

RICE UNIVERSITY

**The Wrongness of Killing**

by

**Rainer Ebert**

A THESIS SUBMITTED  
IN PARTIAL FULFILLMENT OF THE  
REQUIREMENTS FOR THE DEGREE

**Doctor of Philosophy**

APPROVED, THESIS COMMITTEE



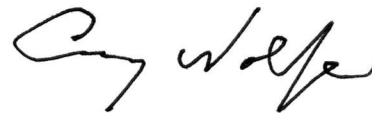
---

George Sher  
Herbert S. Autrey Professor of Philosophy



---

Baruch Brody  
Andrew W. Mellon Professor in  
Humanities & Professor of Philosophy



---

Cary Wolfe  
Bruce and Elizabeth Dunlevie Professor of  
English

HOUSTON, TEXAS  
May 2016

Copyright  
Rainer Ebert  
May 2016

Abstract

## **The Wrongness of Killing**

by **Rainer Ebert**

There are few moral convictions that enjoy the same intuitive plausibility and level of acceptance both within and across nations, cultures, and traditions as the conviction that, normally, it is morally wrong to kill people.

Attempts to provide a philosophical explanation of *why* that is so broadly fall into three groups: *Consequentialists* argue that killing is morally wrong, when it is wrong, because of the harm it inflicts on society in general, or the victim in particular, whereas *personhood* and *human dignity accounts* see the wrongness of killing people in its typically involving a failure to show due respect for the victim and his or her intrinsic moral worth.

I argue that none of these attempts to explain the wrongness of killing is successful. Consequentialism generates too many moral reasons to kill, cannot account for deeply felt and widely shared intuitions about the comparative wrongness of killing, and gives the wrong kind of explanation of the wrongness of killing. Personhood and human dignity accounts each draw a line that is arbitrary and entirely unremarkable in terms of empirical reality, and hence ill-suited to carry the moral weight of the difference in moral status between the individuals below and above it. Paying close attention to the different ways in which existing accounts fail to convince, I identify a number of conditions that any plausible account of the wrongness of killing must meet. I then go on to propose an account that does.

I suggest that the reason that typically makes killing normal human adults wrong *equally* applies to atypical human beings and a wide range of non-human animals, and hence challenge the idea that killing a non-human animal is normally easier to justify than killing a human being. This idea has persisted in Western philosophy from Aristotle to the present, and even progressive moral thinkers and animal advocates such as Peter Singer and Tom Regan are committed to it. I conclude by discussing some important practical implications of my account.

**To the memory of**

my grandmother, **Theresia Ebert**

(15 December 1928, Baranyajenő – 7 October 2008, Gaishardt),

my second uncle, **Stefan Blumenschein**

(3 August 1943, Baranyajenő – 17 July 2010, Ulm),

my grand-aunt, **Katharina Emerling**

(15 February 1925, Baranyajenő – 24 October 2011, Schrezheim),

my grandfather, **Robert Ohr**

(19 March 1928, Pfahlheim – 11 April 2013, Ellwangen), and

my grandmother, **Hedwig Ohr**

(17 January 1929, Neuler – 9 February 2015, Aalen),

who would have been proud. I was very lucky to have had such caring, loving, and kind-hearted people in my life, and I will always remember them fondly.

## Acknowledgements

My thanks go, first and foremost, to my dissertation advisor, George Sher, who has continuously provided essential criticism, advice, encouragement, and support from the first day I have met him at Rice University in 2010. He is an outstanding teacher, and I can only hope to someday live up to his example. I would also like to thank the other members of my dissertation committee, Baruch Brody and Cary Wolfe. When my motivation was down, Baruch urged me to persevere, and to trust him that I am where I belong. I did, and here I am. As a teacher, he was tough but always fair and helpful, and I think I am now a better philosopher for it.

I am grateful to the philosophy faculty and graduate students at Rice University, who made my time there a memorable and worthwhile experience, and from whom I have learned to read, think, write, and speak better. I wish to single out for special acknowledgement Richard Grandy, Jennifer Blumenthal-Barby, H. Tristram Engelhardt Jr., and Hanoch Sheinman, who have provided insightful comments on papers I have written, some of which have since been published, and many much-appreciated words of encouragement. I also want to thank Minranda Robinson-Davis. She keeps things running smoothly at the Department, patiently guided me through bureaucratic hurdles more than once, and always has a warm smile for everyone.

I presented some of the ideas in this dissertation at the *La libération animale, quarante ans plus tard* conference at the University of Rennes 2, in Rennes, and at Hindu College and the Department of Philosophy at the University of Delhi, and the third *Minding*

*Animals* conference at Jawaharlal Nehru University, in New Delhi. I would like to thank the audiences present on those occasions for their valuable questions and comments, which made this dissertation better than it would otherwise have been.

Thanks to Rice University for being my intellectual home for the past six years, and a friendly and stimulating one at that, and to the wonderful Adria Baker and her colleagues from the Office of International Students and Scholars, who have provided invaluable help with immigration issues.

I owe a special debt to Tom Regan, Tibor R. Machan, and Mylan Engel Jr., whose inspiring example led me to deepen my appreciation of the practical significance of philosophical inquiry. If I had not met them, I might have never applied to graduate school in philosophy. Peter Singer – whose book, *Animal Liberation*, was one of the first philosophy books I have read in my life – also had a significant impact on my decision to pursue the life of a philosopher, and so did Peter McLaughlin and Miriam Wildenauer, whom I was very lucky to have as my philosophical mentors as an undergraduate student at Heidelberg University in Germany.

I wish to thank my wife, Maqsuda Afroz, for her love and support, and for countless valuable discussions of matters of philosophy, and pretty much everything else. I could not have done this without her. The same is true for my parents, Anton and Margit Ebert, who have always believed in me and supported me in whatever I wanted to do. I am privileged to have a fantastic family and great friends. I cannot name them all here, but I do want to mention Ankit Sethi, who never got tired of answering my questions about the English language.

Trips to Africa, the Middle East, and South Asia, particularly Bangladesh, during summer and winter breaks provided much-needed perspective, and adventure. I thank my many friends in those beautiful places for their hospitality and friendship.

Finally, I gratefully acknowledge the generous financial support from the Department of Philosophy at Rice University during the time I spent in its doctoral program, and I thank the Humanities Research Center and the Graduate Student Association at Rice University for travel grants that enabled me to attend conferences in France and India, from which this dissertation benefitted.



## Table of contents

Acknowledgements.....	vi
Table of contents.....	ix
Introduction.....	1
Chapter 1: Killing as wrong because of what it does.....	8
1.1. Classical utilitarianism regards as morally right what is wrong .....	11
1.2. Classical utilitarianism arrives at right conclusions for the wrong reasons .....	19
1.3. Classical utilitarianism gets the comparative wrongness of killing wrong .....	29
1.4. Rethinking well-being: non-hedonistic act-utilitarianism.....	39
1.5. Good – but not good for anyone: non-utilitarian maximizing act- consequentialism.....	48
1.6. The pro tanto reason to promote the good.....	52
1.7. Recasting the relationship between the normative status of acts and value.....	55
1.8. Harm-to-the-victim accounts of the wrongness of killing .....	60
1.9. Rule-consequentialism .....	68
Chapter 2: Killing as wrong because of what it is .....	78
2.1. McMahan’s two-tiered account of the wrongness of killing.....	83
2.2. The first tier: the time-relative interest account of the wrongness of killing .....	84

2.3.	The distinction between persons and non-persons, and its moral significance..	96
2.4.	The second tier: the intrinsic worth account of the wrongness of killing .....	108
Chapter 3: Are human beings more equal than other animals? .....		117
3.1.	Two strategies to justify equal human worth .....	120
3.2.	Strategy One: the argument from substantial identity.....	124
3.3.	The genetic basis for moral agency account of rightholding .....	138
3.4.	Strategy Two: the nature-of-a-kind argument.....	153
Chapter 4: Being a world unto one's self: a new perspective on the wrongness of killing, and some of its implications .....		164
4.1.	What we have argued thus far, and an outline of what lies ahead .....	167
4.2.	Being a somebody as a matter of phenomenal consciousness .....	171
4.3.	What sort of things are we essentially? .....	174
4.4.	The binary nature and moral significance of the capacity for phenomenal consciousness .....	193
4.5.	The dignity of subjectivity account of the wrongness of killing.....	201
4.6.	Some implications .....	203
4.7.	Epilogue .....	214
Bibliography .....		217
Appendix: Tom Regan's philosophy of animal rights.....		238

## Introduction

Consensus is rare in matters of philosophy. This is particularly true for ethics. There is widespread and profound disagreement over how we ought to live our lives, both within and across nations, cultures, and traditions. Yet, there are a number of moral values, beliefs, and principles that are almost universally held. Among them is the conviction that, normally, it is seriously morally wrong to kill people like you and me. There is, of course, divergence of opinion about what constitutes normal circumstances. Some people believe that the qualifier “normally” is altogether unnecessary, as it can never be morally justified to kill a person, regardless of the circumstances. Others believe that, in some circumstances, killing a person can be justified, for example, if the one being killed is a serious threat, guilty of a heinous crime, or suffering to an extent that makes his or her life not worth living, or if it is necessary to prevent some great evil. Virtually everybody agrees, however, that it is a grave moral wrong to kill an innocent adult human being who is leading a fulfilled life and is a threat to nobody, for no other purpose than an insignificant material gain, or for some other trivial reason. It is therefore no coincidence that all major religious scriptures contain prohibitions against killing. The Book of Exodus, which is part of both the Torah and the Hebrew Bible, asserts, “Thou shalt not kill.”<sup>1</sup> The Qur’an, in the Medinan surah, al-Ma’ida, states that “whosoever killeth a human being for other than manslaughter or corruption in the earth, it shall be as if he had killed all mankind, and whoso saveth the life of one, it shall be as if he had saved the life

---

<sup>1</sup> Exodus 20:13 (King James Version).

of all mankind.”<sup>2</sup> The Manusmṛti, one of the central texts of Hinduism, echoes this sentiment: “He who commits murder must be considered as the worst offender, more wicked than a defamer, than a thief, and than he who injures with a staff.”<sup>3</sup> Religion here reflects a wide agreement that also finds expression in the law. The criminal laws of all countries, without exception, prohibit the killing of human beings in all but very few narrowly defined circumstances, and Article 3 of the 1948 Universal Declaration of Human Rights guarantees the right to life to everybody. In view of this near-universal acceptance that there is a strong moral prohibition against killing, it is perplexing and in a way unsettling that there is no satisfactory explanation of *why* that is so, a state that, not long ago, has been described by the contemporary American philosopher Fred Feldman as “one of the most notorious scandals of moral philosophy.”<sup>4</sup>

Philosophical attempts to explain the wrongness of killing broadly fall into three groups: *Consequentialists* argue that killing is morally wrong, when it is wrong, because of the harm it inflicts on society in general, or the victim in particular, whereas *personhood accounts* see the wrongness of killing people in its typically involving a failure to show due respect for the victim and his or her intrinsic moral worth as a person. “Person” here is meant in a Lockean sense that is not coextensive with humanity, and both families of views hence make no categorical distinction between the killing of human beings and the killing of other animals. In contrast, the third group of theorists defend, in decidedly secular terms, the radical dichotomy between human beings and other animals that is deeply rooted in Abrahamic religiosity and general Western thought alike, by appealing to the concept *human dignity*. I will show that none of these attempts to explain the

---

<sup>2</sup> Qur’an 5:32 (Pickthall).

<sup>3</sup> Manusmṛti (Laws of Manu) 8.345.

<sup>4</sup> Feldman 1992, p. 157.

wrongness of killing is successful. By paying close attention to the different ways in which they fail to convince, I will identify a number of conditions that any plausible account of the wrongness of killing must meet. Finally, I will propose an account that does, and discuss some of its implications.

Even though almost everybody already accepts that, normally, killing people is seriously wrong, this is a worthwhile project for a variety of reasons, of which I will just mention two. First, there is, of course, an academic interest in finding a plausible justification for the belief that killing is normally wrong. But second, and more importantly, there is an urgent practical need to make progress in contemporary debates about the killing of individuals other than normal, adult human beings, such as fetuses, infants, the profoundly mentally retarded, the permanently suffering and terminally ill, and non-human animals. Despite my narrow focus on the question of why killing is *normally* wrong, my answer to that question will allow us to draw surprising and far-reaching conclusions that are directly relevant to some of these debates. That is, because I will argue that the fundamental reason that typically makes killing normal adults morally wrong *equally* applies to later-stage fetuses, infants, atypical adults, and a wide range of non-human animals. I will hence deny, for example, that there is a categorical moral difference between infanticide and killing innocent adults, which is also the position of proponents of human dignity.<sup>5</sup> However, extending the scope of the egalitarian aspiration at the core of popular theories of human dignity, the account I envision challenges – and

---

<sup>5</sup> This position is not uncontroversial. In a widely-read 2012 article in the *Journal of Medical Ethics*, Alberto Giubilini and Francesca Minerva argued that infanticide – the authors call it “after-birth abortion” – should be permissible in all cases in which abortion is (Giubilini & Minerva 2012). The article caused a media storm. Yet, the view that killing a new-born is a lesser moral wrong than killing an adult is hardly new. For decades, philosophers such as Michael Tooley, Peter Singer, and Jeff McMahan have been arguing for that view, cf., e.g., Tooley 1972 & 1983, Singer 1979a, and McMahan 2002.

this is probably the most interesting and controversial aspect of this dissertation – the view that killing a non-human animal is normally easier to justify than killing a human being. This view, which has persisted in Western philosophy from Aristotle to the present, and to which even progressive moral thinkers and animal advocates are committed, can be understood as a manifestation of the more general view that has been dubbed *perfectionism* in an exciting 2008 book by Paola Cavalieri, *The Death of the Animal*:

“[P]erfectionists hold that there is a hierarchy in moral status. They maintain that conscious beings, and their interests, deserve different consideration according to their level of possession of certain characteristics.”<sup>6</sup>

The first part of Cavalieri’s book is a dialogue between two fictional characters, one of whom is Alexandra Warnock, who notes that nowadays almost everybody rejects the idea that there is a hierarchy in moral status among human beings. Cognitive ability and other characteristics that have traditionally been associated with our special moral status vary widely within humanity, yet we do not think that these differences undermine human equality. Alexandra then argues that it is therefore inconsistent – or, to put it more neutrally, theoretically problematic – to appeal to these characteristics in order to deny equal moral standing to other animals. Perfectionism, she concludes, is a remnant of an obsolete moral worldview: “a ‘living fossil’ in the field of ideas.”<sup>7</sup>

Alexandra has a soulmate in another fictional character, J. M. Coetzee’s Elizabeth Costello, who too has serious doubts about the all-too-commonly-assumed moral significance of cognition. In a lecture about animals, Elizabeth tells her audience that she

---

<sup>6</sup> Cavalieri 2008, p. 3.

<sup>7</sup> Cavalieri 2008, p. 22.

has always been uncomfortable with René Descartes' famous formula, "*Cogito, ergo sum.*"

"It implies that a living being that does not do what we call thinking is somehow second-class. To thinking, cogitation, I oppose fullness, embodiedness, the sensation of being – not a consciousness of yourself as a kind of ghostly reasoning machine thinking thoughts, but on the contrary the sensation – a heavily affective sensation – of being a body with limbs that have extension in space, of being alive to the world. This fullness contrasts starkly with Descartes' key state, which has an empty feel to it: the feel of a pea rattling around in a shell."<sup>8</sup>

Elizabeth fervently believes that the cognitive elitism of perfectionism ought to give way to a radical egalitarianism of some kind, and one of the things I hope to accomplish in this dissertation is to provide a rigorous philosophical defense of the need for that shift in how we see and position ourselves in relation to other conscious beings.

Alexandra's conclusion, that the theoretical underpinnings of perfectionism are deeply problematic, seems right in cases where perfectionism takes the form of drawing a line between the community of moral equals, and those who are said to possess various degrees of a lesser moral status, and hence are excluded from that community. Personhood and human dignity accounts of the wrongness of killing fall into that category. I will argue that our modern scientific understanding of life on earth, according to which all life is interrelated, through evolution, and biological characteristics come in degrees, does not fit well with a bifurcated ethics that radically marks out a certain class of conscious beings. That kind of ethics, I think, is in fact "an atavism that a sound ethics

---

<sup>8</sup> Coetzee 2004, p. 78.

can no longer accept [...],”<sup>9</sup> and insofar the vast majority of people are still holding on to it, the Darwinian revolution is not yet complete. But there are other forms of perfectionism that avoid the theoretical shortcomings of traditionalist ethics. The view, mentioned by Peter Singer in the Foreword to Cavalieri’s book, that there are “graduations of moral status for humans”<sup>10</sup> falls into that category, and so does – at least indirectly – consequentialism.<sup>11</sup> We will come back to the former alternative in a later chapter, but we will start our discussion with the latter, which is by far the more popular one.

Before we begin our argument, a brief note on my use of moral intuition is in order. I have the belief, which I cannot here defend, that we may rely on intuitions as evidence in ethical justification. The method I will use throughout this dissertation is a coherentism of the kind endorsed by John Rawls, which seeks to establish *reflective equilibrium*.<sup>12</sup>

“[W]e begin with a set of moral intuitions about particular cases, filter out those that are the obvious products of distorting influences, and then seek to unify the remaining intuitions under a set of more general principles. We seek principles that both imply and explain our particular judgments. But the match between principles and intuitions will inevitably be very imperfect in the first instance. A candidate principle may imply a great many of our intuitions and yet have some

---

<sup>9</sup> Cavalieri 2008, p. 33.

<sup>10</sup> Cavalieri 2008, p. x.

<sup>11</sup> From the definition given above, it might seem as if consequentialism does not qualify as a perfectionist theory insofar consequentialists typically give equal consideration to interests that are comparable, regardless of the mental characteristics of the interest-bearers. The consequence of this, however, often is that conscious beings who are thought to lose less by dying due to their less sophisticated mental lives are said to have a less stringent claim not to be killed. In that sense, such conscious beings are given less consideration because they possess certain characteristics to a lesser degree. Alexandra calls such attempts to devalue the lives of conscious beings who are perceived to have a mental life less rich than the ones you and I have “subjective defense[s] of perfectionism” (Cavalieri 2008, p. 30).

