamining. Much has already been said in this regard by Singer, who has written a great deal in criticism of the emphasis on principles respecting rights.<sup>111</sup>

#### FAILURE TO PROVIDE ADEQUATE PROTECTION, PART ONE

The bulk of the objections against utilitarianism indicate applications of the theory that fail to provide adequate protection for individual beings. Rights theories and contractualism aim for stronger principles against harming others, regardless of whether these principles would maximize utility. I have already noted that something seems amiss in a moral system that aims at something other than the best possible world. But this is certainly not enough. More must be said in order to address specific objections to the application of utilitarianism.

One objection is made by Carruthers, and involves the relative moral value of humans and animals. He introduces a thought-experiment wherein you enter a burning house, and find one human unconscious on the floor and five dogs locked in cages.<sup>112</sup> You estimate that you only have enough time either to drag the human out, or release the

---

<sup>111</sup> See for instance his *Practical Ethics*, 2<sup>nd</sup> ed., chapter 8, especially 232-246.

<sup>112</sup> Carruthers, 9.

dogs from their cages and lead them out. He writes, “no one would maintain that you ought to place the lives of many dogs above the life of a single human,” and claims that there is a “common-sense belief that human and animal lives cannot be weighed against one another.”<sup>113</sup> Therefore, any theory that suggests otherwise is highly suspect, and we have good reason to dismiss it.

It is worth pointing out that Carruthers’s use of this argument to bolster his contractualist theory, and then to apply that theory in order to persuade us of his beliefs about the moral value of animals, is an undeniably circular strategy. However, if the popular moral belief he appeals to is “central to morality,” as he claims it is, then perhaps he makes a reasonable case.<sup>114</sup> I seriously doubt, however, that he has done so.

First, is it really true that we cannot weigh human and animal lives? When an animal research lab claims that the suffering and deaths of their animals are justified by the benefit to human lives, is this not the weighing of lives against one another, even if the result is in favor of the humans? If only one person, rather than many, would ever be saved by killing millions of animals in research, it does not seem that their work would be justified, or at least *as* justified. Consider Carruthers’s thought-experiment. What if there were not five dogs, but five thousand? Or five million? I find it hard to believe that the commonsense view is that the value of these lives never approaches the value of the human’s life.

It was on this very point that I found Regan’s worse-off principle suspect. Recall that his principle instructs us to sacrifice an infinite number of animals in order to save one (normal) human. Of course, the principle was suspect because Regan also holds the view that all moral patients are equally morally valuable. Carruthers does not hold the latter view; rather, he holds that animal life has no inherent value whatsoever, a view that leads (as we saw) to its own set of problems.

Pluhar, on the other hand, insists that the right action is to save the five dogs rather than the human. She accepts the worse-off principle, but does not agree with Regan that animals are harmed less by death. Given that the lives of animals have any moral value, it appears there are three possible answers as to how to weigh their lives

---

<sup>113</sup> Ibid., 9.

<sup>114</sup> Ibid.

against human lives. We can view their lives to be equal (Pluhar), view their lives as infinitely less valuable (Regan), or recognize a sliding scale of value, with normal humans more valuable—but not infinitely more valuable—than animals.

I believe the third route is the proper choice under utilitarianism, though a utilitarian could conceivably argue for any of the three. As a welfare-utilitarian, my evaluation of the life of any individual comes out of an analysis of its welfare. Now, there are two ways one can understand comparative overall welfare of individuals. One choice is to take the view of Pluhar, who argues that in death, all is lost for the victim, and so each life is equally valuable.<sup>115</sup> Her assessment is derived from her theoretical foundation, which bases moral value on the fact that an individual *cares* about his or her well-being. The other choice is to view overall welfare in such a way that what is lost in death is an amount that can be compared across individuals. That is to say, yes, the victim loses everything in death, as Pluhar puts it. But the *amount* signified by ‘everything’ depends on the individual: some have more to lose than others. On this latter view, some lives are indeed more morally significant than others.

The idea that individual lives can be more or less valuable surely agrees with popular morality when it comes to weighing lives in Carruthers’s thought-experiment, but we are far less likely to want to compare human lives in this way. Pluhar is firmly against the notion that lives are more or less valuable, though she did make exceptions for the very old or near death. Yet, if we can weigh human lives based on the comparative time one has left to live, why not on the quality of that time as well? I see no reason why we cannot justifiably make such comparisons.

We must be cautious not to take this idea too far, as there are clearly dangerous implications. R. G. Frey, as I have mentioned, has accepted the ASO and its demand that we view animals and relevantly similar humans as having lives of similar moral value. Frey’s response, as a proponent of animal experimentation, is to accept similar experimentation on relevantly similar humans.<sup>116</sup> But Frey bases his response on his own utilitarian formulation, which rests on a sophisticated notion of interests. According to Frey, animals and non-rational humans lack these interests to such an extent that their

---

<sup>115</sup> Pluhar, 293.

<sup>116</sup> Frey, 115-116.

lives are far more expendable than normal humans. As a welfare-utilitarian, I do not recognize a wide gap. There is a gradual scale, surely, but the differences between the welfare of a normal human and the welfare of a 'higher' animal or similar human are quite small.

I do not see an alternative to the gradualist answer without accepting the unlikely view that the simplest conscious organism has a welfare of equal value as dogs, apes, and human beings. Many philosophers rely on a method of cushioning gradualism: they postulate an imaginary line dividing moral patients on the basis of one or another morally relevant property. Those individuals above the line are considered maximally morally considerable; those under the line are placed on the gradualist scale. Under a moral theory centered on rational agency, rationality may provide the basis for that line. Singer, to some extent, uses the preference to continue one's life as the line: individuals with no conscious preference regarding their own lives fall below the line.<sup>117</sup> Philosophers have also drawn the line such that all rational agents, as well as all humans (including non-rational agents), are above the line; however, such divisions have been found to be unjustified, as I mentioned in chapter 2. The welfare-utilitarian does not have access to any such property to serve as a magic line.