<sup>12</sup> Cf. Rawls 1972, pp. 19-21 & 48-51.



implications that conflict with other intuitions. In that case we may modify or even abandon the principle; but, if the principle has considerable explanatory power with respect to a wide range of intuitions and cannot be modified without significant sacrifice of this power, we may instead decide to reject the recalcitrant intuitions. In this way we make reciprocal adjustments between intuitions and principles until our beliefs at various levels of generality are all brought into a state of harmony, or reflective equilibrium.”<sup>13</sup>

As far as I can tell, this method is the norm among contemporary analytic philosophers. Unlike some of them, however, I will restrict my use of moral intuitions to my own intuitions, and not rely on the actual or supposed intuitions of “many” or “most” people, as the mere number of people who share some particular moral intuition does not seem to say much about its credibility. Claims about where most people stand on matters of morality also sometimes raise the suspicion of being grounded less in reality than in the argumentative goal of the arguer, as reliable empirical data on moral intuitions is sparse, and insofar the value of such claims anyway is at best unclear it seems more honest and transparent to abstain from making them altogether. I further hope – maybe without basis, which I shall leave up to the reader to decide – that my professional training in philosophy has helped me to become less susceptible to superstition, bias, and historical prejudice than people in general. For these reasons, I will exclusively rely on the critical and careful use of my own moral intuitions, and appeal to what I think is the common moral consciousness only when I intuitively agree with it. This way I will, at best, convince or even inspire the reader, or, at worst, merely offer the reader a new perspective – both seems worthwhile.

---

<sup>13</sup> McMahan 2013a, p. 110.

## Chapter 1: Killing as wrong because of what it does

“Because of the consequences.” That, schematically, is the answer to the question of why killing is wrong, when it is wrong, that is common to the views we will discuss in this chapter.

These views are divided into two categories: consequentialist views, and views that are not strictly consequentialist but nevertheless regard the consequences of wrongful killings as their primary wrong-making feature. I will start with the former, and then show that many objections against it also apply to the latter.

*Consequentialism* is not a single ethical theory but a large class of such theories. As I shall understand the term “consequentialism,” all of these theories are variations on a theme that involves at least the following two claims:

- *Supervenience*: The normative status of acts is *fully* determined by the value of relevant consequences.
- *Agent-neutrality/impartiality*: The goodness or badness of consequences, and states of affairs in general, is to be evaluated by means of some objective standard of intrinsic value; the goodness or badness of consequences does not depend on the personal perspective from which they are evaluated, and there is no essential reference to individuals in the description of valuable states of affairs.<sup>14</sup>

---

<sup>14</sup> For example, a theory of value that implies that my aunt’s joy at finding out she has won the lottery has more value when evaluated from my point of view than it has when evaluated from a stranger’s point of view cannot be the basis of a consequentialist ethical theory, given my use of the term “consequentialism.” Neither can a theory of value that attaches more value to pleasure experienced by Victor than to pleasure experienced by Susan, not because it is more intense or of a longer duration or higher quality, but merely because it is Victor’s.

Consequentialist ethical theories differ from each other in a) which consequences they regard as relevant; b) how they spell out the relationship between the rightness and wrongness of acts and the value of relevant consequences; and c) their theory of value.

Historically, the most influential consequentialist theory has been maximizing hedonistic act-consequentialism, which – following a common practice – I will refer to as *classical utilitarianism*.<sup>15</sup> Classical utilitarianism is the combination of maximizing act-consequentialism and value-hedonism. The former is the view that the consequences relevant to the normative status of an act are those of that particular act and its alternatives, in that, in a given choice situation, an act is morally right, if no alternative act available to the agent would bring about more good. Otherwise, the act is morally wrong.<sup>16</sup> The latter, value-hedonism, is the standard of good by which consequences of acts are to be evaluated. According to value-hedonism, only pleasurable experiences, and the absence of painful or unpleasant experiences, have intrinsic value (are good in themselves), and only painful or unpleasant experiences, and the absence of pleasurable experience, have intrinsic disvalue (are bad in themselves).<sup>17</sup> Hence, in a given choice situation, the consequences of acts open to the agent are to be ranked according to the net

---

<sup>15</sup> Classical utilitarianism is commonly associated with Jeremy Bentham, John Stuart Mill, and Henry Sidgwick, even though it is doubtful that any of these authors actually held that view, cf. Bentham 1789a, Mill 1861, and Sidgwick 1874.

<sup>16</sup> I take the term “morally right” to mean the same as the term “morally permissible,” and the term “morally wrong” to mean the same as the term “morally impermissible.” There might be two or more available acts for which it is true that no alternative act would bring about more good. If so, the agent has a moral obligation to perform one of these acts (as the alternative acts are wrong). If only one optimistic act is available to an agent in a given choice situation, that act is not only right but also obligatory.

<sup>17</sup> As this theory of value is meant to be used to determine the relative value of consequences, the classic utilitarian is committed to the assumption that there is a hedonic measure that allows us to assign numbers to pleasures and pains across individuals and times. The overall value of a state of affairs can then be computed by adding up the numbers assigned to the instances of pleasure and pain it contains. I will not discuss the plausibility of this assumption. I will further abstain from giving precise definitions of the terms “pleasure” and “pain,” as my argument does not depend on such definitions. It should be noted, however, that a sensible classical utilitarian will not restrict them to their narrow, physical or sensory meaning. Emotional pain, anxiety and other mental states that are not liked surely also matter morally, and not only physical or sensory pain. The same holds for pleasure.

pleasure they would bring about, and only the one or few acts that score highest are morally right.<sup>18</sup> All other acts are morally wrong. This fundamental moral principle of classical utilitarianism – again following a common practice – I will call the *principle of utility*.

Classical utilitarianism has been the object of much criticism. It has been argued against classical utilitarianism that it is based on a superficial and impoverished view of the good, that it demands both too much and not enough, that it fails to take persons, and their ontological separateness, seriously, and that it cannot accommodate important considerations about justice and fairness.<sup>19</sup> Such criticism has led Bernard Williams to suggest that “[t]he day cannot be too far off in which we hear no more of it.”<sup>20</sup> And yet, classical utilitarianism is still around and does not seem to be going away any time soon. One reason for this, I believe, is the fact that classical utilitarianism incorporates many features that, at least on the face of it, have great theoretical appeal:

- *Simplicity and scope*: Classical utilitarianism promises to explain the normative status of acts by a single fundamental principle, and straightforwardly applies to all possible acts in all possible situations.
- *Coherence*: Because classical utilitarianism recognizes only one fundamental principle, there is no room for conflict between competing principles. The consequences of an act are either optimific, or not. Hence, after the required

---

<sup>18</sup> My definition of classical utilitarianism is deliberately vague. For example, I do not specify whether what matters are the foreseen, reasonably expected, intended, likely, or actual consequences of acts. I think there are good reasons for the utilitarian to formulate his theory in terms of reasonably expected consequences, somehow weighted in accordance with their degree of probability, but not much depends on this choice in the context of our discussion.

<sup>19</sup> “Person” here is used in a technical sense. I will discuss the difficulties of finding a philosophically sound theory of personhood later. For the moment, it is sufficient to note that, according to every theory of personhood I am aware of, normal adult human beings are persons.

<sup>20</sup> Williams 1973, p. 150.

utilitarian calculation is done, every act will come out as morally either right or wrong.

- *Rigorous impartiality*: Classical utilitarianism counts all pleasures and pains without distinction of the race, gender, nationality, age, species, etc. of the ones who experience them.<sup>21</sup>

There are few moral theorists who do not wish to find a theory that has these features. In particular the promise that all of morality can be traced to a single fundamental principle – a promise that only few ethical theories can make – has tremendous appeal. Another reason why classical utilitarianism continues to be discussed is the reasonable expectation that such discussion can help us better understand and evaluate classical utilitarianism’s more sophisticated descendants and cousins.

Classical utilitarianism deserves to be considered in its own right, despite its well-known shortcomings. I will do that, and show that many of the features that make it problematic are present in other consequentialist ethical theories as well. My goal in this chapter is not to “disprove” classical utilitarianism, or any of the related theories that I will discuss, but to explain why I think that, even though they have many virtues, they are ultimately not plausible and ought to be rejected, particularly as an account of the wrongness of killing.

### ***1.1. Classical utilitarianism regards as morally right what is wrong***

According to classical utilitarianism, an act is morally right, if, and only if, no alternative act results in a greater balance of pleasure over pain. The implication for the morality of

---

<sup>21</sup> Also, it can seem almost trivial that the utilitarian response to value is appropriate. How can it ever be wrong to make the world as good as we can?

killing is straightforward: Whenever an act of killing is morally wrong, it is wrong because there was another act available to the agent at the time of the act that would have brought about a greater balance of pleasure over pain, and the act of killing hence failed to maximize net pleasure. I think it is safe to assume that most people never find themselves in a choice situation in which no other available act would be more conducive to the good than killing another person. Normally, killing another person has devastating consequences. The classical utilitarian can point to the pain and suffering killing victims often go through while dying; the victim's terror of knowing that she is about to die as the attacker prepares to kill her; the utility that is foreclosed by the premature death of the victim, i.e. the excess of pleasure over pain the victim would likely have experienced in the future, had she not been killed; the remorse the killer will experience later on in his life, if he has any conscience at all, as well as the punishment he might receive; the loss and grief inflicted upon the victim's dependents and loved ones; and the fear and feeling of insecurity caused to third parties by the killing. And yet, sometimes, these common adverse consequences do not arise or can be neutralized, and killing in fact seems optimistic. In such cases, not only is there no utilitarian objection to killing, classical utilitarianism may morally require that it be done. This clashes with the common conviction that killing can only be justified under narrowly defined special circumstances, say, in a just war, or when necessary to defend one's own life or the life of another person. Classical utilitarianism is only concerned with the aggregate balance of pleasure over pain and not at all with the way in which it is produced or who stands to benefit or be harmed, and optimistic killings will hence not coincide with killings we

usually regard as justified. It would be astonishing if, e.g., all killings done in self-defense were optimific, but no killings for financial gain.

Many cases of optimific killings have been described and discussed in the philosophical literature that appear to lie outside the boundary of what is commonly considered justifiable homicide: the doctor who can save five patients only by killing a healthy person and harvesting her organs; the bystander who sees a railway trolley rushing toward ten people tied to the tracks and can only save them by pushing a fat man in front of the trolley; etc.<sup>22</sup> Opponents of classical utilitarianism argue that consideration of these cases shows that, in contradiction to what classical utilitarianism stands for, the ends do not always justify the means. The general structure of the argument is the following: Classical utilitarianism implies that a certain act is morally required; that act is monstrous and no sound ethical theory would allow and even less require it; therefore, classical utilitarianism must be false.

Does classical utilitarianism really require agents to perform acts that are clearly morally wrong, despite being optimific? Consider the following scenario, first described by H. J. McCloskey, that is exemplary for the general argument.<sup>23</sup>

*Innocent Scapegoat.* An outrageous crime has been committed, and the public demands that the perpetrator be caught and executed. “[I]n order to stop a series of lynchings which he knew would occur if the guilty person were not immediately found, or believed to have been found [...],”<sup>24</sup> the sheriff frames an innocent scapegoat who is then executed. If there was no other way for him to

---

<sup>22</sup> The first case was introduced by H. J. McCloskey in McCloskey 1957, and made famous by Philippa Foot in Foot 1967. The second case is a variation by Judith J. Thomson (cf. Thomson 1985) of a similar case first introduced by Foot in Foot 1967.

<sup>23</sup> Cf. McCloskey 1957, pp. 468 f.

<sup>24</sup> McCloskey 1965, p. 255.

stop the series of lynchings, it would seem like classical utilitarianism must endorse the view that the sheriff did what he was morally required to do.

The objection is that classical utilitarianism here sanctions what is morally unacceptable. A sheriff, of all people, should not get an innocent person executed, just to maximize the balance of pleasure over pain in the world. The claim that the sheriff is morally required to frame an innocent scapegoat is outrageous, and any ethical theory that implies that claim is surely false, or so the argument goes.<sup>25</sup>

In a direct response to McCloskey, Timothy L. S. Sprigge has argued that, initial appearances to the contrary notwithstanding, classical utilitarianism agrees with the common moral consciousness in all or almost all actual cases of the kind described in *Innocent Scapegoat*.<sup>26</sup> The prominent utilitarians J. J. C. Smart and R. M. Hare later repeated essentially the same point.<sup>27</sup> In reality, framing an innocent person is not likely to have the best results. Surely, there is a significant probability of the public finding out that the person who was executed was in fact innocent. If that happened, the public would lose trust in the sheriff, and law and order in general would deteriorate, the consequences of which would be much worse. Also, how does the sheriff know that the execution of a person who the public believes to be the perpetrator and only such an execution will stop the lynchings? How does he know that the public turmoil is not going to die down soon anyway? How can he be certain that there is no other way to prevent a series of lynchings? For example, why is it that he cannot just pretend to have the perpetrator

---

<sup>25</sup> This objection is closely connected with the so-called demandingness objection against classical utilitarianism (cf., e.g., Scheffler 1982). Suppose there is only one person, Ahmed, who could be credibly framed. Ahmed knows that he is innocent, but he also knows that many other innocent people will be killed by vigilantes unless he allows himself to be executed. It seems like classical utilitarianism demands that Ahmed assists the sheriff in framing him. Some might regard such an incredibly brave and selfless sacrifice as admirable, but it seems wildly implausible to suppose that it is a moral requirement.

<sup>26</sup> Cf. Sprigge 1965, pp. 272 ff.

<sup>27</sup> Cf. Smart 1973, pp. 67 ff., and Hare 1981, pp. 130 ff.



executed? In an actual situation, it is hard to imagine that all of these assumptions would amount to more than a hunch. Having robust knowledge in these matters seems practically impossible. In contrast, the sheriff knows for certain that the execution of an innocent person would cause significant harm to that person and his friends and relatives. A reasonable utilitarian will attach more weight to the bad consequences that are certain than to the uncertain hunch that an execution will stop the lynchings and none of the associated risks will be realized, and find an alternative act with a higher expected utility. Acts that seem optimistic at first glance often turn out not to be optimistic, particularly in unusual situations. But, in actual decision-making situations, there usually is only enough time to have a brief glance at the options. The sheriff has little time to think about risks and potentially disastrous side-effects. The human tendency, pointed out by Hare, to “‘cook’ our moral thinking to suit our own interest”<sup>28</sup> and our weakness in the face of temptation and pressure further limit the possibility of performing a correct utilitarian calculation. We are not what Hare calls “archangels,” beings “with superhuman powers of thought, superhuman knowledge and no human weakness.”<sup>29</sup> Both Smart and Hare, and Jeremy Bentham and Henry Sidgwick too for that matter, hence caution against using the principle of utility as a decision procedure.<sup>30</sup> Most people in most situations are more likely to act rightly if they follow their moral consciousness than if they try to perform utilitarian calculations. For the sheriff, there is hence a strong utilitarian reason to not allow himself to even think of the option of framing an innocent person as a genuine

---

<sup>28</sup> Hare 1981, p. 38.

<sup>29</sup> Hare 1981, p. 44.

<sup>30</sup> Cf. Sidgwick 1874, Smart 1967, Smart 1973, pp. 42 ff., and Hare 1981, pp. 44 ff. Jeremy Bentham wrote that “[i]t is not to be expected that this process [of calculating the expected utility of all available acts] should be strictly pursued previously to every moral judgement, or to every legislative or judicial operation” (Bentham 1789a, p. 311 n. 1).

option; too high is the risk of people finding out that the sheriff's deliberations are guided by utilitarianism rather than the law and hence losing their confidence in criminal justice, and of utilitarian thinking resulting in a disaster rather than the best possible outcome.

How much does this strategy of denial do for the classical utilitarian? Even if we are convinced that it is very unlikely that any sheriff will ever find himself in a situation in which he knows with a sufficiently high probability that framing an innocent scapegoat will prevent a series of lynchings, that the innocence of the scapegoat will never be discovered, and that the side effects will not be bad enough to tilt the utilitarian balance against framing an innocent person, it is still *possible* – at least logically – for a sheriff to be in that kind of situation.<sup>31</sup> That remains true even if we assume that, as Hare sometimes seems to suggest, such a situation never has occurred, and never will occur.<sup>32</sup>

The classical utilitarian must say that, *if* framing an innocent scapegoat were to bring about the best consequences, and *if* the sheriff knew that for sure, he must do it. I think this is enough to discredit classical utilitarianism. Proponents of the view however need not agree.

Smart too is “not *happy* about this consequence of utilitarianism.”<sup>33</sup> But then, as Sprigge observes, it is quite understandable that a classical utilitarian, or any person of conscience for that matter, would not be happy with the idea of getting an innocent person killed: “[A] man who was not sad at producing suffering would lack the basic sentiment which inspires the utility principle, namely a revulsion at the suffering and a delight in the

---

<sup>31</sup> We might not be convinced and still believe that this is much more than a mere logical possibility. Framing and executing an innocent person might well turn out to have the highest expected utility among all available options in a realistic choice situation, even taking into account our epistemological limitations. However, for the sake of argument, we shall refrain from pushing this point.

<sup>32</sup> Cf. Hare 1981, pp. 133 & 164.

<sup>33</sup> Smart 1973, p. 71. Also, cf. Hare 1981, p. 164.

happiness of any sentient being.”<sup>34</sup> We are asked to imagine a situation in which a sheriff is forced to choose between two evils. Either one innocent person dies, or many innocent people die. Both are results that one cannot be happy with, as both involve a great deal of misery. The classical utilitarian recommends the lesser of the two evils, and sadly accepts the sacrifice of the scapegoat, even if that is intuitively revolting. After all, a situation in which the sheriff knows that framing an innocent person will bring about the best consequences is so wholly different from situations we ordinarily encounter, so outlandish, that we are well-advised to be skeptical about our intuitive reaction to such a situation.<sup>35</sup> Our intuitions have evolved to fit everyday situations and are informed by our knowledge of risks normally associated with acts of injustice and the general unpredictability of the behavior of other people, particularly large groups as in the case of the rioting mob. Can we be sure that our intuitive objection to the idea that it can be right to execute an innocent scapegoat does not just stem from the fact that we cannot fully appreciate, not emotionally anyway, the details of the fantastic situation we are asked to imagine, particularly the absence of risk and uncertainty which are such crucial and always-present aspects of our everyday experience? The classical utilitarian says we cannot, and urges that McCloskey’s example does not provide us with a sufficient reason to reject her view. Rather, she argues, it points to a case – a fantastic case – in which we must revise our intuition. Not only is it, as Smart writes, “quite conceivable that there is *no* possible ethical theory which will be comfortable with all our attitudes,”<sup>36</sup> we should

---

<sup>34</sup> Sprigge 1965, p. 280.

<sup>35</sup> The kind of knowledge we are considering here is quite extraordinary. Philip E. Devine pointedly notes that, if we always had such superior knowledge, “our moral life would be very different from what it now is. [...] The moral rules governing the conduct of an omniscient being are not necessarily those which should govern the conduct of a human being” (Devine 1978, p. 150).

<sup>36</sup> Smart 1973, p. 72.

expect any sensible ethical theory to challenge some of our attitudes. One of the main purposes of engaging in moral theorizing is to critically reflect on ordinary morality and advocate for moral reform if inconsistencies are discovered, and surely we should not be surprised to find that ordinary morality is not fully consistent.