However, I believe some sort of line ought to be imagined nonetheless. Because the welfare of one human being is impossible to compare to most others, it is best to consider them equal. This applies to many non-rational humans as well. After all, the richness of one's experience is merely one component of one's welfare. The web of friendships and other relationships is another. Other components include one's lifestyle, moral inclinations, access to necessary resources, and physical health. Facts about an individual's mental life surely cannot give us a complete understanding of the individual's welfare, nor is any final, precise assessment really possible. So, between two highly complex animals (human or otherwise), there may indeed be a difference in moral value. But we have limited means to determine it.

So, unlike Frey, I find reason to object to painful experimentation on many sorts of animals, and on humans. This does not rule out the possibility that some may be

---

<sup>117</sup> Singer, *Practical Ethics*, 2<sup>nd</sup> ed., 95-99. I will have more to say on Singer's position in a later section where I discuss his 'replacement argument'.

justified, but I will discuss this further in the next section, when I reply to the charge that utilitarianism fails to sufficiently protect many basic rights.

To return to Carruthers's objection, I think it is the dogs who ought to be saved, as long as we ignore all welfare considerations outside the six individuals in the thought-experiment. In the larger picture, we need to also be aware of the welfare of those whose lives were tied in with the human, though we must do the same for the dogs. I think it is safe to assume that in most (but certainly not all) cases the death of five dogs is not going to have the same repercussions on others as the death of a human. Moreover, the normal human is a moral agent and, if he is not morally corrupt, might be expected to do more good in his lifetime than the dogs in theirs. Also, like Pluhar, I see the length of one's life as relevant. If the human was young, the longer life expectation becomes another factor in the calculation.

But now, no one is going to expect the person running into the burning house to work all this out. No utilitarian insists that a utility-calculation be performed before every action, especially in an emergency. Sometimes it is necessary to rely on past reflections to produce an intuitive judgement, and this judgement may very well be to save the human. If the number of dogs were greater, perhaps the person would choose otherwise. In no circumstance would I find the person blameworthy; morality does not provide a final answer to every possible dilemma, a fact resulting from the limitations of human reason. All told, I think the welfare-utilitarian handles Carruthers's thought-experiment more successfully, and more in keeping with some of our commonsense moral ideas, than Carruthers himself does.

## FAILURE TO PROVIDE ADEQUATE PROTECTION, PART TWO

Carruthers, Regan, and Pluhar all criticize utilitarianism for its failure to protect rights. When the details of these criticisms are spelled out, I think the utilitarian has the necessary resources to counter their charges.

Regan argues that the utilitarian treats the individual as a mere means to the end of maximizing utility.<sup>118</sup> In doing so, the individual is not treated as an end-in-itself, the Kantian notion of what it is to be morally valuable. The individual does not matter for

---

<sup>118</sup> Regan, 311.

his or her own sake, but for the sake of utility. Regan describes the utilitarian understanding of individuals as “mere receptacles,” valueless containers of utility.<sup>119</sup> As valueless containers, their individual welfares are not important as long as the aggregate utility is maximized. Therefore, the utilitarian would justify shifting all utility into one individual, if slightly more utility is created in the process. If utility is already maximized, no increase in utility would even be necessary to permit such a utility shift.

To illustrate his point, Regan imagines a situation where killing one innocent victim would maximize overall utility.<sup>120</sup> Without the notion of a right to life, the utilitarian must permit the killing. What is worse, the utilitarian must find the killing morally obligatory. Similarly, a principle against lying is good only insofar as it maximizes utility. If by lying I can improve my own welfare without harming anyone to an equal degree, then lying is a morally right action. So it goes for all principles of rights theories; as long as utility is maximized, any rights or principles may be ignored. Pluhar makes a similar point: if two actions will maximize utility, but one action involves killing an innocent individual and the other does not, either action is morally acceptable.<sup>121</sup> Carruthers makes it plain: “Apparently anything can be done to a person, provided that it produces more utility (either total or average) than any alternative course of action.”<sup>122</sup>

A frequent utilitarian response is to appeal to side effects. There are several ways of doing this. For instance, one can appeal to the effects the killing of an innocent individual will have on the individual’s friends and loved ones. Of course, this appeal will not work if the individual has no friends or loved ones. And this defense will not work anyway: we can still suppose that utility is maximized despite any additional suffering. Another possible side effect is the bad character of the killer. The utilitarian can argue that even if the killing maximizes utility in this one instance, a person capable of calculated killing reveals a bad character. In the long run, utility will not be maximized if we permit bad characters to flourish, so we ought not allow the killing. But again, this response will not succeed: there are a variety of ways to reject the appeal to bad character. Perhaps the killer is near death or is otherwise removed from the company

---

<sup>119</sup> Ibid., 205-211. This description is not without merit: Regan credits Singer with the ‘receptacle’ terminology.

<sup>120</sup> Ibid., 202-204.

<sup>121</sup> Pluhar, 182-183.