Now, in response, it must be noted that the principle of utility too is fair game for criticism, and it might well be argued that the classical utilitarian accepts it because of its prima facie intuitive appeal, and does so uncritically, even though examples such as McCloskey's raise significant doubt about whether that acceptance is warranted on critical examination. The classical utilitarian must make sure that her starting point, the principle of utility, can meet the high standard of proof she herself sets. If she is ready to discard our intuitive reaction to fantastic cases in which her view clashes with our sense of justice, to say with Smart, "so much the worse for the common moral consciousness,"<sup>37</sup> then she better is also ready to explain why those who do not share her commitment to classical utilitarianism do not have equal reason to say, "so much the worse for the principle of utility." I for one think that the intuitions behind McCloskey's objection to classical utilitarianism are at least as strong as those behind the principle of utility.

Utilitarians have attempted to meet that challenge. Hare and Mill, for example, have offered elaborate proofs of utilitarianism. A detailed and adequate discussion of these proofs however is beyond the scope of this dissertation – in fact, doing justice to these proofs would likely require its own dissertation – and we have to content ourselves with

---

<sup>37</sup> Smart 1973, p. 68.

the observation that most philosophers have found them unconvincing.<sup>38</sup> Other utilitarians draw attention to cases in which ordinary morality requires individual sacrifice for the common good. They then proceed to explain these moral judgments as instances of the principle of utility. Note, however, that utilitarian generalizations of this kind can only do so much. The moral data taken from cases in which acts in agreement with the principle of utility seem right does not compel us to accept that principle as true. In the philosophy of science, this is known as the underdetermination of theory by evidence. We can agree with the classical utilitarian about particular moral judgments without subscribing to her general view, as there will very likely be an alternative theory that also supports these judgments, e.g., by appealing to a principle that requires us to bring about only certain goods, and that only in certain kinds of situations.

There is no quick and easy way to break this stalemate, but there are other considerations, to which I will turn now, that disfavor utilitarianism. These considerations concern situations that are not fantastic but actual or at least realistic, which allows us to be more confident about our intuitive responses, particularly after we have subjected them to critical reflection.

### ***1.2. Classical utilitarianism arrives at right conclusions for the wrong reasons***

In the previous section, I took issue with classical utilitarianism on the count that there are possible scenarios in which it yields results that clash with the common moral

---

<sup>38</sup> The problems with Mill's proof, offered in in Chapter IV of *Utilitarianism* (Mill 1861, pp. 35-41), are well-known. For criticism of Hare's proof of utilitarianism, see Feldman 1984, and Carson 1986. Note that Hare took himself to have proven preference utilitarianism, not classical utilitarianism.

consciousness. This is only one of a variety of ways in which one may take issue with a given ethical theory. Another way is to object not to its implications for acts in terms of rightness or wrongness, but to the manner in which it arrives at, or the reasons it generates in support of, these judgments. An ethical theory might rightly imply that a certain act is morally right or wrong, yet fail to give a satisfactory explanation of why that is so. In this section, I will argue that classical utilitarianism is such a theory and often fails to properly account for the wrongness of acts of killing it rightly condemns.

Consider a situation in which a large group of people derives much pleasure from the suffering and death of a few unfortunate victims. We need not venture into the realm of the fantastic to find an example for this kind of entertainment. To the displeasure of a great many people who care about animals, dog fighting continues to be practiced on a large scale, even though it is illegal in most parts of the world. A prominent example in history are the Roman amphitheaters, where not only non-human animals but humans too were victimized. Scenes more or less like the following actually took place.

*Colosseum.* At an emperor's command, an innocent slave is brought to the Colosseum. The emperor orders that three fierce lions be released into the arena, who then kill and devour the slave at once. The large audience present on the occasion derives much sadistic pleasure from watching the spectacle unfold.

Given the great amount of pleasure people in the audience derive from attending this scene, must the classical utilitarian not commend the emperor's acts? No, quite the contrary: She must condemn the emperor's acts. That is because, as Hare reminds us, "what ought to be done [...] depends on the alternatives to doing it."<sup>39</sup> According to classical utilitarianism, it is not enough to show that an act brings about more good than

---

<sup>39</sup> Hare 1981, p. 142.

bad to establish its moral permissibility or obligatoriness. What needs to be shown is that no other act available to the agent would bring about a greater balance of good over bad, and it

“would be absurd to suggest that there was no other way in which the Roman populace could get its pleasures. The right thing to have done from the utilitarian point of view would have been to have chariot races or football games or other less atrocious sports; modern experience shows that they can generate just as much excitement.”<sup>40</sup>

It is hard to dispute Hare’s claims, and I feel no urge to try. Surely, a more desirable and useful course of action for the emperor to take would be to facilitate a change in the Romans’ taste in pleasure, a move away from their sadistic delight in cruelty to more innocent joys, such as those of watching football games, and then to provide these pleasures. Not only would the slave be spared but there would likely also be positive effects on society in general. Discouraging malicious glee and promoting cooperative sports fosters compassion and community spirit and can reduce violent and aggressive tendencies, which will likely promote utility in the long term. Whatever act available to the emperor is in fact optimific, whether it is starting a campaign promoting public appreciation for football or something else, we can confidently say that it is not the act of having lions dismember a slave in front of an audience of roaring savages. Hence, classical utilitarianism rightly condemns the emperor’s acts in *Colosseum*.

An opponent of utilitarianism might wonder, what if the Romans cannot rid themselves of their sadistic desires? What if their disdain for slaves is so deeply ingrained that the principle of utility does demand the sacrifice of slaves for their entertainment after all? A

---

<sup>40</sup> Hare 1981, p. 142.

world in which these conditions are fulfilled would be a grim place indeed. It would be very different from the world we live in, or so I sincerely hope, and I promised to keep clear from the fantastic, as the reliability of our intuitions is doubtful in that realm. Instead, it is worth having a closer look at the way classical utilitarianism arrives at its uncontroversial conclusion regarding the normative status of the emperor's acts in a realistic instance of *Colosseum*.

Classical utilitarianism holds that the normative status of an act is determined solely by the value of the consequences of that act and its alternatives. We cannot know whether the emperor's acts in *Colosseum* are immoral unless we ask, for example, how much sadistic pleasure the audience derives from watching the killing of the slave, how much pleasure they would derive from watching a football game instead, etc. The utilitarian explanation of the wrongness of the emperor's acts appeals to the answers to these questions. Having a slave killed by lions for the entertainment of the people is morally wrong because of the fact that other forms of entertainment would bring about less pain, yet as much or more pleasure. The sadistic pleasure enjoyed by the audience is not a *decisive* reason to have the slave killed, but notice that it nevertheless figures as *a* moral reason to kill the slave in the classical utilitarian's deliberation. In order to better understand what that means, we need to have a closer look at utilitarian ethical deliberation.<sup>41</sup>

Classical utilitarianism is a teleological ethical theory which declares that the promotion of utility is the sole ultimate aim of morality. All moral reasons for action it recognizes hence are grounded in pleasure and pain. That  $\phi$ -ing will bring about pleasure or prevent

---

<sup>41</sup> Here, of course, what is meant is not the deliberation classical utilitarianism implies moral agents should engage in before acting. Rather, we will discuss the utilitarian deliberation of reasons relevant to the normative evaluation of acts.



pain is *always* a moral reason to  $\phi$ , and the strength of that reason is proportional to the amount of that pleasure or pain. If an act available to an agent in a given choice situation,  $\phi_1$ , would bring about a particular episode of pleasurable experience that an alternative act,  $\phi_2$ , would not bring about, then  $\phi_1$  ranks higher in the utilitarian calculus, and therefore is more likely to come out as morally permissible, than  $\phi_2$ , all else being equal. A moral reason to  $\phi$  hence is a consideration that counts in favor of the permissibility of  $\phi$ -ing. Similarly, that  $\phi$ -ing will bring about pain or prevent pleasure is *always* a moral reason not to  $\phi$ , and a moral reason not to  $\phi$  is a consideration that counts in favor of the impermissibility of  $\phi$ -ing. Restated in terms of moral reasons, the principle of utility is the claim that an act is morally permissible, if, and only if, no alternative act is better supported by moral reasons.<sup>42</sup>

As having the slave killed brings about pleasure, namely, the sadistic pleasure enjoyed by the audience, the classical utilitarian is committed to the claim that there is a moral reason to have the slave killed that is grounded in that pleasure. In other words, the fact that the murderous spectacle provides entertainment is claimed to do *something* to morally justify or legitimize it, rather than nothing. I regard this as absurd, and as evidence that classical utilitarianism is committed to a defective theory of moral justification.<sup>43</sup> Killing a person is too serious an offense for such consequences as amusement – sadistic amusement, no less – to matter. That is not to say that benefits to others can never justify, or be a moral reason for, killing a person. It may well be morally permissible to kill a person in order to

---

<sup>42</sup> If there is only one permissible act, it is also required. If there is more than one permissible act, the agent is morally required to perform one of them.

<sup>43</sup> R. E. Ewin implicitly registers a similar complaint about this aspect of utilitarianism. He writes that, “[w]hen one man wantonly kills another, not in self-defense or anything like that, but, say, simply because he enjoys killing people, we have no need to wait for the consequences before judging the act to be a murder and therein wrong” (Ewin 1972, p. 129).

avoid a disastrous war, or some other significant evil. What I think is unacceptable is the idea that *each and every* benefit a third party would derive from the death of a person constitutes a moral reason to kill that person. Because of that feature of classical utilitarianism, its proponents must find and cite moral reasons outweighing the utilitarian moral reason to have the slave killed in order to make the determination that, and explain why, the emperor's actions are immoral. In contrast, I suggest that there is no need to do such outweighing in the case at hand, as it is deeply implausible that there is a moral reason to kill the slave for the entertainment of the masses in the first place. Classical utilitarianism generates too many moral reasons.<sup>44</sup>

So far, I have shown that classical utilitarianism implies that, if others would get a benefit from a person's death, no matter how trivial or unimportant, then that is a moral reason to kill that person. I argued that this implication is implausible and reveals a flaw in the way classical utilitarianism determines and explains the normative status of acts of killing relevantly similar to the one occurring in *Colosseum*. The problem however is not limited to such situations; it is more general than that. In order to be able to more fully appreciate the extent of the problem, it is instructive to consider Richard G. Henson's (in)famous paneuthanasia argument.<sup>45</sup>

In a widely-discussed article, Henson has claimed that "a consequence of utilitarianism is a sort of cancerous euthanasia, growing wildly out of control."<sup>46</sup> His main argument is easily stated. Consider a person at some point in time,  $t_1$ . If he is painlessly killed in an

---

<sup>44</sup> Classical utilitarianism also generates not enough moral reasons. If you have made a promise, that is a moral reason to do what counts as fulfilling that promise, regardless of whether doing so brings about any good.

<sup>45</sup> "Paneuthanasia argument" is Henson's term, not mine, cf. Henson 1971, p. 326.

<sup>46</sup> Henson 1971, p. 331.

instant at  $t_1$ , he will not experience any pleasure or pain at any time after  $t_1$ .<sup>47</sup> If he is not killed at  $t_1$ , he will die at some later point in time,  $t_2$ . Disregarding effects on third parties, a comparison between these alternatives in terms of utility turns on the value of the remainder of the life he would lead, if he were not killed at  $t_1$ . Killing him scores zero on the utility scale, and hence outranks not killing him, if the life he would lead between  $t_1$  and  $t_2$ , were he not killed at  $t_1$ , contains more pain than pleasure. Hence, Henson concludes, classical utilitarianism implies that he “ought to be killed merely to save him from moderate unhappiness hereafter, provided he would not be too much missed.”<sup>48</sup> What is true for one person is true for others as well, and if we accept Henson’s assumption that only “one out of two persons will enjoy a favorable hedonic balance for the remainder of his natural life,”<sup>49</sup> then we get a cancerous euthanasia indeed.<sup>50</sup> Whether or not that rather pessimistic assumption is justified, the classical utilitarian should be troubled by Henson’s argument.<sup>51</sup>

On closer inspection, however, doubt quickly arises about how much damage Henson’s argument really does to classical utilitarianism. First, it must be noted, as L. W. Sumner reminds us in a response to Henson, that classical utilitarianism “is fully comparative. It requires computation of the utilities of *all* the alternatives open to the agent at the time of

---

<sup>47</sup> I am assuming throughout this dissertation that one ceases to exist at death.

<sup>48</sup> Henson 1971, p. 332.

<sup>49</sup> Henson 1971, p. 329.

<sup>50</sup> According to Henson, the implications of *average utilitarianism* are even harder to accept. Since average utilitarianism requires us to maximize the average utility per person, it would require us to kill any person who is worse off than the average person. “Eventually, the conscientious utilitarian will be playing ‘Ten Little Indians’ with the last ten people alive” (Henson 1971, p. 325). Robert Nozick makes a similar point: “Maximizing the average utility allows a person to kill everyone else if that would make him ecstatic, and so happier than average” (Nozick 1974, p. 41).

<sup>51</sup> Henson calls his assumption that one out of two persons leads a life not worth living “optimistic” (Henson 1971, p. 329).

decision.”<sup>52</sup> Henson specifies two broad alternatives: kill, or do not kill. This obfuscates the complexity of the choice situation we are considering, and we will see that much of the force of the paneuthanasia argument depends on Henson’s inadequate presentation of the choices available to the agent. Typically, there is not one but a large number of available acts that would count as killing, and the same is true for not killing. For example, you can kill by gun, or you can kill by knife. If both are done quickly and more or less painlessly, nothing of interest depends on it. Not killing however can typically take a variety of forms which vastly differ in terms of the value of their consequences. Suppose Asif is the kind of unhappy person Henson has in mind. You can let Asif live but punch him in the face, or you can let him live and become his friend, or you can just not interfere with his life at all, etc. The prudential value of Asif’s future will vary significantly with how you choose not to kill him. Henson’s argument seems to rely on the fact that *nothing* you can do will result in Asif experiencing more pleasure than pain between now and the time of his natural death. Otherwise, Sumner asks, “[w]hy not make him happy instead”<sup>53</sup> of killing him? From a utilitarian perspective, that Asif’s future will *in fact* be bleak if you do not kill him is not enough to morally justify killing him, even if there were no negative side effects. It might be true that what you will in fact do if you choose not to kill him will not make Asif happy, but the relevant question to ask is whether there is anything you *could* do that would make him happy. If there is, that is what you *should* do, other things being equal, and classical utilitarianism rightly condemns killing Asif. For most people whose futures seem bleak, there is hope, and there are more effective ways to benefit them than bringing their life to an end. A

---

<sup>52</sup> Sumner 1976, p. 148.

<sup>53</sup> Sumner 1976, p. 149.

pervasive utilitarian euthanasia campaign hence hardly seems warranted. But, and this is where Sumner's defense of classical utilitarianism fails, there are people who, if they continue to live, will *inevitably* experience more pain than pleasure during the remainders of their lives. Many people who suffer from painful and terminal diseases, for example, fall into that category. Say Susan is in such a regrettable condition. Doctors are confident that she has no more than a few months to live, and those months will very likely contain much misery and little joy, as effective palliative care is not available for patients with her rare disease. Nevertheless, Susan does not want to die and would strongly resist any attempt to kill her. Does classical utilitarianism imply, as Henson argues, that Susan must be killed against her will? This question brings me to the second point that I want to make about the paneuthanasia argument.

In a realistic choice situation, we cannot disregard effects on third parties. At first glance, classical utilitarianism might seem to imply that we are morally required to kill Susan against her will – and surely that would be a very damaging implication. But, if we take a closer look, we will come to think of the grief her family and friends would experience if Susan were to be killed, and the damage the killing, if discovered, would do to the public's sense of security. Imagine a society in which you have to fear being killed at any time, because somebody might come to sincerely believe that there is no hope for you and you would be better off dead and not be missed. Such a society is hardly worth aspiring to. In anything like normal circumstances, killing Susan against her will is not the most useful thing one can do, regardless of the quality of the life that lays ahead of her. In general, involuntary euthanasia typically has many significantly negative actual

and possible side-effects, and classical utilitarianism hence practically always rightly condemns instances thereof.

Nevertheless, Henson's criticism points to a serious problem. That problem however does not lie in that classical utilitarianism cannot account for the wrongness of involuntary euthanasia – I have argued it can – but in that it gives a wildly implausible account of the wrongness of involuntary euthanasia. As Susan will experience more pain than pleasure if she continues to live, the utilitarian reasons to kill her that arise from facts about Susan and her well-being are stronger than so-grounded reasons not to kill her. The utilitarian condemnation of killing Susan hence depends on the negative effects that her death would have on the people who care about her and society in general, which give rise to moral reasons that tip the scale otherwise tilted in favor of killing Susan by facts about Susan. It seems clear, to me anyway, that classical utilitarianism condemns killing Susan largely for the wrong kinds of reasons, by appealing to features of the situation that are, at best, marginally important. Explaining the wrongness of killing Susan in terms of her death's effect on others misses the main point. The wrongness of killing Susan has something to do with Susan, not with the consequences for others. The fact that Susan lives an unpleasant life and has little hope for an enjoyable future, by itself, is no justification *whatsoever* for killing Susan.<sup>54</sup> Hence, there should be no *need* to appeal to effects on others to explain the wrongness of killing Susan.<sup>55</sup>

Classical utilitarianism generates too many moral reasons in situations in which we have the option of killing somebody with an inevitably bleak future, just like it generates too

---

<sup>54</sup> This is not to say that the fact that Susan lives a life not worth living cannot be part of a moral justification for killing Susan, for example, when combined with a competent request by Susan to be killed.

<sup>55</sup> The point made in this paragraph is closely related to Thomas L. Carson's observation that classic utilitarianism falsely "implies that we have a *prima facie* duty to kill people who would be better off dead [...]" (Carson 1983, p. 55).

many moral reasons in *Colosseum*. The result is a wildly implausible explanation of the wrongness of involuntary euthanasia. Any plausible explanation of the wrongness of involuntary euthanasia must appeal primarily, if not exclusively, to facts about the victim. Classical utilitarianism does not have the resources to provide such an explanation, and hence is defective in that regard.

### ***1.3. Classical utilitarianism gets the comparative wrongness of killing wrong***

We generally expect every theory of right and wrong to answer, for any act, the questions, “Is it morally permissible, or impermissible?” and “Why does it have the permissibility status it has?” In the previous two sections, I have argued that there are acts about which classical utilitarianism cannot satisfactorily answer these two questions. I have objected to classical utilitarianism on the grounds that it has unpalatable implications for rare yet possible acts in terms of rightness or wrongness, and that it arrives at right conclusions for the wrong reasons. Now, there is another question one can ask about acts. If an act is morally wrong, one can ask *how* wrong it is. In this section, I shall argue that classical utilitarianism also fails to give a plausible answer to this question in many cases that involve killing.

In one sense of “wrong,” wrongness is an all-or-nothing matter: An act is wrong just in the case that it is not morally permissible, and either it is the case that an act is permissible, all things considered, or it is not. In another sense of “wrong,” an act is more seriously wrong (“more wrong”) than another, if the reasons why it morally ought not to be done are stronger, and it hence would take weightier countervailing considerations to

counterbalance or outweigh these reasons.<sup>56</sup> The idea that wrongness comes in degrees has an important place in moral thinking. For one, the degree of wrongness of typical instances of a kind of act tells us something about how hard or easy it is to morally justify acts of that kind. Also, the more seriously wrong an act is, the more guilty one should arguably feel about performing it, and, if retributivism is true, the more severe a punishment one deserves.

For example, virtually everybody accepts that shoving another person to the ground, causing minor injury, and killing a person usually are both morally wrong. It is equally true for both acts that, in normal circumstances, they should not be done. In that sense, they are equally wrong. But everyone also accepts that there is an important moral difference between the two acts. In general, killing is a graver moral offense than battery. Suppose you see that a child is about to get hit by a car while crossing the road, and you have only very little time to get to the child, pull her aside, and save her from serious injury, or even death. Somebody not aware of the emergency is standing in your way and there is no time to explain. While it is typically wrong to shove people to the ground, in this situation, it seems permissible to do just that to the person blocking your way in order to save the child, but impermissible to push that person in front of the approaching car in order to achieve the same end. It is easier to outweigh the reasons that typically make battery wrong than those that typically make killing a person wrong. We can imagine, without much difficulty, many such situations in which battery might be morally justified. But there are only few situations, if any, in which killing a person is justified. The prohibition against killing is more stringent than the prohibition against battery. Killing is more seriously wrong than battery.

---

<sup>56</sup> Cf. McMahan 2002, p. 190 & Lippert-Rasmussen 2007, p. 717.



The fact that classical utilitarianism seems well-equipped to accommodate such judgments is often cited in its favor. Shoving somebody to the ground is less seriously wrong than shoving somebody in front of a car because the consequences of the former act typically contain a more favorable balance of pleasure over pain than the consequences of the latter act. In fact, if shoving somebody to the ground is the only way to save another person from serious injury or death, then doing so is not morally wrong at all. In general, the consequences of killing a person are far worse than the consequences of battery. Classical utilitarianism hence supports our intuitive judgment that killing is more seriously wrong than battery, and almost all other wrongs too for that matter. Killing a person typically has disastrous consequences, paralleled only by the consequences of very few acts of other kinds, which lends initial credence to the claim that classical utilitarianism can properly account for the extraordinary moral gravity of killing.

According to classical utilitarianism, an act of killing is wrong to the degree that it falls short, in terms of the value of its consequences, of the utility-maximizing act or acts available to the agent at the time of the act. Even though, in most choice situations, all acts of killing available to the agent are a far cry from the best she can do to promote utility and hence seriously wrong, the value of their consequences may vary a great deal. The future of an always-cheerful person,  $P_1$ , contains a more favorable balance of pleasure over pain than the future of a person disposed to depression,  $P_2$ . The future of  $P_1$  hence provides more utilitarian reason against killing  $P_1$  than the future of  $P_2$  provides against killing  $P_2$ . If  $P_1$  and  $P_2$  contribute equally to the well-being of others and are equally loved, and if their deaths would be equally mourned, etc., then classical

utilitarianism implies that it would be more seriously wrong to kill  $P_1$  than it would be to kill  $P_2$ . This is profoundly counterintuitive and at odds with the widely held belief that all persons are equal, and hence deserving of equal protection. The wrongness of killing is commonly thought not to depend on the prudential value of the victim's future, perhaps unless that value is deep in the negative.

The quality of a person's future life is not the only factor that affects the value of the consequences of that person's death, yet seems irrelevant to the degree of wrongness of killing that person. Here are a number of other such factors:

- *Age*: Classical utilitarianism implies that, other things being equal, it is more seriously wrong to kill a young person than an old person, as the former could have expected to live longer and hence aggregate more good in the remainder of his life than the latter.
- *Life expectancy*: Classical utilitarianism implies that, other things being equal, it is more seriously wrong to kill a person who would otherwise have died shortly – for example, in a traffic accident – than a person who would otherwise have continued to live for a long time, for the same reason that it is more seriously wrong to kill a young person than an old person.
- *Popularity*: Classical utilitarianism implies that, other things being equal, it is more seriously wrong to kill a popular and widely loved person than an unpopular person, with no friends and family, as the former's death will cause more grief than the latter's death.<sup>57</sup>

---

<sup>57</sup> Nobody lives forever. Hence, a popular person will be mourned regardless of whether she is murdered while in the prime of her life or dies of natural causes when old. The point is that grief is typically more severe when a person dies prematurely. Also, if one lives to an old age, many close friends will have died already by the time one dies. Hence, there will be a smaller number of mourners.

- *Usefulness*: Classical utilitarianism implies that, other things being equal, it is more seriously wrong to kill a highly useful member of society, such as an exceptionally skilled surgeon, than a person who has little effect on the well-being of others, such as an eccentric and reclusive retiree who spends most of his waking hours on the couch watching TV, as the former would have brought about more utility in the remainder of his life than the latter.
- *Social status*: Classical utilitarianism implies that, other things being equal, it is more seriously wrong to kill a privileged than an underprivileged person, as it is reasonable to expect that the former would have had a more pleasant future than the latter.
- *Killer's conscience*: Classical utilitarianism implies that, other things being equal, a killer who regrets that he killed his victim and consequentially finds himself drowning in remorse committed a more serious wrong than a remorseless killer who derives great pleasure from his deed.

Not everybody finds the idea that the degree of wrongness of a killing varies with the victim's age intuitively objectionable.<sup>58</sup> Dale Jamieson correctly observes that

“there is a thread in our untutored judgments that seems to support [...] [classical utilitarianism]. We seem to believe that it is a particularly bad thing when a child is struck down. We defend this belief by saying that ‘she had her whole life ahead of her’. We adopt special protections and policies for children because we seem to believe that their lives are especially valuable.”<sup>59</sup>

---

<sup>58</sup> Arguments for the relevance of age can be found in Lippert-Rasmussen 2007 & Soto 2013.

<sup>59</sup> Jamieson 1984, p. 215. However, Jamieson also points out that “many people think that killing an adult is worse than killing an infant” (Jamieson 1984, p. 215), and that “[w]e don't think that it is worse that a 7 year-old dies than a 9 year-old, nor do we think that it is worse that a 40 year-old dies than a 42 year-old”

Presumably, one could similarly defend life expectancy as a factor relevant to determining the wrongness of killing. If a person is certain to die of an incurable disease within a few weeks, that person does not have much of a life ahead of her. In contrast, other people look at engaging in long-term projects that give their lives meaning and purpose, and decades of harmonious family life. Does it follow that it is less seriously wrong to kill a person who is terminally ill? I think not, but I do not want to argue for that here.

It is sufficient to note that the fact that classical utilitarianism renders the remaining factors in the above list relevant is quite enough to safely conclude that classical utilitarianism fails to account properly for the wrongness of killing. Surely, whether or not you may use deadly force to defend yourself against an attacker, for example, does not at all depend on the popularity, usefulness, or social status of the attacker, but rather on considerations of necessity and proportionality. If, in a particular self-defense situation, you are morally justified to kill an unpopular, less socially useful, or underprivileged attacker, then you would also be justified to kill a popular, useful, or privileged attacker in an otherwise identical situation, and vice versa. In general, the fact that the death of an unpopular, less socially useful, or underprivileged person forecloses a smaller amount of utility than does the death of a popular, useful, or privileged person does not make it easier to justify killing the former than killing the latter. It is also obviously false that the degree of wrongness of a killing is inversely correlated with the amount of pleasure the killer derives from that killing.

---

(Jamieson 1984, p. 215). Hence, at best, the common moral consciousness lends only partial support to utilitarianism but also opposes it.

More counterintuitive implications emerge if we compare acts of killing to other kinds of acts. As classical utilitarianism judges acts only by the value of their consequences, the fact that an act is an act of killing, as such, is irrelevant to the normative status of that act. It also makes no difference to the normative status of an act whether it brings a certain amount of disutility to one or a few, making the burden per person large, or the same amount of disutility to many, making the burden per person small. Suppose John, an American, is on holiday in Yemen and gets into a heated argument with Sameer, a local tea vendor, who he feels has grossly overcharged him because he is a foreigner. John has a horrible temper, and a disdain for Arabs. The argument quickly escalates and John starts beating the poor man and does not stop until he is dead. Nobody is around to witness the murder. As it happens, Sameer does not have any family or friends, and it is unlikely that anybody will bring charges against John, and particularly unlikely that anybody will do so before John is back home and out of the reach of the Yemeni authorities. Even if police gets to know about the crime and starts to suspect him while he is still in Yemen, he knows that he can stop any investigation against him by paying a generous bribe. John quickly forgets this episode and never feels any remorse about what he has done. Killing Sameer, of course, was a grave wrong, and the classical utilitarian readily agrees. The murder did not have any significant bad side-effects, but, even though old and poor, Sameer was in good health and would have continued to live a mostly pleasant life for many more years if John had not killed him. John deprived Sameer of a future worth living for. Almost all of the acts available to John at the time, such as paying the seemingly excessive price and walking off, would have had much better consequences than killing Sameer. – Nothing interesting so far, but now consider this

event: On New Year's Eve in 1994, Rod Stewart performed a concert in Rio de Janeiro. It was one of the largest concerts in history. Three and a half million people attended. If Mr. Stewart had decided to cancel the concert just minutes before its scheduled starting time, a great many people would have been very upset and disappointed. Some concert attendants would have found another way to celebrate the new year but many would have gone home or back to their hotels in frustration. For the latter, the night would have been spoiled. The unexpected cancellation would also have generated bad publicity for the city of Rio de Janeiro and the event management firm that has organized the concert, resulting in loss of future revenue and possible layoffs. Cancelling the event ranks very significantly lower on the utility scale than going through with the concert. In fact, it seems clear that cancelling the New Year's Eve concert would have been even more of a failure to maximize utility on Mr. Stewart's part than killing Sameer was on John's part. To illustrate this, let us assign the number +10 to the average amount of pleasure minus pain that Sameer would have experienced per day in the remaining, say, ten years of his life. Accordingly, by killing Sameer, John foreclosed a total amount of pleasure minus pain of  $10 \text{ day}^{-1} \cdot 365.2425 \text{ days/year} \cdot 10 \text{ years} \approx 36,524$ . If cancelling the concert would have upset the average concert goer for just half a day, and if the amount of pleasure minus pain he would have experienced during that half a day would have been 2 units smaller than the one he actually experienced, then Mr. Stewart would have foreclosed a total amount of pleasure minus pain of  $2 \cdot 3,500,000 = 7,000,000$  by cancelling the concert. That number is almost two hundred times larger than the number representing the amount of utility John foreclosed, and we did not even take into account the bad consequences a cancellation would have had for the event management firm, its

employees, the city, etc. Therefore, classical utilitarianism implies that it would have been more seriously wrong for Mr. Stewart to cancel his historic concert in Rio de Janeiro than it was for John to murder Sameer. That cannot be right. The vicious murder of Sameer was a very grave moral wrong, particularly because it was partly motivated by racial hatred and there were no mitigating circumstances. In contrast, cancelling a concert and thereby upsetting people, regardless of their number, appears to be a minor misdeed, wrong not so much because it upsets people but because it involves a failure to make good on a contractual promise to provide entertainment.

The foregoing example shows that it is sometimes less seriously wrong to foreclose more hedonistic good. Classical utilitarianism cannot accommodate that fact, and consequentially sometimes gets the comparative wrongness of acts, including acts of killing, wrong. This problem for classical utilitarianism appears to have its roots in its defective account of moral reason which we briefly discussed in the last section. According to that account, each and every good or bad that results from an act affects that act's comparative normative status. If the expected consequences of two wrong acts available to an agent in a given choice situation,  $\varphi_1$  and  $\varphi_2$ , are value-isomorphic, except for the fact that there is one additional intrinsic good (bad) that  $\varphi_1$  would bring about but  $\varphi_2$  would not, then the classical utilitarian must conclude that it would be less (more) seriously wrong to  $\varphi_1$  than it would be to  $\varphi_2$ , regardless of the respective natures of  $\varphi_1$ ,  $\varphi_2$ , and the additional good (bad). This yields implausible results as there are kinds of acts, such as wrongful killing, that are too serious for trivial side-effects to affect their degree of wrongness. The fact that sensation-mongering onlookers get some excitement out of watching a murder taking place in front of them does not make that murder any less

wrong. Nor does it typically lessen the moral gravity of a wrongful killing that the victim's premature death saves his health insurer some money. Also, no ordinary person would try to downplay the gravity of killing another person by pointing out that the killer enjoyed the act, that the victim was not of much use to others, that the killing was done painlessly, or that the victim did not have much to expect from a continued life.<sup>60</sup>

Taken together, the three objections developed up to this point constitute a sufficient reason to move away from classical utilitarianism and look for another ethical theory that can provide a more plausible account of the wrongness of killing. I have argued that classical utilitarianism renders morally right what is wrong, arrives at accurate moral judgments for the wrong reasons, and cannot properly account for the comparative wrongness of killing. Having achieved a good understanding of these objections, we can now ask whether they also apply to consequentialist ethical theories other than classical utilitarianism. In doing so, I shall focus on the last two objections as I think that they pose a more serious problem than the first objection. It is incredibly hard to know, if it is at all practically possible, how realistic a situation would be in which it were optimistic for a law enforcement officer to frame an innocent scapegoat who is then executed, and it seems wise not to build one's argument against an ethical theory on one's intuitive response to that theory's implications for such unusual cases that potentially are far removed from our ordinary experience and moral life. If there is a consequentialist ethical theory that avoids some or all of our objections to classical utilitarianism and is not subject to similarly damaging new ones, it is surely to be preferred – at least as long as a sound proof of classical utilitarianism is still lacking.

---

<sup>60</sup> Henson makes a closely related point in Henson 1971, cf. p. 330.



There are three issues on which consequentialist ethical theories divide. First, consequentialists can disagree on what sort of consequences are relevant to the normative status of acts. Second, they can disagree on how to spell out the relationship between the rightness and wrongness of acts and the value of relevant consequences. Third, consequentialists can subscribe to different theories of value. We will take classical utilitarianism as our departing point, from which we will move to other consequentialist ethical theories by adjusting one or more of these variables.

#### ***1.4. Rethinking well-being: non-hedonistic act-utilitarianism***

Even though value-hedonism still has its defenders, it has lost much of its currency in contemporary moral philosophy. One reason for that is that there seem to be instances of people being harmed or benefitted by things that never enter their experience and do not cause or prevent any pleasure or pain. For example, is your being pleased with falsely believing that your partner is faithful less good for you than that pleasure would be if it was based on a true belief? If so, then well-being must have an objective component. Robert Nozick's famous experience machine thought experiment is often taken to establish just that point and thereby decisively refute value-hedonism. Suppose you have the chance to hook yourself up to a machine that intercepts the usual sensual connection between you and the world around you and feeds your brain an alternative reality in which you experience being a rock star or a successful athlete, or both, or whatever else gives you pleasure. Would you take that chance and hook yourself up? Nozick suggests that most people would decline, for a variety of reasons. People want "to *do* certain

things, and not just have the experience of doing them [and they] want to *be* a certain way, to be a certain sort of person [instead of being] an indeterminate blob [vegetating next to a machine].”<sup>61</sup> He concludes that there are things that “[matter] to us in addition to experience,”<sup>62</sup> and most commentators further infer that the reason why these things matter to us is that they make our lives go better or worse.<sup>63</sup> But what are these things that are not mental states yet have a non-instrumental prudential value?

The by far most popular response to this question explains the common intuitive reaction to Nozick’s thought experiment by appealing to the fact that much of what we get when hooked up to the experience machine is not what we actually want. A life in the experience machine, proponents of that view argue, is bad for us and hence not an attractive choice, because, even though it is full of pleasure, many of our most central desires remain unfulfilled. You wanted to be a rock star and you think you got what you wanted, but in fact you did not. What you got instead is the *illusion* of being a rock star, which is not what you wanted. You wanted to be a successful athlete and you think you have won numerous competitions, but in fact your body has not moved in years. Etc. According to *desire-satisfaction theories of well-being*, actual states of affairs are important to the quality of our lives, not our perception of reality. Our lives go well to the extent that we get what we want, and badly to the extent that we do not get what we want. If we combine this standard of good with maximizing act-consequentialism, we get

---

<sup>61</sup> Nozick 1974, p. 43.

<sup>62</sup> Nozick 1974, p. 44.

<sup>63</sup> Unlike these commentators, I think that the lesson we should take away from Nozick’s thought experiment is that people care about things other than, and have wants that are irrelevant to, their own well-being.

desire-utilitarianism, or *preference-utilitarianism*, an ethical theory most prominently associated with Hare and Peter Singer.<sup>64</sup>

Preference-utilitarianism provides a strong direct reason against killing. Almost all people have numerous future-directed desires, many of which are of considerable strength. They are engaged in projects they want to complete, they want to see their children grow up, they have plans for their retirement, etc. Most importantly, they want to continue living. “To kill a person is therefore, normally, to violate [...] a wide range of the most central and significant preferences a being can have. Very often, it will make nonsense of everything that the victim has been trying to do in the past days, months or even years.”<sup>65</sup>

When confronted with the objection that classical utilitarianism cannot properly account for the wrongness of killing Susan, moving to preference-utilitarianism may seem like the obvious strategy for a utilitarian to adopt. Susan, we recall, would inevitably experience more pleasure than pain in the remainder of her life if she continued to live. Nevertheless, she has a strong desire to stay alive. According to preference-utilitarianism, the fact that killing Susan would frustrate her desire to stay alive is a weighty moral reason not to kill her, regardless of how bright or bleak her future may be. Pleasure and pain of course are still relevant insofar Susan presumably prefers pleasure over pain – a preference that will largely be frustrated if she continues to live –, but the utilitarian can plausibly argue that the preference to stay alive rather than die and other future-directed preferences are typically strong enough to outweigh the preference for pleasure over pain and hence

---

<sup>64</sup> Cf. Hare 1981, and Singer 2011.