<sup>122</sup> Carruthers, 28.

of others, and will not have time for his bad character to cause any harm.<sup>123</sup> Or perhaps the character of the killer is so well developed that he can distinguish between situations where killing is justified and situations where it is not. This is no extraordinary claim. Nations routinely send soldiers out to kill the enemy, but still welcome them back into society without fear that they have become dangers to society. The response to any appeal to side effects is the same. Side effects can be accounted for, but as long as utility is maximized, they will not significantly affect the utilitarian's evaluation of killing, lying, stealing, or any other actions strongly prohibited by rights theories.

Another response to the rights theorist's challenge is to propose the adoption of a rights scheme as the best means of maximizing utility. Pluhar discusses one such attempt in her criticism of utilitarianism, and calls the strategy "the best case that can be made for genuine rights within a utilitarian framework."<sup>124</sup> This defense begins with a point I made earlier, that moral agents cannot be expected to make utility calculations quickly and with precision. Add to this premise the claim that some choices must rely on taking certain risks, and that what is at risk is sometimes extraordinarily bad. For instance, suppose that a group of terrorists are planning an operation to kill many civilians. You are an officer in the military and intelligence reveals that there is a likelihood that the terrorists are holed up in a certain area containing innocent civilians. Suppose further that you are assured that the terrorists will disappear if they detect any military presence in the area. The only way to defeat them would be a remote strike on the area using missiles or bombs, which will kill innocent civilians as well. Now, the simplistic utilitarian calculation might be to weigh the number of innocent casualties lost in the missile strike with the number of lives threatened by the terrorists. If fewer lives are involved in the missile strike, then we should launch the strike. However, this simplistic calculation omits the very important risks involved. Perhaps the terrorists are not hiding there after all. Or, even if they are, perhaps their plans can be foiled in another way, with less loss of life. The addition of these risk considerations generates a strong presumption against launching the missile strike. After all, the strike is *guaranteed* to generate a great amount of disutility; whatever disutility is prevented is merely a matter of probability,

---

<sup>123</sup> This is suggested in Carruthers, 32.

<sup>124</sup> Pluhar 219-221. The particular argument she presents is borrowed from L. W. Sumner.

and the precise level of probability is only guesswork. The presumption against killing, in this instance and others with relevantly similar choices, can be construed as *rights* held by the innocent individuals. This notion of rights, as strong presumptions against certain harmful actions, matches up with our common use of the term.

There is merit in this argument for deriving rights from utilitarianism. I think it can be strengthened by additional considerations regarding the welfare of individuals. First, there is a frequent misconception about utility calculations, which is captured in Regan's criticism that individuals are mere 'utility containers' mattering only insofar as the utility they hold. The idea is that units of utility can be moved between individuals, even if one individual ends up with most or all of them, and that this would be preferred if utility was maximized in doing so. To an extent this is a correct characterization, in the sense that total utility is what counts, in the end. However, one tends to ignore the complexities contained in the notion of utility. Utility cannot be exchanged like currency: taking something valuable away from one individual and giving it to another does not necessarily maintain a constant amount of utility. This is plainly true in the case of wealth in our society, where more equal distributions allowing more people to rise above the poverty line result in greater utility than more unequal distributions. The rising out of poverty simply counts for more than, say, the rising from a comfortable income to an even higher income.

What we have here is a distinction between *trivial* and *necessary* nonmoral goods. A trivial good contributes to welfare, but not to any important extent. Trivial goods are often expendable, and easily replaced by similar goods. Necessary goods, on the other hand, are things like food, shelter, and adequate resources for a healthy life. Necessary goods often have no suitable replacement, and cannot be sacrificed without extreme consequences for the individual's welfare. There is a difference between trivial and necessary goods other than one of quantity; there is a qualitative difference. The use of a limb, for instance, is simply not the same *kind* of good as excessive wealth. No one would reasonably trade the use of a limb for a great deal of money, unless the person was so destitute that similarly necessary goods could be acquired thereby. Trivial goods, then, are worth *infinitely less* than necessary goods. There is simply no amount of trivial goods that can warrant the loss of a necessary good. This is an argument against the

notion that utility can easily be exchanged between individuals: if the exchange results in a loss of necessary goods by one individual and no gain in necessary goods for another, then the exchange results in great disutility.

There is a related point. Let us judge my losing an arm to equal ten units of disutility. Therefore, my losing an arm and a similar person losing his arm will add up to twenty units. The idea of ‘utility containers’ might suggest that I could keep the disutility constant by transferring all twenty units to the other person, costing him his other arm and leaving mine intact. But this is not right. Being without both arms is substantially more than twice as bad as living with one, due to the much more extreme handicap it causes to an individual. The same is true for necessary goods in general: the loss of multiple necessary goods adds up to more than the sum of those losses taken individually. Therefore, we ought to favor a wide distribution of harms among many individuals over a concentration of the same number of harms on just a few.

This principle comes fairly close to the worse-off principle favored by Regan and Pluhar. In fact, it can be seen as more successful. Recall that the worse-off principle prevented us from harming any one individual rather than harming many individuals to a lesser extent. I have just showed how utilitarianism generates a similar principle via two methods, a distinction of trivial and necessary goods and a claim regarding the additional disutility of multiple harms. However, the worse-off principle suffers from problems that mine does not. For instance, the worse-off principle gives us the same result concerning my lost limb example above, but would further suggest that it is preferable to cause a million people to lose a limb than to cause one person to lose two. Even worse, it makes it preferable to cause a million people to lose a limb than to cause one person to lose a limb and a finger! Or a limb and a dollar (or any other nonmoral good). One expects there would be some point where it is acceptable to cause a greater harm to prevent many lesser harms, at least when both harms involve necessary goods. The utilitarian can accept this possibility, but a rights theorist who accepts the worse-off principle cannot.

The claim by opponents of utilitarianism that the aggregate utility is all that matters, regardless of which individuals ‘contain’ the utility, misses a general point about how welfare is measured. An individual’s welfare is bound up in the individual, and can only be measured with respect to that individual. The welfare-utilitarian does not simply

add up all the happiness (or preferences) in the world to generate a measure of utility; rather, it is essential to take each individual separately, consider what sort of being it is and what comprises its welfare, and then evaluate the individual's overall well-being. Now, after considering the individual, utility ought to be aggregated. But this summing-up does not prevent the individual from being the primary locus of moral consideration. Rather, it reflects the fact that one individual's welfare is only as important as that of any other moral patient, and no more.

Despite these arguments, rights theorists are likely to object that rights remain inadequately protected. Consider my earlier example of the choice to kill innocents in order to stop terrorists. In the example I noted that the chance that utility would be maximized was unknown, but that we should not guarantee the loss of life in order to prevent a merely possible loss of life. But what if we were sure that the terrorists were in the area, and what if they were guaranteed to kill many more people if we did not launch a strike? In the unlikely event that these facts are absolutely known to us, then the utilitarian would launch the strike.

Pluhar seems to find this result unacceptable. She mentions a parallel situation, where medical experimentation is guaranteed to prevent more harm than it causes. "Even if we have decided that sentient nonhuman animals have a right to have their interests respected, a case could be made [under utilitarianism] for painful, fatal experimentation on them if overall utility would be maximized." She goes on to say, "[T]he same would apply to humans."<sup>125</sup> Her objection is that utilitarianism justifies rights violations too easily. Under her rights view, medical experimentation, regardless of benefits, is wrong. Her argument that the utilitarian fails to provide this stronger kind of protection is correct. But how weak is the utilitarian's protection? Pluhar asserts that the utilitarian prohibits experimentation unless "overall utility would be maximized," but she does not provide an example of when this would be the case. I argue that the case is rarely, if ever, actualized. We can never guarantee the results of experimentation—that is why they are experiments! If, somehow, we knew that more lives would be spared pain and death, and we knew that the injury and death spared would certainly outweigh the harms of experimentation, then I think painful, fatal experimentation would be justified. But it

---

<sup>125</sup> Ibid., 222.

should be emphasized that this is a highly extraordinary, if not utterly implausible situation. Under more realistic conditions, utilitarianism has no problem extending the kind of protection the rights theorist wants.

## REPLACEABILITY

A final challenge to utilitarianism is the ‘replacement argument’. Pluhar uses the notion of replaceability in order to reveal a weakness of utilitarianism.<sup>126</sup> The argument is supposed to show that one individual can justifiably be replaced by another, so long as utility is maximized in doing so, and can be understood as follows:

1. Relevantly similar moral patients possess lives of similar moral value (measured in utility).
2. The existence of one such moral patient is therefore equivalent to the existence of another.
3. The death of an individual is measured as the loss of whatever utility the individual possesses, as well as whatever harms the death causes in others.
4. Therefore, if an existing moral patient has a life worth living (i.e., counts as positive utility), and the existence of a future moral patient depends on the destruction of the existing individual, then we are permitted to kill and replace the first individual with the second, assuming no extra disutility (aside from the harm of death to the individual, but including the suffering of the individual before death) occurs in the process.

The conclusion, if sound, represents a serious challenge to the utilitarian. Since future lives add to total utility, they must be counted as part of the utilitarian’s moral evaluation of actions. The argument is typically applied to such things as the raising of animals for food. Since future animal lives are created by the same industry that kills them for their meat, the replacement argument points the way toward preferring a meat-eating society. This result is not terribly objectionable to many people in our society; however, the ASO reminds us that we must apply similar results to humans as well. And those implications are unacceptable to most.

What are the implications? First, let us note what is *not* implied. Our current animal food industries are not justified by the replaceability argument, on the grounds that the existing moral patients in factory farms do not have lives worth living. This fact has been well documented, and there is little need at this point to rehash the description of the suffering involved in factory farms. Pluhar claims that hunting (and then replacing

---

<sup>126</sup> Ibid., 185-190.

the hunted animals) is justified by the replacement argument,<sup>127</sup> but this is clearly wrong. Hunting fails to satisfy the condition that no additional disutility attend the death of the individual, disutility such as the fear and pain experienced by the hunted animal. Replacing individuals is also not acceptable when other individuals will suffer as a result. Human deaths tend to cause additional disutility in those that knew the victim, and the same is true for some animals.

However, the argument does apply in cases where animals—or humans—are raised humanely and killed painlessly, and where no others are harmed by their deaths. According to the argument, a life as a well-treated food animal, or as a subject of painless medical experiments ending in an early but painless death, is acceptable as long as the individual is replaced to avoid a loss of utility. The additional utility generated by tasty meat or increased chances of new medicines make the replacement choice obligatory, though only marginally (trivial and probable benefits do not count for much), and only if utility is actually maximized (other courses of action may do better).

Many find it highly counterintuitive to justify the killing of one individual by bringing a similar individual into existence. But that justification is exactly what seems to come out of the ‘total view’ of utilitarianism. Rights views, on the other hand, do not have the same result: presumably, the existing individual has a right to life that nonexistent beings do not, and so replacement is an unjustified rights violation. Pluhar holds that the “upsetting” problem of replaceability by itself is enough to drive many people away from utilitarianism.<sup>128</sup>

There is a temptation for the utilitarian to bite the bullet on this issue. If replaceability is permissible under very stringent conditions, then at a practical level it is rarely a real worry. Humans are in even less danger than animals of real replacement, since popular sentiments will not condone any killing of humans, regardless of whether they are replaced. Secret killings, of course, would bypass that problem. In the end, the uncertainty that any practical replacement industry will really maximize utility is probably enough to warrant a utilitarian prohibition on it.