<sup>65</sup> Singer 2011, p. 80.

make room for an explanation of the wrongness of killing Susan that is primarily based on facts about Susan and does not depend on side-effects.<sup>66</sup>

Utilitarianism's euthanasia problem reemerges however when we distinguish between the desire-satisfaction theory of well-being in its simple, unrestricted form and its restricted variations. The unrestricted desire-satisfaction theory, unlike its restricted counterparts, holds that *all* of a person's desires are relevant to her well-being, regardless of their origin or content. It is vulnerable to a variety of counterexamples, which are intended to reveal that getting what one wants is not always conducive to one's self-interest. One category of such counterexamples concerns other-regarding desires. Its most famous representative in the literature is arguably the following thought experiment by Derek Parfit:

“Suppose that I meet a stranger who has what is believed to be a fatal disease. My sympathy is aroused, and I strongly want this stranger to be cured. We never meet again. Later, unknown to me, this stranger is cured. On the Unrestricted Desire-Fulfillment Theory, this event is good for me, and makes my life go better. This is not plausible. We should reject this theory.”<sup>67</sup>

The unrestricted desire-satisfaction theory of well-being overreacts to the shortcomings of value-hedonism. While it has the desirable feature of accommodating the

---

<sup>66</sup> One could object that this does not work if the misery that lies ahead is very great. For example, consider a man who is in constant pain and has only few future-directed desires that are mostly insignificant, yet wants to stay alive because he is scared of death. Is a theory of value plausible that attaches more value to the satisfaction of his desire, however weak, to stay alive (and the few other future-directed desires he has) than to the satisfaction of his desire, however strong, to be free of suffering? I think not, but it would be futile to push this point here, for at least two reasons. First, I do not know how to put numbers on the strengths of desires as diverse as those relevant here, and even if I had a method of assigning numbers to the strengths of preferences, that method would likely be very controversial. Second, I will develop another objection that I think is less controversial, more interesting, and decisive by itself.

<sup>67</sup> Parfit 1984, p. 494. The same objection has been made by James Griffin and Shelly Kagan, cf. Griffin 1986, pp. 16 f., and Kagan 1998, p. 37.

unexperienced goods and bads that motivate a move away from value-hedonism, it also renders prudentially relevant things that have little to nothing to do with one's life, are not *about* one's own life in the relevant sense, and seem well beyond the bounds of what can affect one's well-being. If I see a sniper taking aim at a senator, want the politician not to get hurt, and satisfy that desire by throwing myself in front of him and being shot in his stead, then I am not thereby better off.<sup>68</sup> While other-regarding desires that closely relate to one's life and therefore also have a significant self-regarding component – such as the desire that *one's own* children will have a fulfilling career – may count towards one's well-being, the satisfaction of *purely* other-regarding or altruistic desires should not.<sup>69</sup>

Another restriction a credible desire-satisfaction theory of well-being must incorporate concerns desires that are based on false beliefs, or are otherwise ill-informed. As James Griffin puts it, “notoriously, we mistake our own interests. It is depressingly common that when even some of our strongest and most central desires are fulfilled, we are no better, even worse, off.”<sup>70</sup> Suppose you have, and always had, a strong desire to be a member of the most exclusive country club in town. You work hard, get a well-paying job, socialize with the right people, make your way into high society, and finally get offered a membership, just to find out that the people at the country club bore you terribly. You are put off by the snobbish atmosphere at the club and, contrary to what you thought, golf

---

<sup>68</sup> Mark Overvold discusses this special case of other-serving behaviour in Overvold 1980, where he argues that an unrestricted desire-satisfaction account of well-being is inconsistent with the possibility of self-sacrifice.

<sup>69</sup> Arguably, purely altruistic desires constitute a subclass of the larger class of remote desires, the satisfaction of which does not affect one's well-being. Examples for desires that are not altruistic but remote in the sense I intend are the desire that people living three hundred years from now will be in contact with extraterrestrials and the desire that the number of water molecules in the Atlantic Ocean is odd.

<sup>70</sup> Griffin 1986, p. 10.

and polo are just not for you. In hindsight, you much regret having spent so much time and effort to become a member. If the satisfaction of each and every desire was to count towards your well-being, then being a country club member would be good for you. Similarly, it would be good for you to satisfy your desire to drink from a river, even if, unknown to you, doing so would make you sick.<sup>71</sup> None of the two seem right.<sup>72</sup> That has led Richard B. Brandt – to name just one among many – to suggest that only desires you would have if you were fully informed be counted.<sup>73</sup>

If we accept that purely altruistic and ill-informed desires, or at least those among them that are relevantly like the desires in our examples, need to be discounted, we see that, pace Dale Jamieson and other utilitarian apologists, it is not at all “close to incoherent [...] [to imagine] people whose lives are not worth living and will never be, [but] who want to live anyway.”<sup>74</sup> There is no reason why the general problem purely altruistic and ill-informed desires pose for the unrestricted desire-satisfaction theory should not also manifest itself in situations relevantly similar to Susan’s. Other future-directed desires

---

<sup>71</sup> Carson gives this example in Carson 2000, pp. 72 f.

<sup>72</sup> For a detailed and useful discussion of different strategies to restrict desire-satisfaction theories of well-being, see Griffin 1986, and Sumner 1996.

<sup>73</sup> Cf. Brandt 1979. In contrast, Krister Bykvist argues that there is no need to impose an information constraint to deal with ill-informed desires. Instead, he makes a distinction between intrinsic and instrumental desires, and suggests that desires such as the ill-informed ones we just discussed should be discarded because they fall into the latter category. “I have an intrinsic desire for something if I desire it as an end in itself, in virtue of what it is in itself. I have an instrumental desire for something if I desire it as a means to an end” (Bykvist 2010, p. 44). Your desire to drink from a river that unknown to you carries poisonous water, for example, depends on your belief that the river carries clean water and your intrinsic desire to drink clean water. Hence, it is instrumental rather than intrinsic, and its satisfaction does not contribute to your well-being. Fortunately, we do not need to decide whether Brandt’s or Bykvist’s solution to the problem of ill-informed desires is to be preferred. For our argument, it is enough to note that Bykvist and Brandt, and most other philosophers who have thought about the issue, agree that certain ill-informed desires must be discounted. Two other categories of desires that a credible desire-satisfaction theory of well-being should arguably discount I will only mention. One concerns desires that are the result of brain-washing, manipulation, or addiction, and the other is made up of desires that are based on errors in reasoning, or are otherwise irrational: “[A] person might know that going to the dentist is in his interest, but still he prefers and chooses not to go, because he is weak-willed. The claim is that desire satisfactionism implies, incorrectly, that since he desires not to go, and all desire satisfactions are good for a person, it is good for him not to go to the dentist” (Heathwood 2006, p. 545).

<sup>74</sup> Jamieson 1984, p. 213.

aside, preference-utilitarianism loses its competitive advantage over its classical utilitarian rival in cases in which the desire to continue living must be discounted because it arose entirely out of concern for others or is based on a false belief.

Suppose Susan has a terminal cancer and is in great pain, and that a quack doctor has convinced her that she can beat the cancer through a raw food diet. Her belief in the efficacy of adopting a raw food diet is the only thing that keeps her going. If she believed that her condition will progressively worsen no matter what she eats, as it in fact will, she would commit suicide, or ask to be killed. In this case, Susan's desire to continue living is a paragon of an ill-informed desire that should be discounted in a credible desire-satisfaction theory of well-being. Susan might have other future-directed desires that do count towards her well-being, but most of them will be frustrated regardless of whether she dies from cancer a little later or at the hands of a moral agent now. The preference-utilitarian calculus hence mainly revolves around Susan's standing desire to enjoy life and avoid pain, and may well yield the classical utilitarian result that Susan would be better off if her desire to continue living were frustrated. Yet, killing her remains immoral at least as long as she keeps wanting to live. In order to accommodate that judgment, preference-utilitarianism has to recourse to side-effect, and we are back to an explanation of the wrongness of killing Susan that largely misses the point.

Next, consider a man with a very rare disease who participates in a clinical study that aims to find a cure for that disease. As the study progresses, it becomes clear that a cure will not be found in time to save him, and that he will live in agony for the short remainder of his life. However, with every day the man stays in the study, the researchers learn more about his disease. For the sake of others who might contract the disease in the

future, and only for their sake, the man wants to go on living, despite the suffering that he knows will come with staying alive. If he did not care about others, he would want to die. The man's desire to continue to live is based entirely on concern for others and aimed at a goal that is so far removed from his life that it would be implausible to think it relevant to his well-being. It seems not to be in his interest that it be satisfied, and yet killing the man against his will would be seriously wrong.

Every plausible desire-satisfaction theory of well-being allows for people who want to stay alive yet would be better off dead, just like value-hedonism does. Hence, a utilitarian ethic based on any of those theories of well-being has to resort to side-effects to establish and explain the wrongness of killing such people against their will, which falsely draws the attention to third parties instead of the victims, where it belongs. In the standard classification of theories of well-being, this leaves us with *objective-list theories*.<sup>75</sup> Unlike their competitors, objective-list theories of well-being recognize a plurality of ultimate goods. The list virtually always includes pleasure or desire-satisfaction, or both, but it also includes things that are said to be intrinsically good for people independently of whether they are enjoyed or wanted, such as achievement, friendship, mental and physical functioning, knowledge, and autonomy. If an objective-list utilitarian chooses to include autonomy in his list, he thereby gains additional leverage in explaining the wrongness of killing the cancer-ridden Susan, or the man who heroically bears his pain for scientific and medical progress. While death would spare both of them much pain and deprive them of only little pleasure, it also thwarts their autonomous wish to stay alive

---

<sup>75</sup> For a useful discussion of objective-list theories of well-being, see Griffin 1986.



and prevents them from exercising their autonomy ever again.<sup>76</sup> Hence, if sufficient weight is attached to the value of autonomy, the objective-list utilitarian's explanation of the wrongness of killing Susan, or the man, need not appeal to the consequences for third parties. Objective-list theories also offer an attractive alternative to the desire-satisfaction theory's solution of the problem that non-experiential value poses for value-hedonism. For example, an objective-list theorist can argue that what is bad about living one's life hooked up to Nozick's experience machine is not that one's desires are frustrated, but that it is just bad for people to be deceived, and directly good for them to actually achieve things, be certain sorts of people, etc. This is attractive because it seems more intuitive, for example, to explain the desirability of being a successful athlete in terms of its being independently valuable than to explain the value of being a successful athlete in terms of its being desired.

Objective lists of intrinsic prudential goods are often compiled, not by means of some general principal, but with recourse to considered judgments about particular cases. Because of that, objective-list theories of well-being are sometimes quickly dismissed as arbitrary. That, however, hardly seems warranted, as their major rivals – value-hedonism and desire-satisfaction theories – too must rely on reflective judgment or intuition to justify their respective one-item lists. In order to reach a fair evaluation of objective-list theories of well-being, we would have to look closely at how each one of them decides what goes on the list. We further would have to see how well these theories fare when applied to test cases, both in comparison to each other, and in comparison to value-hedonism and desire-satisfaction theories. For our purposes, however, we need not decide

---

<sup>76</sup> Other items on the list might be relevant, too. If achievement is on the list, and if making it through yet another round of tests and thereby providing valuable medical data counts as an achievement for the man, then his death would deprive him of that achievement.

whether objective-list theories offer the most plausible account of well-being. It is enough to note that, if they are combined with maximizing act-consequentialism, the resulting objective-list varieties of utilitarianism are *all* subject to earlier-mentioned objections regarding comparative wrongness and the excessive generation of moral reasons. The same is true of non-utilitarian maximizing act-consequentialism, to which I shall now move at once in order to avoid uninteresting repetitions.

***1.5. Good – but not good for anyone: non-utilitarian maximizing act-consequentialism***

Non-utilitarian maximizing act-consequentialism agrees with act-utilitarianism that an act is morally permissible, if, and only if, its consequences are no less valuable than those of any alternative act, but denies that all that matters to the value of an act's consequences is the well-being of those affected. This opens a vast range of possibilities. One could assign intrinsic value to autonomy as such, regardless of whether it is good in itself for people, or one could be of the opinion, as G. E. Moore in fact was, that beauty and knowledge are good, even when they are not good for anyone, etc.<sup>77</sup>

Outside the narrow constraints of utilitarianism, a consequentialism that affords substantial protection to the life of persons, without relying on contingencies such as side-effects, can be found much more easily. There are ample ways in which such protection might be realized – for example, by assigning a sufficiently large value to

---

<sup>77</sup> Cf. Moore 1903, pp. 83-85 & 194, and Moore 1912. Moore's ethical view is often referred to as "ideal utilitarianism." Note that this is not how I use the term "utilitarianism" in this dissertation. I use the term "utilitarianism" only for consequentialist ethical theories that hold that all that matters for the value of consequences is well-being.

(some notion of) justice, life or autonomy, or a sufficiently large disvalue to the killing of innocent persons. While this promises an ethical theory whose evaluation of acts of killing in terms of rightness and wrongness largely coheres with our moral intuitions, none of it guarantees acceptable judgments in all cases, especially if we allow highly unusual cases. Say, we hold that the occurrence of a killing of an innocent person is intrinsically very bad, and that a killing is consequently much worse than a natural death, other things being equal. That makes it generally more difficult to justify killing one to save the lives of many, but it does not help in cases such as *Innocent Scapegoat* where many lynchings – i.e., *killings*, not natural deaths – can be prevented by killing one.<sup>78</sup> More importantly, however, moving away from utilitarianism does nothing, or only little, to address some of the other problems we have discussed earlier.

Any plausible theory of value I can think of – including welfarist objective-list theories and theories that assign positive value to justice, life or autonomy, or negative value to killing the innocent – will recognize at least one of the following things as intrinsically valuable: the pleasure of watching a sunset, enjoying a cup of tea, finally getting that book one has wanted for so long, learning something new about European history, making a new friend. Yet, the fact that killing a person would bring about any of these things does not even begin to justify killing that person. The root of classical utilitarianism's problem with *Colosseum* is not, as one might initially think, that *sadistic* pleasure is falsely recognized as intrinsically valuable. Rather, the problem arises from the fact that, in classical utilitarianism, each and every benefit a third party would derive from the death of a person constitutes a moral reason to kill that person. That, however, is

---

<sup>78</sup> Cf. Pettit 1997.

not only a feature of classical utilitarianism, but of maximizing act-consequentialism in general.

A fundamental flaw common to all forms of maximizing act-consequentialism is that there is no room for considerations of value to be muted. The fact that an act brings about some intrinsically good or bad state of affairs is always to be taken as a consideration that is relevant to that act's normative status, regardless of the kind of act and the kind of good or bad. Particularly in the case of killing, that has implications that are profoundly implausible.

First, as we just noted, a maximizing act-consequentialist must say that one ought to kill a person if that would allow somebody to enjoy a cup of coffee (or bring about beauty or knowledge, etc.), if other things are equal. Of course, normally, other things are *not* equal, and killing has plenty of negative consequences, some of them maybe even necessarily, e.g., if life is intrinsically valuable. But that does not change the fact that the cup of coffee has *some* weight in the consequentialist calculus that must be counterbalanced. A plausible ethical theory would say that there is no standing reason to bring about such trivial pleasures as that of having a cup of coffee, or that there is such a reason but that that reason is *muted* if acting on it would involve the killing of an innocent person.

Second, this consequentialist promiscuity of moral reasons, which is counterintuitive in itself, also has implications for the comparative wrongness of different killings that clash with common intuitions. Maximizing act-consequentialism implies that, other things being equal, a killing that enables a third party to enjoy a cup of coffee is slightly less seriously wrong than a killing that does not, and a killing that enables thousands to enjoy

a cup of coffee is more-than-slightly less seriously wrong than a killing that does not. Similarly, many of the factors we earlier observed classical utilitarianism deems relevant to the degree of wrongness of killing will reemerge and have the same relevance in any credible form of act-consequentialism. Surely, for example, it is true that a reasonably happy life of eighty years is better – both in itself and for the one who lives it – than an equally happy life of forty years, and that some people bring about more good than others. If so, an act-consequentialist must conclude that it is more seriously wrong to kill a young person than an old person, and more seriously wrong to kill a person who can be expected to continue producing much good than a relatively unproductive person. All this is fundamentally at odds with the belief that all people are equal, which has assumed a place at the very center of modern moral life. If that belief means anything, it means that it is equally wrong to kill a person regardless of whether the trivial side-effects of doing so are more or less valuable, and of how old or useful that person is. In turn, that means that, at least in certain cases, such as those that involve wrongful killings, certain goods and bads do not have any moral weight – and this, act-consequentialism cannot accommodate.

The intuitions behind the widely-held belief in the moral equality of persons are strong, but not beyond reproach. However, it would take a very strong reason for us, or me anyway, to allow the degree of wrongness of wrongful killings to vary with factors that are almost universally seen as irrelevant. Shelly Kagan has offered an argument that, if sound, would provide such a reason, and we shall now briefly discuss that argument.

### 1.6. *The pro tanto reason to promote the good*

Ordinary morality obligates us to make some sacrifices, but also holds that we are at least sometimes allowed to do what would not have the best consequences overall, and that, at other times, we are even required *not* to do what would have the best consequences overall. Kagan refers to people who hold any such position as “moral moderates,” and claims that they are committed to the existence of “a pro tanto reason to promote each individual good”<sup>79</sup> (or, shorter: a pro tanto reason to promote the good). A pro tanto reason to  $\phi$  is an always-present moral reason to  $\phi$ , which translates into a decisive reason to  $\phi$ , and thereby creates a moral requirement to  $\phi$ , unless it is outweighed by a stronger reason.<sup>80</sup> Accordingly, to say that there is a pro tanto reason to promote the good amounts to the claim that the fact that some particular good is attainable in a given choice situation *always* provides *some* ground for acting such that that good is brought about. Hence, in cases in which one is not obligated to promote the good, that is not because the pro tanto reason to promote the good has vanished, but because of some other reason.

Kagan’s argument for the existence of a pro tanto reason to promote the good is an inference to the best explanation. He argues that there are a number of judgments the moral moderate wants to make that are best explained with reference to a standing reason to promote the good, which he accordingly takes to be common ground for moral moderates and consequentialists.

---

<sup>79</sup> Kagan 1989, p. 57.

<sup>80</sup> W. D. Ross, somewhat misleadingly, used the term “prima facie duty” to refer to pro tanto reasons, cf. Ross 1930, pp. 19 f. & 28 f. Note that Kagan subscribes to the theory of moral reason we have discussed earlier, in the context of classical utilitarianism. In that picture, reasons are like forces which pull into different directions and can be added together to form a resultant.

Kagan starts with a group of cases that all involve saving lives. One case in that group is Singer's well-known example of the child drowning in a shallow pond.<sup>81</sup> Here, Kagan claims,

“the moderate certainly recognizes the existence of a reason to promote the good. Now imagine that the cost of saving the child somehow becomes so significant that ordinary morality would no longer require the necessary sacrifice. Surely the moderate does not want to hold that suddenly there is no longer any reason to save the child. [...] The reason to promote the good does not pop into and out of existence as the cost of promoting the good varies: rather there is a standing reason to promote the good, which may or may not be overridden.”<sup>82</sup>

The moderate may concede the point that a reason to promote the good that is there when the cost is low and suddenly disappears when the cost exceeds a certain threshold would be a strange animal indeed. But it is unlikely that a moderate would be ready to admit the existence of a reason to promote the good in the first place. If there is a moral requirement to save the child in Singer's example, that requirement can equally well be explained in terms of a standing reason to save lives, a standing reason to save lives if it can be done at low cost to oneself, or a standing reason to prevent harm. All of these alternative pro tanto reasons do the same job as the pro tanto reason to promote the good, yet none of them commits the moderate to the claim that one has a reason to kill somebody whenever doing so is the only way to bring about some good, no matter how trivial. That I think is a significant advantage, and a good reason to deny that there is a pro tanto reason to promote the good.