---

<sup>127</sup> Ibid., 185.

<sup>128</sup> Ibid., 190.

Nonetheless, the opponents of utilitarianism stress that the replaceability issue indicates a theoretical flaw within the utilitarian theory, and specifically the total view. The theoretical problem is that utilitarianism seems to allow the goodness or rightness of creating one individual to cancel out the badness or wrongness of killing another individual. Our common sense, of course, tells us that the wrongness of killing far outweighs the rightness of creating a replacement individual.

One attempt to escape the replaceability problem is the ‘prior-existence view’.<sup>129</sup> This is an alternative to the total view, and provides moral consideration only to existing beings, not future individuals. Singer held this view, but has since discarded it.<sup>130</sup> The prior-existence view successfully explains the notion that bringing an individual into existence is not tantamount to benefiting that individual, whereas killing an existing individual counts as a harm. This view, however, is too flawed to accept. Quite simply, it is not acceptable to exclude future generations from the moral community. An exclusive focus on present individuals, with no regard for future generations, is a shortsighted morality. For instance, under the prior-existence view there would be no reason to protect the environment, so long as the conditions did not greatly deteriorate in our own lifetimes. And how can it be that future generations matter in the future, but deserve no consideration now?

A specific thought-experiment has been used by Singer and Pluhar to illustrate the problem with the prior-existence view.<sup>131,132</sup> Pluhar calls it the ‘wretched child’ scenario, and it runs as follows. Suppose that a couple is considering having a child, but they are told by a doctor that there is a very high probability the child would be afflicted with a birth defect that would render the child’s life miserable and brief; the child would have a life not worth living. The couple has a very good reason not to conceive a child, and that reason is the child’s suffering. However, this child does not yet exist. Under the prior-existence view, there would be nothing wrong with conceiving the child *up until the point that it exists*. Once the child exists, it would be wrong to let the child live. Therefore, the prior-existence view permits the conceiving of a miserable life, but obligates us to

---

<sup>129</sup> Ibid., 190-193.

<sup>130</sup> Peter Singer, *Animal Liberation*, 2<sup>nd</sup> ed. (New York: Avon Books, 1990), 228. Also mentioned in Pluhar, 190-191.

<sup>131</sup> Ibid.

<sup>132</sup> Pluhar, 193-195.

destroy that life once it appears. A moral principle that allows this is absurd, so the prior-existence view must be rejected.

Singer's answer to the replacement argument is to retain the total view, but under his preference utilitarianism. Recall that preference utilitarianism measures utility in terms of satisfied preferences, rather than happiness or pleasure. Singer argues that complex beings such as humans have a preference for continued existence, and that this preference is thwarted when the individual is killed.<sup>133</sup> A thwarted preference is not compensated by the creation of new, initially unsatisfied preferences; an unsatisfied preference is understood as negative utility, which disappears when the preference is satisfied. (Whether we should construe a satisfied preference as 'resetting to zero' or as positive utility is a matter for debate. Singer prefers the latter option.)<sup>134</sup> In a replacement industry, individuals will consistently have preferences thwarted, so utility is not maximized.

Such is Singer's argument against replacing humans, and other animals capable of a preference for continued existence. However, individuals with no preference for continued existence gain no protection from this argument. Because conscious preferences are the issue, individuals with no *concept* of a continued existence cannot have a preference for continued existence. Therefore, individuals must be self-conscious, reflectively aware of themselves, in addition to being merely conscious, to have the preference for continued existence. We believe many animals lack self-consciousness. These animals, therefore, are replaceable. Singer mentions fish as an example:

We can presume that if fish became unconscious, then before the loss of consciousness they would have no expectations or desires for anything that might happen subsequently, and if they regain consciousness, they have no awareness of having previously existed. Therefore if the fish were killed while unconscious and replaced by a similar number of other fish who could be created only because the first group of fish were killed, there would, from the perspective of fishy awareness, be no difference between that and the same fish losing and regaining consciousness.<sup>135</sup>

---

<sup>133</sup> Singer, *Practical Ethics*, 2<sup>nd</sup> ed., 126-127.

<sup>134</sup> *Ibid.*, 128-129.

<sup>135</sup> *Ibid.*, 126.

Many will find Singer's distinction between replaceable and non-replaceable individuals preferable to the application of the prior-existence view, and to that of utilitarian theories that equate utility to pleasure or happiness.

There are still problems for Singer's view. First, the ASO demands that humans receive consideration equal to the consideration granted to relevantly similar animals. So Singer must admit that since some animals are replaceable, similar humans are replaceable as well. Some humans are not self-conscious. Newborn infants are certainly in this class, as are the most extreme cases of mental retardation and senility. Singer concludes that these humans are just as replaceable as the fish in his example.<sup>136</sup> This problem is not fatal to Singer's position, but many people find it highly objectionable. On the other hand, Pluhar argues that the class of self-conscious individuals may be larger than Singer realizes, and may include birds and other 'lower' animals.<sup>137</sup> If this is true, fewer humans are replaceable as well.