---

<sup>81</sup> Cf. Singer 1972.

<sup>82</sup> Kagan 1989, p. 49.

Kagan foresees this objection, and moves on to consider a case that neither involves saving lives nor the prevention of harm:

“Imagine that I am considering giving a copy of Kilgore Trout’s short stories to someone whom I believe might enjoy them. Now it may well be that I am not *required* to give the book to some such worthy individual, but I take it that the moderate does not want to claim that morally there is *no* reason at all for me to give it.”<sup>83</sup>

This is hard to believe. In a sense, that there is no reason at all for me to give the book is exactly what the moderate wants to claim. Much of the *prima facie* appeal of Kagan’s argument goes away once we carefully distinguish between insistent and non-insistent reasons.<sup>84</sup> Insistent reasons are what I have so far simply referred to as moral reasons: considerations that count in favor of the moral obligatoriness of acts. In contrast, non-insistent reasons are “reasons for action that do not mandate action,”<sup>85</sup> and are as such irrelevant for the rightness or wrongness of acts. To use an example by Thomas Nagel, “if I buy a book because it is cheap, that is a reason for buying it, but it did not have to be my reason.”<sup>86</sup>

The moderate may agree that there is a non-insistent reason to promote the good. In Kagan’s example, if the Kilgore Trout fan is my friend, or if I am concerned for her well-being for some other reason, and hence choose to make the reason to promote the good my reason for giving her the book, then that reason can explain and justify my conduct. The fact that she will enjoy the book however is irrelevant to the question of whether I

---

<sup>83</sup> Kagan 1989, p. 52.

<sup>84</sup> I borrow this distinction from Kamm 1996, p. 231, and Vallentyne 2006, p. 24.

<sup>85</sup> Kamm 1996, p. 231.

<sup>86</sup> Kamm 1996, p. 231.



morally ought to give the book. No countervailing considerations are required to prevent the reason to give the book from translating into a moral requirement, as it is not insistent. I might simply lack interest in the Kilgore Trout fan's well-being. If I therefore failed to give her the book, I would do no wrong, even in the absence of any moral reasons not to give the book. Kagan's argument has some appeal because there is in fact some reason to give the book, *if* I make it my concern to bring about the good that can be achieved by doing so. But, if I do not, then I simply do not have a reason to give the book.

Unless one is already committed to some form of consequentialism, Kagan's argument is less likely to convince of the existence of a pro tanto reason to promote the good than to be rejected once the crucial distinction between insistent and non-insistent reasons is made.

### ***1.7. Recasting the relationship between the normative status of acts and value***

It is time to take stock. So far, we have argued that all forms of maximizing act-consequentialism fail to provide a persuasive account of the wrongness of killing, for at least two reasons. First, they cannot accommodate there being goods unfit to even begin to justify killing someone. Second, they offend the almost-universal belief in the moral equality of persons, as they cannot make room for considerations of value to be muted, or otherwise prevent obviously irrelevant factors from altering the degree of wrongness of wrongful killings.

We have introduced maximizing act-consequentialism as the set of ethical theories that includes classical utilitarianism and all ethical theories that can be obtained from classical utilitarianism by substituting an alternative theory of value for value-hedonism. As we noted earlier, there are two more generalizations we can make to expand that set, while staying within the boundaries of consequentialism. One is to move away from act-consequentialism, and tie the normative status of an act to consequences other than the consequences of that particular act and its alternatives. We will discuss this move in a later section. The other is to recast the relationship between the rightness and wrongness of acts on the one hand and the value of their consequences, and the consequences of their alternatives, on the other hand. This recasting can take at least two forms that we shall now very briefly consider.

*Aggregation.* So far, we have implicitly assumed that outcomes are to be ranked according to the total sum of value they contain. Consequentialists and their opponents alike have objected to this way of ranking outcomes because it ignores the separateness of persons and is insensitive to the distribution of well-being. Parfit, for example, has pointed out that, if the overall value of an outcome is the total sum of value it contains, an act that brings into existence a large number of people who will have lives barely worth living ranks higher than an act that brings into existence a small number of people who will have lives of a high quality, other things being equal.<sup>87</sup> This strikes many as false, and is known as the *repugnant conclusion*. A common response is to move from total to average act-consequentialism, which ranks outcomes according to the average net good per person, or per sentient being, they contain. Unfortunately, average act-consequentialism is not free of problems itself. It implies that a world where all people

---

<sup>87</sup> Cf. Parfit 1984, Part 2, Chapter 17.

are truly miserable is made better by adding another person who is slightly less miserable, yet also suffering greatly. Further, our objections concerning the relevance of trivial goods for the justification of killing and the comparative wrongness of killings equally apply to average act-consequentialism. They also apply to distribution-sensitive forms of act-consequentialism that have been devised to accommodate concern for certain kinds of equality, such as *prioritarianism*, which assigns more weight in the consequentialist calculus to the well-being of people who are worse off. If two people are equally well off, yet one is more useful than another, then average and prioritarian act-consequentialism, and the combination of the two views, imply that killing the one is more seriously wrong than killing the other, other things being equal.

*Satisficing*. Maximizing act-consequentialism demands that we always bring about the best results we possibly can. That leaves little room for personal choice, as it requires us to abandon our projects and commitments whenever doing so maximizes the good. Maximizing act-consequentialism is hence often criticized as too demanding, and Williams and others have called it an attack on the integrity of persons, for that reason.<sup>88</sup>

“Consider a manager of a resort hotel who discovers, late one evening, that a car has broken down right outside its premises. In the car are a poor family of four who haven’t the money to rent a cabin [...], but the manager offers them a cabin gratis [...]. [S]he doesn’t go through the complete list of all the empty cabins in order to put them in the best cabin available. She simply goes through the list of cabins till she finds a cabin in good repair that is large enough to suit the family. [...] In such circumstances, optimizing act-consequentialism [...] would presumably hold that the manager should look further for a better room. [...] But I

---

<sup>88</sup> Cf. Williams 1973.

think ordinary morality would regard her actions as benevolent and her choice of a particular room for the family in question as morally acceptable, not wrong.”<sup>89</sup>

In light of cases such as this one, and in order to make act-consequentialism seem less out of line with ordinary morality, Michael Slote proposes that maximization be abandoned in favor of satisficing. Satisficing act-consequentialism sets a threshold for the value of consequences and permits moral agents to perform any act whose consequences meet that threshold. Both the maximizer and the satisficer are committed to the promotion of value, but while only the best is good enough for the former, the latter is sometimes satisfied with less. Satisficing act-consequentialism “makes it permissible to pursue non-optimific personal projects and commitments,”<sup>90</sup> as long as the outcome is good enough. How close it is to ordinary morality seems to depend on the threshold of what is good enough. However, there is reason to be skeptical about the common-sense appeal of satisficing act-consequentialism, regardless of where the threshold is set. For one, the theory does not take into account cost, which somewhat alienates the satisficer’s use of “good enough” from the common meaning of the term. If I can save ten thousand and one people from death at very little cost to myself, then saving ten thousand people is not “good enough,” even if doing so meets some threshold of what is good enough that is cost-insensitive and only looks at the value of consequences. In contrast, if the personal cost of saving others is very great, ordinary morality is likely not to require the sacrifice, even if by not making the sacrifice one falls short of said threshold. Satisficing act-consequentialism sometimes demands too much, and sometimes too little.<sup>91</sup>

---

<sup>89</sup> Slote 1984, p. 149.

<sup>90</sup> Slote 1984, p. 158.

<sup>91</sup> Also, it makes the injustice objection to act-consequentialism much worse. In Mulgan 2001, Tim Mulgan presents a “thought experiment, in which an agent must choose whether to save the lives of ten innocent

Further, satisficing act-consequentialism does little to address our general objections to act-consequentialism. Satisficers hold that killing is morally wrong, when it is wrong, because, and to the extent that, it falls short of some threshold of what is good enough. Any kind of good a killing would bring about, no matter how trivial, has a tendency to elevate that killing above the threshold and thereby make it permissible. The problem with trivial goods hence remains. Satisficing act-consequentialism also does not give us any additional leverage in preventing obviously irrelevant factors from affecting the degree of wrongness of wrongful killings. The more socially useful a person is, the less satisfactory, and hence the more seriously wrong, killing that person would be, other things being equal.

Admittedly, forms of act-consequentialism that recast the relationship between normative status and value have been given a short rift here, and it is hard to rule out the possibility of there being such a theory that avoids our objections. For lack of space, however, what has been said must suffice, and we have to content ourselves with the observation that we have considered many or most of the most popular forms of act-consequentialism, and leave the discussion of more exotic act-consequentialist ethical theories – that may or may not avoid our objections – to others.

---

people by using a sand bag or by killing an innocent person,” (Mulgan 2001, p. 41) and argues that satisficing consequentialism “must allow such an agent to kill. [...] [T]his result is much more counter-intuitive than the fact that Maximizing Consequentialism permits agents to kill in order to produce the best consequences” (Mulgan 2001, p. 41).

### 1.8. *Harm-to-the-victim accounts of the wrongness of killing*

Before wrapping up our discussion of consequentialism, this is a good point to briefly turn our attention to another class of accounts of the wrongness of killing that has received considerable attention in the literature. The accounts in that class are not strictly consequentialist but nevertheless regard certain consequences of wrongful killings as their primary wrong-making feature, which is why much of what we said about consequentialism is relevant here, too.

Act-consequentialism does not allow us to give special importance to the consequences of a killing for the one who is killed, and many have thought that that is one place where it goes wrong. As R. E. Ewin puts it, “the reason why it is wrong to kill somebody has something to do with him, not with his mother or maiden aunt.”<sup>92</sup> *Harm-to-the-victim accounts of the wrongness of killing* are built on that intuition and hold that killing is wrong, when it is wrong, because it harms the one who is killed. James Rachels is one of the philosophers who defend such an account. He writes that, “[i]f we should not kill, it is because in killing we are harming someone. That is the reason killing is wrong.”<sup>93</sup> This immediately raises the question of what that harm consists in. Don Marquis, another proponent of a harm-to-the-victim account, gives the following answer:

“What primarily makes killing wrong is [...] its effect on the victim. The loss of one’s life deprives one of all the experiences, activities, projects, and enjoyments that would otherwise have constituted one’s future. Therefore, killing someone is

---

<sup>92</sup> Ewin 1972, p. 126.

<sup>93</sup> Rachels 1986, p. 6. Also, cf. Glover 1977.

wrong, primarily because killing inflicts (one of) the greatest possible losses on the victim.”<sup>94</sup>

In Clint Eastwood’s words, “it’s a hell of a thing to kill a man – you take away all he’s got and all he’s ever gonna have”<sup>95</sup> – killing is harmful and therefore wrong, because it is the ultimate form of theft. This explanation of the wrongness of killing, which deserves a thorough discussion because of its frequent occurrence in both the literature and the classroom, has some obvious virtues. First, if there are any non-human animals who have futures relevantly similar to ours, Marquis’ account equally applies to them.<sup>96</sup> Secondly, unlike explanations of why killing is wrong that make use of the concept of personhood, some of which we will discuss in subsequent chapters, Marquis’ account straightforwardly applies to infants and young children and condemns killing them for the same reason that it condemns killing adults. Thirdly, it explains why killing is more seriously wrong than most other harmful acts: “Killing is especially wrong, because it deprives the victim of more than perhaps any other crime.”<sup>97</sup> Here, Marquis makes the plausible assumption that the wrongness of a harmful act varies with the amount of harm it does to the victim, if all other things are equal. Taking into account this implicit assumption, we can formulate Marquis’ central claim more clearly: Typically, killing is wrong, when it is wrong, primarily because it harms the person killed by depriving that person of the value of his or her future, and the degree to which a wrongful killing is wrong varies with the amount of harm it does to the victim.

---

<sup>94</sup> Marquis 1989, p. 189.

<sup>95</sup> Clint Eastwood, in the 1992 movie, *Unforgiven*.

<sup>96</sup> On the other hand, Marquis’ account does not apply to human beings who do not have such futures.

<sup>97</sup> Marquis 1989, p. 190.

Marquis' harm-to-the-victim account of the wrongness of killing is based on the *deprivation account of the badness of death*, versions of which are championed by Thomas Nagel and Fred Feldman.<sup>98</sup> According to the deprivation account, death is bad for the one who dies, when it is bad for the one who dies, because it deprives her of the good things that otherwise would have fallen into her life. Not everybody accepts this. Epicurus famously wrote in a letter to Menoeceus that

“death, the most terrifying of ills, is nothing to us, since as long as we exist death is not with us; but when death comes, then we do not exist. It does not then concern either the living or the dead, since for the former it is not, and the latter are no more.”<sup>99</sup>

Epicurus inspired, and gave name to, a line of philosophers, extending to the present day, who claim that, contrary to common belief, death is not bad for the one who dies. Elsewhere, I have argued in support of this claim and developed an account of extrinsic prudential value that further implies that death is almost always good for the one who dies.<sup>100</sup> If death is never a harm, the deprivation account of the badness of death fails, and so does Marquis' account of the wrongness of killing that crucially depends on the deprivation account. But, for the sake of argument, let us assume that common sense is right, and that death is usually a great harm. What else can be said about Marquis' proposal?

Like act-utilitarianism, Marquis's account runs into difficulties when one tries to apply it to cases of involuntary euthanasia. Marquis leaves open the question of just what it is about our futures that, normally, makes it morally wrong to kill us. He does not subscribe

---

<sup>98</sup> Cf. Nagel 1970, and Feldman 1992.

<sup>99</sup> Epicurus 1940, p. 31.

<sup>100</sup> Cf. Ebert 2013.



to any specific theory of well-being. However, as I have argued earlier, virtually every such theory will allow for cases in which a person would benefit from death, yet wants to live. If all the future has in store for a person is pain and despair, and nothing good to compensate, death for that person would not mean the loss of a valuable future like ours, and the primary feature that makes killing wrong is absent. This does not commit Marquis to the claim that nothing is wrong with involuntary euthanasia. He is careful to stress that his account is not to be understood as a complete account of the wrongness of killing, and he concedes that there may be other factors that also contribute to the wrongness of killing. Hence, Marquis can argue that involuntary euthanasia is wrong not because it harms the victim – it does not –, but for some other reason. However, as he does not say, we are left wondering what this other reason might be. I do not wish to speculate. Instead, I want to express a worry about this solution. Whatever it is about involuntary euthanasia that makes it wrong, it seems like Marquis must regard that thing as less significant than the loss of a valuable future which he sees as *the* primary wrong-making feature of killing. If, regardless of whether the future of a person is bleak or bright, there is an equally or nearly equally strong presumption that killing that person against his or her will is seriously wrong, then Marquis must admit that there are at least two “primary” wrong-making features of killing. The result would be a theory that, in my mind, is much less attractive than a theory that relies on just *one* explanation why it is seriously wrong to kill a person in every case in which it is seriously wrong to kill a person. Putting this worry aside, there are two more straightforward objections to Marquis’ account of the wrongness of killing. One concerns scope, the other – which we

have already encountered in the context of consequentialism – concerns degrees of wrongness.

If there is a strong presumption against killing typical adults because they have a valuable future like ours, then there is a strong presumption against killing fetuses for the same reason.<sup>101</sup> Marquis is happy to accept this conclusion. In fact, establishing it is the main goal of his argument. But, if abortion is within the scope of his argument, why not contraception, too? After all, contraception patently prevents valuable futures like ours. Marquis responds to this objection by pointing out that “[n]othing at all is denied such a future by contraception [...]”<sup>102</sup> His point here is that there is no non-arbitrary way to identify the thing that is deprived of a valuable future by contraception. Assigning the harm to some sperm or other is arbitrary because one could as well choose some ovum or other, and assigning the harm to some ovum or other is arbitrary because one could as well choose some sperm or other.<sup>103</sup> Making a sperm and an ovum each a separate subject of harm would give us one future too many. The one candidate for a subject of harm we are left with is a set of a sperm and an ovum. Marquis argues that this does not work either because there are millions of such sets and there is no non-arbitrary way to pick out one of them. While it is true that there are millions of such sets, there in fact *is* a non-arbitrary way to pick out one of them. In each case in which the use of contraception

---

<sup>101</sup> “The future of a standard fetus includes a set of experiences, projects, activities, and such which are identical with the futures of adult human beings and are identical with the futures of young children” (Marquis 1989, p. 192).

<sup>102</sup> Marquis 1989, p. 201.

<sup>103</sup> David Benatar interestingly observes that Marquis’ rejection of some sperm or other as the victim of contraception seems to depend on a biological contingency that can hardly be morally relevant. If human sperm “contained all the genetic material and required the ovum only for nutrition, then the relationship between sperm and ovum would be relevantly like the current relationship between zygote and uterus” (Benatar 2006, p. 158), and some sperm or the other could be identified as the victim of contraception. “Thus the moral issue, on Professor Marquis’ view, rests on whether sperm is haploid or diploid” (Benatar 2006, p. 159).

successfully prevents a pregnancy, there is one identifiable set of a sperm and an ovum that would have combined and formed a zygote if contraception had not been used (or two, if a twin pregnancy is being prevented, etc.). Hence, if an instance of contraception-use prevents a future like ours, it deprives the sperm and the ovum that would have formed a zygote of that future and hence falls into the same moral category as abortion and killing an innocent adult human being. This I regard as absurd.

As noted earlier, Marquis' position implies that, other things being equal, a killing that is wrong primarily because it deprives the victim of the value of his or her future is more seriously wrong than another killing that is wrong for the same reason but deprives the victim of a less valuable future. This has profoundly counterintuitive implications. If harm played the central role in explaining the wrongness of killing that Marquis thinks it does, then, other things being equal, an abortion would be more seriously wrong than killing a child or a twenty-year-old, as the fetus would be deprived of more. Similarly, it would generally be more seriously wrong to kill always-jolly optimists than people disposed to depression, and less seriously wrong to kill members of an oppressed minority than the privileged. It would also not be a serious moral matter at all to kill somebody a minute before he otherwise would have been run over by a bus or struck by a lightning, as the effect on the overall value of his life would be negligible.

The contraception objection does not straightforwardly generalize to other harm-to-the-victim accounts of the wrongness of killing. Rather, I think it shows the inadequacy of Marquis' theory of harm. One could defend the position that only those who have been sentient, are sentient, or will be sentient can have interests and be harmed. Successful contraception-use prevents a sperm and an ovum from forming a zygote and hence a

future like ours, but does not harm the thing that is now that sperm and that ovum taken together, as that thing in fact has never been, is not, and will never be sentient. Similarly, this position would imply, plausibly I think, that early abortion does not victimize the early fetus, because an early fetus that is killed while still being an early fetus has never been, is not, and will never be sentient, and hence cannot be harmed.<sup>104</sup> One could also argue for a theory of harm that explains the harm of death in terms of frustrated future-directed desires.<sup>105</sup> If death harms the one who dies, if, and only if, the one who dies has desires about his own future, which ground a desire to continue to live, then the primary wrong-making feature of killing is absent in cases of contraception, and cases of abortion for that matter.

The objections concerning involuntary euthanasia and comparative wrongness, however, readily apply to all harm-to-the-victim accounts of the wrongness of killing that allow for death sometimes being a benefit and recognize that the harm of death varies from person to person, and from circumstance to circumstance, respectively. Further, such accounts seem to give the wrong *kind* of explanation of why killing is wrong. While it might be true that killing usually deprives the victim of goods from which she would have benefited otherwise, an explanation of why killing is wrong that primarily appeals to that deprivation leaves out something important. Worse, there is a vague but powerful sense that it misses the main point. Harm-to-the-victim accounts are built on an intuition that demands an intimate link between the wrongness of a killing and its victim. Consequentialism cannot meet that demand, as its explanation of the wrongness of killing

---

<sup>104</sup> Elizabeth Harman argues for a position along these lines in Harman 1999. She spells that position out in terms of moral status: “An early fetus that will become a person has some moral status. An early fetus that will die while it is still an early fetus has no moral status” (Harman 1999, p. 311).

<sup>105</sup> Cf., e.g., Singer 1979b & 2011.

does not essentially involve the one who is killed. I will argue for this point, which I think further discredits consequentialism, in the next section.<sup>106</sup> Harm-to-the-victim accounts fare better insofar as their explanation of why killing is wrong *does* contain an essential reference to the victim: Killing is wrong because it makes *the victim* worse-off. Yet, even if a bit less so than the consequentialist explanation, that explanation still seems superficial. It focuses our attention on what killing takes away from the victim, rather than the seemingly more important fact that the victim is annihilated, almost as if the victim were relevant not as such, but only as a container or receptacle for what really matters.<sup>107</sup> The suspicion that this focus is misplaced is strengthened by considering cases of wrongful killing where the foreclosed futures would not have been worth living for. If the victim in such a case is wronged by being killed no less than, and in the same way as, somebody who is killed in the middle of a life that would have continued to be rewarding, then it seems we must recognize that the wrongness of killing does not primarily lie in its being the ultimate form of theft. Killing is a different kind of wrong than stealing somebody's car, rather than a more severe instance of the same kind of wrong. Killing is more personal, and we hence expect the explanation of its wrongness to have the one who is killed *herself* at the center, not contingent facts about her well-being.

---

<sup>106</sup> “The problem for [...] utilitarianism is that an explanation of the wrongness of killing in terms of negative side effects, diminution in pleasure, and frustration of desires just does not seem to provide the intimate link to the destruction of individual, valuable, autonomous persons that [...] our intuitions demand” (Frey 1984, p. 15).

<sup>107</sup> “In killing, the main point is not that something is taken away from someone, but that *the someone* is taken away” (Chappell 2004, p. 111).

### 1.9. *Rule-consequentialism*

Returning from our brief excursion to consequentialism proper, there is one more issue on which consequentialists disagree that we have not yet talked about. We shall remedy that at once. At the beginning of this chapter, we imagined a sheriff who is confronted with the choice of whether or not to frame an innocent scapegoat in order to prevent a series of lynchings. In terms of well-being, one death seems preferable to many deaths, yet justice – or, anyway, our sense of it – requires that the sheriff does not allow the innocent scapegoat to be executed. Unconvinced by the way Sprigge, Smart, and Hare respond to this thought experiment, and other act-consequentialist strategies to avoid the conclusion that the sheriff ought to frame the innocent scapegoat, a number of consequentialists have moved away from act-consequentialism, and instead endorse some form or other of *indirect consequentialism*. A consequentialist ethical theory is indirect, if it ties the normative status of an act to the consequences of something other than, yet related to, that particular act and its alternatives. The classical and perhaps most prominent example is *rule-consequentialism*, with which we shall end our discussion of consequentialism.<sup>108</sup>

According to rule-consequentialism, an act is right, if, and only if, it is performed in accordance with a set of rules which is chosen solely on the basis of the value of its consequences, relative to the value of the consequences of alternative sets of rules. This phrasing is both vague and too general to be useful, and hence needs to be unpacked. For example, strictly speaking, a rule – being an abstract entity – does not have any

---

<sup>108</sup> Other, less prominent examples are motive and virtue consequentialism, cf., e.g., Adam 1976, and Sverdlik 2011. Much of the subsequent discussion is relevant to these forms of indirect consequentialism as well.

consequences. Hence, the question arises, what is meant by “the consequences of a set of rules”? Further, how exactly do the consequences of rules determine which rules are relevant, and, in turn, which acts are right, and which are wrong? A simple answer to these questions is that an act is morally right, if, and only if, it is performed in accordance with a set of moral rules, *universal conformity* with which would have the best consequences. This answer, however, has well-known problems.

No course of action would result in a greater amount of good than everybody always doing, of all the things they can do in whatever situations they are in, what has the best consequences. In other words, most good would be produced if everybody only performed acts that accord with the rule, “Among the acts available to you, perform the act, or one of those acts, whose consequences are at least as good as those of all the other acts.” Accordingly, universal conformity with the set of rules that contains that and no other rule would have the best consequences. There may be other sets of rules with the same property, with rules likely to be very complicated and allowing for many exceptions, but all of these sets have in common that, as part of the simple rule-consequentialism we just formulated, they require precisely the same acts maximizing act-consequentialism does. While intensionally distinct, simple rule-consequentialism is extensionally equivalent to, and hence seems to collapse into, maximizing act-consequentialism. Yet, even if that collapse can somehow be avoided, the criticism continues, that does not make for a plausible rule-consequentialism either, as then there will be situations in which rule-consequentialism requires that the rules be followed even though doing so means bringing about less good than one could. That, in turn, seems to

be at odds with the intuition that drives consequentialist thinking, that there is a fundamental duty to promote the good.

Partly in an attempt to meet the criticism levelled at simple rule-consequentialism, contemporary theorists have developed more sophisticated forms of rule-consequentialism. Brad Hooker, for example, argues that

“[a]n act is wrong if it is forbidden by the code of rules whose internalization by the overwhelming majority of everyone everywhere in each new generation has maximum expected value in terms of well-being (with some priority for the worst off). The calculation of a code’s expected value includes all costs of getting the code internalized.”<sup>109</sup>

Richard B. Brandt, another contemporary proponent of rule-consequentialism, argues that

“[a]n act is right if and only if it conforms with that learnable set of rules the recognition of which as morally binding – roughly at the time of the act – by everyone in the society of the agent, except for the retention by individuals of already formed and decided moral convictions, would maximize intrinsic value.”<sup>110</sup>

Both Hooker’s and Brandt’s account do not collapse into act-consequentialism, as the code or set of rules consisting only of the one rule, “Maximize the good,” does not meet their respective requirements. The cost of internalizing that rule, just like the cost of getting people to recognize it as morally binding, would be very high, as people are naturally biased towards themselves and others they love, or care about. “[T]he time, energy, attention, and psychological conflict that would be needed to get people to

---

<sup>109</sup> Hooker 2000, p. 32.

<sup>110</sup> Brandt 2003, pp. 232 f.



internalize an overriding impartial altruism would be immense.”<sup>111</sup> Further, people would lose trust in each other, if they came to expect, as they would, that others will kill, rob, or lie whenever doing so would maximize the good. The consequences of that loss of trust would be disastrous. Finally, there is no reason to expect, and much reason to doubt, that the internalization or recognition of said rule, and no other moral rule, would maximize value. We are flesh-and-blood human beings, and not archangels à la Hare. Often, likely more often than not, we would not do what leads to the most good if we tried doing so, sincerely and to the best of our abilities. Our intelligence, knowledge, and willpower are limited, and we are almost certain to do better if we internalize, or recognize, not the rule, “Maximize the good,” but a set of rules that has been devised with due regard to our limitations.

Suppose there was a widely internalized, or recognized, rule that allows people in situations relevantly similar to the one the sheriff is in to sacrifice an innocent person if it is believed that doing so will maximize the good, by preventing the deaths of many. Given our cognitive, motivational, and epistemological limitations, there would be a significant risk of that rule being misused. There would be situations in which people mistakenly believe that the rule applies, whereas in fact it does not. Also, innocent people would have to live in constant fear of somebody else rightly or mistakenly coming to believe that he or she is morally justified, or required, by the rule, to kill an innocent person, as that person might be one of them. Weighing these costs against the unlikely benefits of the rule, realized only on extremely rare occasions, Hooker and Brandt can convincingly conclude that, according to their accounts, framing and executing an innocent scapegoat in order to prevent a series of lynchings is morally impermissible,

---

<sup>111</sup> Hooker 2000, p. 95.

even if doing so would result in more good than any alternative course of action – which is in line with common moral intuition.<sup>112</sup> In general, rule-consequentialism requires us not to calculate the value of the consequences of this killing or that killing, but the value of the consequences of the internalization or recognition of rules that apply to acts of killing. It seems practically impossible to identify, and write down, a complete set of rules that fulfills Hooker’s or Brandt’s conditions, and then investigate what these rules imply for various acts of killing.<sup>113</sup> For the sake of argument, however, let us assume that there is a rule-consequentialist ethical theory that does not collapse into act-consequentialism and forbids all killings of innocent people that are commonly seen as clearly immoral. How does such a theory fare as an account of the wrongness of killing? Rule-consequentialism will not seem like an improvement to those who are satisfied by the act-consequentialist response to *Innocent Scapegoat* offered by Sprigge, Smart and Hare. That is, because they would rather have us abandon our moral intuitions in cases far removed from ordinary experience, and instead hold on to a direct form of consequentialism, which is more closely aligned with the fundamental consequentialist

---

<sup>112</sup> It should be noted that this does not commit Hooker and Brandt to the claim that it is *never* right to kill an innocent person. While those who believe that there are certain kinds of act that are always wrong might find that claim innocuous (e.g., Anscombe 1958, p. 16, and Thomson 1990, p. 168), I think most people reject that kind of moral absolutism and would intuitively be willing to make an exception if enough is at stake. Both Hooker and Brandt regard it as one of the virtues of their respective theories that they can accommodate that intuition. They argue that a “prevent disaster” rule that overrides all other rules must necessarily be included if a code or set of rules is to be ideal. Brandt formulates that rule as follows: “If following an otherwise optimific set of rules would cause a very large utility loss or forgo a very large utility gain on this occasion, the agent is permitted to perform an act that is (close to) utility-maximizing on the occasion” (Brandt 1992, p. 151). The question immediately arises, how bad do the consequences of following some important rule, such as the one against killing an innocent person, have to be before the “prevent disaster” rule applies? Hooker and Brandt never say exactly, but they give a few examples that suggest that what they have in mind when talking about disasters are large-scale catastrophes: “the blowing up of New York City” (Brandt 1992, p. 87); “the ending of our species” (Hooker 2000, p. 135). I hence made the assumption that a series of lynchings does not qualify, and rule-consequentialism accordingly fares better than act-consequentialism from the perspective of those who insist that an adequate ethical theory must yield the judgement that the sheriff ought not frame and execute an innocent scapegoat.

<sup>113</sup> That is a problem, not only for rule-consequentialism, but for consequentialism in general. The few actual consequentialist calculations, with numbers, that I have come across were overly simple at best, and totally unconvincing at worst.

intuition that the good ought to be promoted. On the other hand, those who insist that a credible ethical theory should deliver verdicts that accord with our considered convictions even in very unlikely circumstances have no reason to be too enthusiastic about rule-consequentialism either. It seems plain that there are cases, such as rare instances of *Innocent Scapegoat*, in which the verdicts of rule-consequentialism are more agreeable to the common moral consciousness than those of (at least the more common forms of) act-consequentialism, but there are also cases in which rule-consequentialism seems to have no advantage. Imagine the day comes when science or natural selection has freed human beings from their manifold biases, particularly the one towards themselves, thereby directing their motivational set more towards the general good, and furthermore has increased their cognitive abilities drastically. These future human beings are not quite Harean archangels, but they are close enough to perform act-consequentialist calculations at a high degree of reliability, and getting them to internalize or recognize a moral code that consists only of one rule, asking them to always do what maximizes the good, involves little cost. In such a possible, future society, the internalization or recognition of the rule, "Maximize the good," might well have maximum expected value, and even sophisticated forms of rule-consequentialism, such as those developed by Hooker and Brandt, collapse into act-consequentialism. That a situation arises in a world populated by people like us, in which framing an innocent scapegoat would maximize the good, is as much a possibility as such a situation arising in world populated by people that do not share our limitations. Rule-consequentialism yields the exact same verdict in the latter case that the rule-consequentialist regards as fatal to act-consequentialism in the former case. How much of an improvement over act-

consequentialism is rule-consequentialism, really, if it merely reduces the number of possible situations in which consequentialism conflicts with our considered moral judgement, which is very small anyway, but does not bring it down to zero – at the cost of complicating a theory whose appeal owes much to its simplicity?

Whatever the answer to that question, the thought experiment that prompted it draws our attention to an important characteristic that rule-consequentialist and act-consequentialist accounts of the wrongness of killing have in common. Both are contingent on empirical facts in a way that is problematic. A credible account of the wrongness of killing not only must delineate wrongful and justified killings in accordance with our considered moral convictions, but also deliver the right verdicts for the right reasons. I have argued in previous sections that act-consequentialism cannot provide a plausible explanation of why killing is wrong, when it is wrong. The same I think is true of rule-consequentialism, to an even greater extent.

Once again recall the case of Susan who would be better off dead, yet wants to live. Classical utilitarianism explains the wrongness of killing Susan mainly in terms of consequences for third parties, such as the grief of those left behind and expectation effects. That misses the main point, which we expect to be more intimately connected with Susan, and what killing does to her. If one evaluates rules solely in terms of well-being, as for example Hooker does, then the justification of a rule against killing people relevantly like Susan too depends on consequences for third parties. According to Hooker's rule-consequentialism, killing Susan is morally wrong because acts of that kind make people other than the victims worse off and a rule allowing acts of that kind is therefore very unlikely to be found in a code of rules whose internalization has maximum

expected value. If humanity knew neither grief nor fear of death, a rule requiring that people whose lives are not worth living be killed, maybe with an exception for exceptionally useful people, might well fulfill Hooker's conditions. The fact that Susan would be better off dead, yet wants to live, supports such a rule, and hence tends to justify killing Susan. That is implausible. Said fact should on the contrary play an essential role in explaining why killing Susan is wrong. Surely, a plausible explanation of why it is wrong to kill Susan must focus on Susan, rather than the effect of her death on others, or the consequences of completely unrelated killings.

Like in the case of act-consequentialism, the problem is neither restricted to the issue of involuntary euthanasia, nor to theories that are based on a welfarist conception of the good. The problem is more general, and has its roots in the fact that every intrinsic good or bad that would result from the internalization or recognition of a rule is as relevant to the consequentialist evaluation of that rule, and therefore indirectly to the normative status of individual acts, as any other. Consider, to again take up an earlier example, a rule that permits that people be fed to a group of lions in a public arena for the entertainment of the masses. That rule should be a non-starter, as mere amusement does nothing to justify the brutal killing of innocent people. Yet, in a rule-consequentialist framework, it is not, as the intrinsic good of amusement lends initial support to any rule that permits acts fit for bringing it about. In order to be able to explain the wrongness of murderous mass entertainment, proponents of rule-consequentialism hence must build a case for the claim that the optimal set of rules contains a rule prohibiting such entertainment, despite the fact that it brings a considerable amount of joy to a very large number of people at the relatively small cost of a few lives. In support of this claim, the

rule-consequentialist can argue, for example, that the optimal set of rules also contains a rule urging providers of entertainment to find equally or more effective means of entertainment that harm nobody. That, however, I dare claim, will strike everybody who does not already believe that consequentialism is true as wholly irrelevant to the question of why killing a *particular* person, or group of people, for the entertainment of others is wrong.

In general, according to rule-consequentialism, what ultimately explains the wrongness of a particular wrongful killing is the fact that no set of rules that would allow such an act has a higher internalization or recognition value than all other possible sets of rules. This explanation is even further removed from the victim than that of act-consequentialism, and hence likewise unfit to accommodate the common intuition about the wrongness of killing that demands an intimate link to the one who is killed.<sup>114</sup> By definition, consequentialism, in both its direct and indirect forms, cannot provide an explanation of the wrongness of a particular act of killing that contains an essential reference to the victim.<sup>115</sup> That is because all forms of consequentialism are built on the claims that the normative status of acts is fully determined by the value of relevant consequences, and that there is no essential reference to individuals in the description of valuable consequences. A fully explicit, consequentialist explanation of why killing a particular person is wrong hence cannot but fail to give adequate prominence to that person, and instead will talk about how the world as a whole will stand as a result of the killing, or of

---

<sup>114</sup> R. G. Frey and Timothy Chappell have described and discussed that intuition at some length in Frey 1984, pp. 15 ff. and Chappell 2004, pp. 111 f., respectively.

<sup>115</sup> “[A]ny view of rightness and wrongness that links them to consequences, whether directly or indirectly, wholly or partly, moves us away from the person-centered exercise of, so to speak, looking at the autonomous person killed to the person-neutral exercise of trying to determine whether the world is a better or worse place, with more or less net pleasure or desire-satisfaction or whatever, as a result of the killing [or the internalization or recognition of a set of rules that allows the killing]” (Frey 1984, p. 17).

the internalization or recognition of a code or set of rules that is not in conflict with the killing.

This concludes our discussion of consequentialism. None of what has been said outright disqualifies consequentialism as an account of the wrongness of killing. Rather, we have accumulated a number of considerations that speak against consequentialism, with varying degrees of strength. That justifies a rejection of consequentialism only if we can find an account that is better supported by reason. In the next chapter, we discuss a promising candidate that more intimately connects the wrongness of killing a person with that person, and hence has appeal in at least one way in which consequentialism has not.

## Chapter 2: Killing as wrong because of what it is

Virtually everybody believes that there are certain things that may not be done to people. Most importantly, innocent people may not be killed, regardless of whether or not it serves the greater good (maybe unless the greater good consists in preventing a very great evil). When a person is wrongfully killed, the extent to which the killing is wrong does not vary with the harm the victim suffers by being killed, or the aggregate value of all consequences of the killing. Other things being equal, it is no more or less seriously wrong to kill an old person than to kill a young person, or to kill a loved and highly productive member of society than to kill a person who lives in seclusion from society and is unlikely to make any significant contributions to the betterment of others. People have rights, and they are each other's moral equals.<sup>116</sup> As we have noted to the discredit of act-consequentialist and harm-to-the-victim accounts of the wrongness of killing in the previous chapter, both families of theories are fundamentally at odds with these common convictions.