Second, Pluhar argues that preference utilitarianism still fails to explain why we cannot replace even a self-conscious individual.<sup>138</sup> If we kill one individual in order to replace it with another, we can counteract the disutility of the former individual's unsatisfied preferences by satisfying the preferences of the new individual. Now, in the case of self-conscious beings, the replacement individual cannot itself be replaced, because the preference for continued existence must be satisfied in order to counteract the previous disutility. Pluhar notes that this whole line of thought is strange, since death comes to all of us, and therefore the preference for continued existence is ultimately thwarted anyway.<sup>139</sup> Singer avoids this criticism by adding a level of sophistication to the idea of a preference for continued existence: he describes self-conscious lives as "arduous and uncertain journeys, at different stages, in which various amounts of hope and desire, as well as time and effort have been invested in order to reach particular goals or destinations."<sup>140</sup> The preference for continued existence is therefore not as morally significant at the end of one's life as it is in earlier stages. "Towards the end of life, when most things that might have been accomplished have either been done, or are now

---

<sup>136</sup> Peter Singer, "Killing Humans and Killing Animals," *Inquiry* 22 (1979): 153.

<sup>137</sup> Pluhar, 203-4.

<sup>138</sup> *Ibid.*, 208-11.

<sup>139</sup> *Ibid.*, 209.

<sup>140</sup> Singer, *Practical Ethics*, 2<sup>nd</sup> ed., 130.

unlikely to be accomplished, the loss of life may again be less of [a] tragedy than it would have been at an earlier stage of life.”<sup>141</sup> Pluhar notes that, at most, this strategy only prevents replacement of younger individuals.<sup>142</sup> Moreover, she argues, the description of self-conscious lives as uncertain journeys is a false generalization. Plenty of self-conscious beings have no complicated plans or long-term goals.<sup>143</sup>

Most importantly, Singer has failed to show why the creation of new individuals is not morally obligatory under the total view, regardless of how the notion of utility is construed. It is a mistake to fall back on the prior-existence view, and hold that the satisfaction of future preferences does not count in the moral calculus. Future generations, and their preferences, do matter. Likewise, it is morally wrong to bring the ‘wretched child’ into existence. The total view, then, must be accepted. Yet, are we not therefore obligated to bring a child into existence, under the condition that his or her preferences are satisfied? Or, to take things to their final conclusion, should we not create as many happy lives as possible? Singer admits that he has no satisfactory answer to this question.<sup>144</sup> “If the pleasure a possible child will experience is not a reason *for* bringing it into the world, why is the pain a possible child will experience a reason *against* bringing it into the world? A convincing explanation of this asymmetry has not, to my knowledge, been produced.”<sup>145</sup>

I believe that the replacement argument can be successfully addressed by the welfare-utilitarian. First, something should be noted about the moral ‘asymmetry’ Singer notes in the creation of the two possible children, one happy and one miserable. Singer presumes that the conditions of the children are symmetrical. The symmetry is actually an illusion. In the case of the miserable child, nothing can be done (aside from euthanasia) to relieve the child of misery. The child *cannot* be made happy, or have satisfied preferences. The happy child, on the other hand, is not necessarily happy. Parents generally succeed in raising children well enough that the children have lives worth living, but surely not always, and the child’s life is never entirely positive. The miserable child’s life, on the other hand, is always entirely bad. A situation truly

---

<sup>141</sup> Ibid.

<sup>142</sup> Pluhar, 210.

<sup>143</sup> Ibid., 211.

<sup>144</sup> Singer, *Practical Ethics*, 2<sup>nd</sup> ed., xi.

<sup>145</sup> Singer, “Killing Humans and Killing Animals,” 148. His italics.

symmetrical to that of the miserable child would require a child that *cannot* be made miserable. Now, if the couple's doctor predicted that a certain couple would produce such a perfectly happy child, then perhaps it is not so radical to suppose that conception would be obligatory! Singer's asymmetry, therefore, has been given at least a partial explanation.

The welfare-utilitarian does not have the same answer to the replacement argument as the preference-utilitarian. A salient difference is that the welfare-utilitarian will not place moral weight on the preference for continued existence. Even if an individual has no concept of continued existence, the individual's welfare depends upon continued existence. Welfare is determined solely by investigating what sort of being the individual is, not merely the individual's particular preferences. Therefore, no distinction between self-conscious beings and merely conscious beings can be highly relevant in regards to replaceability.

To consider an individual's welfare, it is not enough to determine how well an individual is doing at the present point in time. One's welfare is a measure of one's life as a whole. Living a full natural life, being free to develop one's abilities, and flourishing as the kind of thing one is all comprise an individual's welfare. For the welfare-utilitarian, then, replacement harms each individual's welfare in a way that is not counteracted by the addition of a new being. This is especially true when the replacee will face the same harm as well. This line of thought is similar to Singer's application of preference utilitarianism to self-conscious beings, but, as I noted, the welfare-utilitarian extends this consideration to all conscious beings, regardless of whether they have a conscious preference for continued existence.

We can follow Singer's view further, and hold that killing an individual nearer to death is not as significant a harm. Pluhar, you will recall, also holds this view. In old age, or in an otherwise near death condition, one's welfare is already diminished. There is simply less to lose. So perhaps replacement is permissible for near-death individuals. This is not a particularly dangerous view to hold. For one thing, recall that the replacement argument requires that replacement is only justifiable when the replacement individual's existence depends on the killing of the first individual. Replacement scenarios that truly conform to utilitarian demands are hard enough to come by. One

would certainly be hard-pressed to imagine a situation where a near-death individual must be killed, rather than allowed to die a short time later, in order to create a new life.

Finally, something must be said in response to the criticism that the total view obligates us to increase utility by bringing more beings (with lives worth living) into the world. This criticism applies to welfare utilitarianism, at least my version, which utilizes the total view. In the present conditions in many parts of the world, overpopulation has caused great suffering, and many people are calling for population control for that reason. The same can be said for many nonhuman species, an example being deer overpopulation in many parts of our country. We should therefore understand this criticism of utilitarianism as applying only to a world where overpopulation is not a problem. In such a world, is reproduction morally obligatory?

The answer is not simple. If we focus solely on the individuals who are produced, and a life worth living is guaranteed, then the answer is yes. The total view requires us to admit that much. And the admission is not counterintuitive: it makes perfect sense that a world with more happy individuals is a better place than one with fewer. If we imagine that there is another planet in the universe with a population of beings whose lives are fulfilling, and we ask if this is a better universe because that planet exists in addition to ours, few would say no.

However, there are other important considerations that affect the ultimate answer as it pertains to *this* planet. First, an individual's well-being is never guaranteed. In the right environment, we can consider it likely that the life will be one worth living, so the risk may not be substantial. But we must consider it, regardless.

Second, recall the asymmetry between the creation of a happy child and a miserable child: the miserable child cannot be made happy, but the happy child must be cared for in order to be happy. It is not the case that we can simply bring these happy individuals into existence. When a new being appears, one or more individuals receive the duty of looking after the individual's well-being. This duty may be taken on with pleasure, as it is with many parents, or it may be highly undesirable. When the duty to care for a new individual is not desired, we must take into account the disutility generated by the negative affect on the caregivers' welfares. How much disutility is generated will vary greatly, based on the impact the acquired duty has on each individual's well-being.

Moreover, if the duty becomes so much of a burden that the caregivers relinquish their roles, the child's well-being is endangered as well.

Finally, even if reproduction is morally obligatory, enforcing reproduction does not appropriately respect the kind of individuals people (and other animals) are. It would certainly bring a different tone to family existence if the family unit was created out of obligation. Our welfare, as I have mentioned, includes our free activity to pursue our goals, flourish as individuals, and create the kind of lives that are right for us. Enforced reproduction would surely detract from our welfare.

Problems with welfare utilitarianism's answer to the replacement argument may still exist. Why is the limited welfare of a near-death individual sufficient to permit replacement, yet it is impermissible to replace someone with slightly more time left? I do not know how one is to draw the line, but we should always err on the side of caution. Perhaps replacement should always be (legally) prohibited, despite the existence of exceptional cases where it is morally permissible. Such gray areas always exist for gradualist moral theories, but gray areas are a common enough feature when applying any theory that this particular instance is not terribly damaging to utilitarianism.

In contrast to classic utilitarianism and Singer's preference version, welfare utilitarianism provides an answer to the replacement argument that serves humans and animals equally, and in a manner that maintains the utilitarian's total view without suffering from the typical obstacles of the total view.

### **Welfare Utilitarianism and a Final Response to the ASO**

Near the beginning of this chapter I produced an initial utilitarian response to the ASO. The conclusion of the ASO was accepted: some humans and some animals are similarly morally valuable. Exactly how valuable different individuals are was left undetermined. One could conclude that animals are more valuable than our current attitudes suggest, or one could conclude that relevantly similar humans are less valuable. As a welfare-utilitarian, I believe the former is correct.

Humans are incredible animals, capable of mental activities that no other animal on Earth can perform. Humans create splendid works of art, vast cities, and even make valiant attempts at philosophical reasoning. These activities certainly contribute to our

welfare, in a way that other animals cannot achieve. At the same time, I do not believe that these high-order mental activities contribute so much to one's welfare that the welfares of 'lower' beings (non-rational humans included) pale in comparison. Taking an evolutionary viewpoint, the complexity of human primates is only a few steps away from our nearest surviving primate neighbors. Why should we think that the latest evolutionary step has produced an altogether new and unique form of well-being? An individual's well-being is too complex, made up of as many diverse elements as it is, to be compared to the welfare of any closely related individual, even across species borders. Distinctions are not impossible to make, so long as the individuals are substantially different and easily distinguished, in ways that are morally relevant. If Singer's description of 'fishy awareness' is correct, then the welfare of a fish is simple enough to be looked after by a mind that fails to persist over time. Simpler animals such as fish may in some sense be said to flourish over time, but not in the sense that a human, or even a dog, cat, or pig may flourish. Mental richness does count for something, just not as much as many moral theories propose.

In extreme cases, it is admittedly possible to compare the relative value of two humans. An obvious case is an anencephalic human, who is born with no brain whatsoever, only a terminating brain stem. Another case is a victim of an irreversible coma. These extreme cases clearly lack a welfare equivalent to a normal human. There is simply less of an individual present in those cases for us to consider, much less than in many animals. In the case of brain stem babies, there is no consciousness to speak of. Welfare utilitarianism provides no moral consideration whatsoever for such individuals. Neither, for that matter, do Regan's and Pluhar's rights theories, which similarly draw the line at consciousness.

Unlike Regan's answer to the lifeboat scenario, in which he is willing to sacrifice an infinite number of dogs (or non-rational humans) to save a single normal human, my view is closer to Pluhar's. A dog's life is not so expendable. While Pluhar holds that each individual always deserves equal consideration, the welfare-utilitarian's answer is: it depends on the individuals. Which humans are we talking about? Which dogs? Often, welfare comparisons will be highly difficult or impossible. Like Pluhar, I suggest that we give preference for individuals not near death. Other factors, such as the ability to

successfully bring the lifeboat to land, considerations about the individuals' moral characters (toss the villain overboard!), also make a difference. Because such comparisons will be difficult if not impossible to make, a group in such a terrible situation may very well prefer to draw lots, and cannot be blamed for doing so.

The foregoing points regarded our moral evaluation of individual lives. It is important to recognize that the gradualist scheme I am proposing does not produce gradualist answers in every situation. For example, pain is a disutility that can affect two very different individuals in a very similar manner. A chicken may be sufficiently simple that its life is less valuable than mine, yet pain for the chicken detracts from its welfare just as much as a similar pain detracts from mine. The moral value of one's life does not act as a 'multiplier' such that harming me matters more than harming a chicken to the same extent. In fact, some harms may matter more when they are inflicted on the chicken! I can get by, more or less, with a broken leg. A broken leg for a chicken can drastically reduce its welfare, perhaps fatally. Moreover, a human's higher awareness of the source of pain can often make pain more tolerable than it would be for an animal, who might experience extreme fear out of ignorance. Only when we consider the death of an individual, or other harm that detracts from most or all aspects of one's well-being, do we grant the more complicated individual additional consideration.

## **Conclusion**

Carruthers, Regan, and Pluhar all describe utilitarianism as an attractive theory, but find certain implications surrounding the ASO unacceptable. In response, each develops a theory of rights: Carruthers argues for a contractualist position, and Regan and Pluhar each argue for a more generic rights theory. Contractualism, while based on a firm theoretical foundation, generates no consideration for non-rational agents, despite Carruthers's effort to include all humans. Regan's theory has initial appeal, but lacks a solid foundation and, despite his reliance on intuitions, results in highly counterintuitive applications. Pluhar's theory appears most successful, having overcome Regan's foundational obstacles while serving much better than contractualism for explaining our attitudes towards animals and non-rational humans.

Each of these philosophers devotes significant attention to utilitarianism as the ‘runner-up’ theory, and they view Singer’s preference utilitarianism in particular as among the most successful formulations. By proposing the alternative view of welfare utilitarianism, I have attempted to overcome the criticisms that lead to the rejection of utilitarian theories. I have by no means described a complete theory, and much more needs to be said in regards to the nature of welfare, and how one can make decisions that affect it. What I hope to have done is indicate the way toward a moral theory that 1) provides an acceptable answer to the ASO, 2) rests on a solid theoretical foundation, and 3) overcomes the obstacles both of rights-based theories and of other utilitarian formulations.

## BIBLIOGRAPHY

- Carruthers, Peter. *The Animals Issue*. Cambridge, UK: Cambridge University Press, 1994.
- DeGrazia, David. *Taking Animals Seriously. Mental Life and Moral Status*. Cambridge, UK: Cambridge University Press, 1996.
- Edward Regis, Jr., ed. *Gewirth's Ethical Rationalism*. Chicago and London: University of Chicago Press, 1984.
- Frey, R.G. *Rights, Killing, and Suffering*. Oxford: Basil Blackwell, 1983.
- \_\_\_\_\_, ed. *Utility and Rights*. Minneapolis: University of Minnesota Press, 1984.
- \_\_\_\_\_ and Christopher W. Morris, eds. *Value, Welfare, and Morality*. Cambridge: Cambridge University Press, 1993.
- Gewirth, Alan. *Human Rights*. Chicago and London: University of Chicago Press, 1982.
- Goodin, Robert E. "Utility and the Good." In *A Companion to Ethics*, edited by Peter Singer, 241-248. Oxford and Malden, Mass: Blackwell Publishers Ltd., 1993.
- Miller, Harlan B. "Introduction: 'Platonists' and Aristotelians'." In *Ethics and Animals*, edited by Harlan B. Miller and William H. Williams, 1-14. Clifton, N.J.: Humana Press, 1983.
- \_\_\_\_\_. "A Terminological Proposal." *SSEA Newsletter* 30 (March 2002).
- Mill, John Stuart. *Utilitarianism*. Edited by Oskar Piest. New York: Macmillan Publishing Company, 1957.
- Nozick, Robert. *Anarchy, State, and Utopia*. New York: Basic Books, Inc., 1974.
- Parfit, Derek. *Reasons and Persons*. Oxford: Clarendon Press, 1984.
- Pluhar, Evelyn. *Beyond Prejudice. The Moral Significance of Human and Nonhuman Animals*. Durham and London: Duke University Press, 1995.
- Regan, Tom. *The Case For Animal Rights*. Berkeley and Los Angeles: University of California Press, 1983.
- Singer, Peter. *Animal Liberation*. 2<sup>nd</sup> ed. New York: Avon Books, 1990.
- \_\_\_\_\_. "Killing Humans and Killing Animals." *Inquiry* 22 (1979): 145-156.
- \_\_\_\_\_. *Practical Ethics*. 2<sup>nd</sup> ed. Cambridge, UK: Cambridge University Press, 1993.
- \_\_\_\_\_. "A Utilitarian Population Principle." In *Ethics and Population*, edited by Michael D. Bayles. Cambridge, Mass.: Schenkman Publishing Company, 1976.

## VITA

Born in New Jersey, Jesse Ehnert received his B.A. from Connecticut College in 1995, with a major in English and a minor in philosophy. After graduation, he worked at a variety of computer- and website-related jobs in North Carolina and Virginia, pursuing his philosophical interests in his free time. In 1997, Jesse happened upon a copy of Peter Singer's *Animal Liberation*, and made the startling discovery that his prior moral beliefs about nonhuman animals had been absolutely, undeniably incorrect. This revelation, combined with a lack of interest in computer- and website-related jobs, led him to the philosophy program at the Virginia Polytechnic Institute and State University, where he received his M.A. in philosophy in 2002. Jesse is currently seeking a career where he can make use of his philosophical education. Until then, he will likely continue to work at a variety of computer- and website-related jobs.