Curiously, non-human animals are usually not thought of in the same way. They are considered neither inviolable, nor our moral equals, nor even each other's moral equals for that matter. Human beings kill more than two thousand animals for food per second, not including fish and other sea animals.<sup>117</sup> The benefits we derive from that practice are

---

<sup>116</sup> I use the term "rights" merely as a stand-in. The purpose of this and the following paragraphs is to sketch the general idea that morality is divided into two parts, a "morality of respect" and a "morality of interests." That idea underlies a variety of moral theories, some of which spell out the "morality of respect" in terms of rights, while others talk about moral side constraints, or respect for the dignity or intrinsic worth of persons. I will discuss some of these theories explicitly.

<sup>117</sup> The latest estimates can be found on the statistics page of the Food and Agriculture Organization of the United Nations, URL = <<http://faostat3.fao.org/>> [accessed on March 4, 2016].



largely trivial, as most people could lead equally healthy lives without eating meat, and yet the mass killing of animals for food persists without much controversy.<sup>118</sup> The relatively little controversy there is almost exclusively revolves around the immense amount of suffering imposed on animals in factory farms, whereas the killing of animals for food, as such, is rarely questioned. In contrast, it would be unthinkable to inflict death on a human being for the purpose of deriving benefits comparable to those we derive from eating meat, regardless of whatever method of killing we may choose. The strength of these widely shared intuitions remains undiminished even if the benefits are great, rather than minor. While I cannot imagine that anybody I know would regard it as morally permissible to conduct lethal, or potentially lethal, medical experiments on human beings, especially if done without their consent, such experiments on non-human animals are readily and frequently justified by appealing to the interests of those who stand to benefit. Non-human animals – maybe with the exception of very few animals with humanlike cognitive abilities, such as chimpanzees and dolphins – may be sacrificed for the greater good. Unlike us, they do not have rights, or so the conventional wisdom goes. They are neither members of our community of moral equals, nor typically seen as each other's moral equals, in that, for example, it is widely assumed to be less seriously wrong to kill a mouse than to kill a dog, and to kill an old and fragile dog than to kill a dog who is young and healthy.

Common conscience seems to call for a moral theory that consists of at least two parts: one for beings like you and me who are thought to have a special moral status, and whom

---

<sup>118</sup> “It is the position of the American Dietetic Association that appropriately planned vegetarian diets, including total vegetarian or vegan diets, are healthful, nutritionally adequate, and may provide health benefits in the prevention and treatment of certain diseases. Well-planned vegetarian diets are appropriate for individuals during all stages of the life cycle, including pregnancy, lactation, infancy, childhood, and adolescence, and for athletes” (Craig & Mangels 2009, p. 1266).

it is hence especially wrong to kill, and one for most or all non-human animals, who are thought to lack that status. In his book, *Anarchy, State, and Utopia*, Robert Nozick crudely formulates such a theory, which he dubs *utilitarianism for animals, Kantianism for people*. According to this theory, “[h]uman beings may not be used or sacrificed for the benefit of others [...] [whereas] animals may be used or sacrificed for the benefit of other people or animals [but] *only if* those benefits are greater than the loss inflicted.”<sup>119</sup> This I think very well captures how people ordinarily think about human beings and other animals, which one can see reflected in legal systems across the globe, and yet Nozick stops short of endorsing *utilitarianism for animals, Kantianism for people* because he thinks that, “[e]ven for animals, utilitarianism won’t do as the whole story [...]”<sup>120</sup> One reason, among others, is the possibility of utility monsters.<sup>121</sup> Yet, he nevertheless believes, as most people do, that morality ought to draw a line between human beings and other animals: Human beings have rights, in particular the right not to be killed, other animals do not. This calls for an explanation. If we are justified in treating human beings differently from how we treat other animals, then there must be a morally relevant difference between the members of the two groups. In his review of Tom Regan’s *The Case for Animal Rights*, Nozick suggests that mere species membership might be that morally relevant difference:<sup>122</sup>

“[P]erhaps it will turn out that the bare species characteristic of simply being human [...] will command special respect only from other humans – this is an instance of the general principle that the members of any species may legitimately

---

<sup>119</sup> Nozick 1974, p. 39.

<sup>120</sup> Nozick 1974, p. 42.

<sup>121</sup> Cf. Nozick 1974, p. 41.

<sup>122</sup> Regan 1983.

give their fellows more weight than they give members of other species (or at least more weight than a neutral view would grant them). Lions, too, if they were moral agents, could not then be criticized for putting other lions first.”<sup>123</sup>

This seems to be little more than a not-very-well-disguised appeal to prejudice, which becomes apparent when we restate Nozick’s proposal, so as to make it about race, white people, and black people, rather than species, humans, and lions.<sup>124</sup> The result is an apology of racism, which nobody should accept. Members of any race may *not* give their fellows more weight than they give members of other races, and black people certainly *can* be criticized for putting other black people first.<sup>125</sup> Speciesism has been widely denounced as a form of wrongful discrimination in the philosophical literature, and I shall not rehearse the arguments here.<sup>126</sup>

Most moral philosophers today accept that speciesism cannot be defended. But there is a more interesting way to draw a line between humanity and the rest of the animal kingdom. Philosophical proponents of the common view I just sketched typically agree that mere membership in the human species is morally irrelevant, and cannot justify

---

<sup>123</sup> Nozick 1983, p. 29.

<sup>124</sup> This test of Nozick’s “general principle” was first proposed by James Rachels, cf. Rachels 1990, pp. 183 f., and Rachels 2007, p. 22. It is also employed in Cavalieri 2001, cf. p. 80.

<sup>125</sup> I suspect the limited appeal Nozick’s proposal has derives from the fact that it reminds readers of the case of family. Just like it is morally legitimate to attach more weight to the interests of members of one’s family than to the interest of strangers, it might be argued, one may attach more weight to the interests of members of one’s species than to the interest of members of other species. The problem with this is that we are members of many groups, for most of which the argument is not plausible. We are members of families, races, religious groups, districts, nations, species, etc. “If the argument works for both the narrower circle of family and friends and the wider sphere of the species, it should also work for the middle case: race. [...] Conversely, if the argument doesn’t show race to be a morally relevant boundary, how can it show that species is?” (Singer 2012, p. 186.) Also, notice that appeal to species membership as a special relation gives only those who are so specially related a reason to attach extra weight to the interests of each other. If there were moral agents who are not human beings, they would be justified in treating human beings in ways we treat non-human animals (cf. McMahan 2005, pp. 359 ff.). This stands in conflict with the common understanding of rights as giving rise to agent-neutral reasons.

<sup>126</sup> Cf., e.g., Singer 1975, Chapter 2, Cavalieri 2001, Chapter 4, McMahan 2002, Chapter 3, and Singer 2011, Chapter 3.

attributing greater moral significance to what happens to human beings than to what happens to members of other species. They argue, however, that there is a special-status-conferring property, or a set of such properties, which is contingently bound up with membership in the species *homo sapiens*, and not shared by any other animals (maybe with very few exceptions). This view is immensely important, as it is, in my opinion, the best articulation and defense of folk morality, and continues to inform policy decisions everywhere. We will discuss it in detail in the next chapter.

For the purpose of this chapter, we will assume that there is no intrinsic difference between *all* human beings and *all* other animals that could morally justify applying one set of moral principles to one and a different set of moral principles to the other. We will discuss a prominent view that, in agreement with that assumption, draws a line elsewhere instead, between persons and non-persons, where the term “person” is used not in the ordinary sense of “human being,” but in the technical, Lockean sense often employed in philosophy. This view comes in many variations. In recent times, however, one articulation of the view stands out, in terms of both its rigor and careful argumentation, and its impact on contemporary moral philosophy. It constitutes the core of Jeff McMahan’s 2012 book, *The Ethics of Killing*, and is of particular interest in the context of this dissertation because, as the title of the book indicates, the focus is on the act of killing.

### 2.1. McMahan's two-tiered account of the wrongness of killing

Following Warren Quinn, McMahan distinguishes between two realms of morality. One he calls the *morality of respect*, the other he calls the *morality of interest*.<sup>127</sup> All of our obligations towards non-persons, he says, fall into the latter realm, which is concerned with how our actions affect the well-being of others. In contrast, when dealing with persons, we not only ought to consider their well-being, but must also recognize certain deontological constraints that arise from the worth that comes with personhood. Persons are “mature agents on an equal moral footing with ourselves [...],”<sup>128</sup> and they command respect, in virtue of being persons. Killing a person, McMahan argues, is wrong, when it is wrong, primarily because it involves a failure to respect that person and his or her moral worth.<sup>129</sup> The extent to which killing a person is wrong does not vary with how bad death is for the victim, and the wrongful killing of one person hence is normally as seriously wrong as the wrongful killing of any other person. This reflects the status of persons as moral equals with equal worth who are equally deserving of respect. Non-persons are not worthy of the respect that is due to persons, and the ethics of killing non-person hence is wholly governed by the morality of interests. For them, a harm-to-the-victim-account, in form that slightly differs from the ones we discussed in the previous chapter, is taken to be adequate. As much of what we said about explanations of the wrongness of killing in terms of the badness of death also applies to McMahan's ethics of killing non-persons, we will start our discussion of his theory with this, its first tier. We

---

<sup>127</sup> Quinn uses the term “morality of humanity” instead of “morality of interests,” cf. Quinn 1984.

<sup>128</sup> Quinn 1984, p. 49.

<sup>129</sup> This makes room for the possibility that, sometimes, killing a person is not morally wrong, namely, when killing a person is not a failure to respect that person's worth. McMahan argues that that possibility is realized in certain cases of euthanasia, cf. McMahan 2002, Chapter 5.

will then move on to consider problems that arise from the division of those whom it is sometimes wrong to kill into persons and non-persons, and the moral importance McMahan wishes to attach to it. Finally, we will have a closer look at the second tier, and ask whether it provides a plausible explanation of why it is normally wrong to kill persons.

## 2.2. *The first tier: the time-relative interest account of the wrongness of killing*

In the absence of justifying reasons, it is morally wrong to kill a sentient animal.<sup>130</sup> That seems to be true for animals who are persons, such as you and me, just as it is true for animals who are not. While McMahan argues that sentient non-persons do not have the kind of moral worth that commands respect, he believes, as most people do, that they nevertheless matter, in a way that stones and trees do not, and he thinks that the best way to account for the wrongness of killing them is in terms of the harm that killing does to them. The first step in developing such an account is to find an account of the badness of death that says what makes death harmful, and how we are to determine the severity of the harm of death.

Earlier, we already got to know the deprivation account of the badness of death, which has been advocated most skillfully by Thomas Nagel and Fred Feldman.<sup>131</sup> Let us very

---

<sup>130</sup> Not everybody agrees with this. Immanuel Kant is a notable exception. He thought that non-human animals cannot be wronged in any way, and that any moral obligations we may have with respect to non-human animals must arise from moral obligations towards moral agents. For example, he (in)famously argued that animal cruelty is immoral not because we ought to be concerned for the well-being of non-human animals, but because animal cruelty might lead to cruelty against human beings. As McMahan notes, “on one plausible interpretation, [Kant] held that the morality of respect is in fact coextensive with the *whole* of morality” (McMahan 2002, p. 246).

<sup>131</sup> Cf. Nagel 1970, and Feldman 1992.

briefly recall how this account determines the badness for a sentient being (B), of dying at some given point in time ( $t$ ). First, we calculate the prudential value of B's life as a whole if it ends at  $t$ . That value ( $V$ ) is the sum over all that has ever been good for B and all that has ever been bad for B.<sup>132</sup> Second, we consider the closest possible world ( $W^*$ ) in which B lives beyond  $t$ , and calculate the prudential value of B's life as a whole in  $W^*$  ( $V^*$ ). If  $V^*$  minus  $V$  is positive, a greater balance of good over bad would have fallen into B's life if B had not died at  $t$ . Death at  $t$  then was bad for B because it deprived B of prudential value, and the larger  $V^*$  minus  $V$ , the worse it was. Conversely, if  $V^*$  minus  $V$  is negative, it was good for B to die at  $t$ , and the larger the modulus of  $V^*$  minus  $V$ , the better it was. This can be generalized in that we can say that the value of an event other than death for B too is given by the difference between the prudential value of B's life as a whole, and the prudential value B's life would have had if the event had not occurred.<sup>133</sup>

The deprivation account implies that it is a greater misfortune to die earlier rather than later, other things being equal, because a greater amount of life worth living is foreclosed. For the most part, that seems plausible. While we feel sorry for people who die old, that feeling is usually stronger if the person who dies is in the prime of his or her life. If we move towards the earlier stages of human life, however, the deprivation account loses much of its appeal. In the previous chapter, when we discussed the deprivation account in conjunction with the claim that the wrongness of killing varies with the badness of death

---

<sup>132</sup> Following McMahan, we will not assume any particular theory of well-being. We will assume, however, that the prudential value of a life as a whole varies across individuals and circumstances. This is in line with popular theories of well-being, such as welfare hedonism, and desire-satisfaction theories. Other things being equal, a life scarred by frustrated desires and much pain is worse for the one who lives it than a life filled with desire-satisfaction and joy.

<sup>133</sup> Cf. Bradley 2008, p. 292.

for the victim, we brought out a number of hard-to-swallow implications. If, other things being equal, it is more seriously wrong to kill those who are thereby deprived of a greater balance of good over bad, then it is generally more seriously wrong to kill a fetus than an infant, and more seriously wrong to kill an infant than a twenty-year-old. This almost none of us believe, and I think that is not only because of our belief in human equality but also because almost all of us reject the underlying claims about death's badness. Dying seems worse for a twenty-year-old than for an infant, and worse for an infant than for a fetus.

McMahan believes that our intuitions about the comparative badness of death for human beings at different stages of life can be defended. He argues, against the deprivation account, that the extent to which we are harmed or benefited by death, or some other event, does not solely depend on the net prudential value we lose or gain. There is another factor, which is easily missed because it often is inconsequential. In the case of the death of immature human beings, however, it is crucially important. McMahan invites us to consider the following thought experiment:

*"The Cure.* Imagine that you are twenty years old and are diagnosed with a disease that, if untreated, invariably causes death (though not pain or disability) within five years. There is a treatment that reliably cures the disease but also, as a side effect, causes total retrograde amnesia and radical personality change. Long-term studies of others who have had the treatment show that they almost always go on to have long and happy lives, though these lives are informed by desires and values that differ profoundly from those that the person had prior to treatment. You can therefore reasonably expect that, if you take the treatment, you



will live for roughly sixty more years, though the life you will have will be utterly discontinuous with your life as it has been. You will remember nothing of your past and your character and values will be radically altered. Suppose, however, that this can be reliably predicted: that the future you would have between the ages of twenty and eighty if you were to take the treatment would, by itself, be better, as a whole, than your entire life will be if you do not take the treatment.”<sup>134</sup>

The deprivation account implies that it would be bad for you to reject the treatment, because your life as a whole would contain considerably less value if you did – you only get five more years, rather than sixty. To McMahan, however, it seems that it would in fact be egoistically rational for you to reject the treatment, and he suggests that this intuition is best explained with reference to the psychological gulf there would be between your current self and your future self if the treatment was administered.

I do not share McMahan’s intuition about *The Cure*. If I were the one who is ill, I too might reject the treatment, but neither because I think doing so is good for me (I do not), nor out of irrationality. There is a distinction to be made between rationality and egoistical rationality. While it is necessarily true that an egoistically rational choice will have, or can reasonably be expected to have, good consequences for me, the same is not true for rational choices in general. It can be rational to make a choice that is bad for oneself, for example if doing so advances a cause one believes in. Say I am working on a political campaign whose goal I strongly support because it reflects values I deeply hold. If I accept the treatment, my personality and value system will undergo a radical change, and I will likely lose my motivation to take part in the campaign. Anyway, amnesia will rob me of all the valuable knowledge that made me an effective campaigner. In this case,

---

<sup>134</sup> McMahan 2002, p. 77.

it may well be rational for me to sacrifice the additional net prudential value that would have fallen into my life if I had accepted the treatment and instead reject it in order to be able to bring the campaign to a successful end, which I am now convinced is morally desirable.

Others similarly have reservations. In the opinion of Ben Bradley, for example,

“[t]he decision to refuse treatment is shortsighted and irrational [and] [...] seems in many ways similar to the decision of a child to ignore the consequences of his behavior on his adult self, since he does not currently care about the things his adult self will care about.”<sup>135</sup>

Not everyone though will agree. It is hence worthwhile to further pursue McMahan’s line of argument – also because his reasons for thinking that it is good for you to reject the treatment have implications for the comparative badness of death that very well align with the sense of a great many people.

McMahan thinks that you have less prudential reason to care about the future goods you lose by refusing the treatment than you would have if your post-treatment self was more closely related to your current self.

“[While] [t]he future you would have with the treatment would contain vastly more good than you will have if you refuse the treatment, [...] the future offered by the treatment is too much like someone else’s future. In that future, you would be a complete stranger to yourself as you are now.”<sup>136</sup>

He proposes that, when determining what is *now* in your best interest, future goods hence be weighted in proportion to the strength of the psychological connections that establish a

---

<sup>135</sup> Bradley 2008, p. 293. For the reasons given in the previous paragraph, I think Bradley should have written “egoistically irrational” rather than “irrational.”

<sup>136</sup> McMahan 2002, p. 78. It would still be *your* future, but you are not very much invested in it.

relation between your current self and yourself at the time at which those future goods would occur. If you decide to undergo the treatment, your personality will change drastically. The psychological connections between your post-treatment self and your pre-treatment self will be very weak. Accordingly, the goods that will fall into the remaining sixty years of your life ought to be heavily discounted, from your present point of view. On the other hand, if you refuse the treatment, you will only live five more years. But, from your present point of view, the goods that will fall into your life during these five years matter greatly. McMahan's proposal therefore implies that, at the moment at which you must decide whether or not to accept the treatment, it would be in your best interest to refuse.

If this is how one ought to think about harm, then the deprivation account of the badness of death is too simple, as it does not take into account the prudential unity relations that connect different parts of one's life and ground rational egoistic concern. On McMahan's account of harm, the badness of death for an individual is not determined solely in terms of its effect on the value of that individual's life as a whole, but by the strength of that individual's *time-relative interest* in continuing to live. McMahan introduces the notion of time-relative interests to capture what is in one's best interest – that is, “what one has egoistic reason to care about”<sup>137</sup> – at some particular time. “The strength of an individual's time-relative interest in continuing to live is [...] the extent to which it matters, for his sake now or from his present point of view, that he should continue to live.”<sup>138</sup> In addition to the good his future life would contain if he continued to live, it also takes into account “the strength of various psychological relations, such as desire,

---

<sup>137</sup> McMahan 2002, p. 80.

<sup>138</sup> McMahan 2002, p. 105.

belief, memory and so on, that would have bound an individual at the time of death to himself at those later times at which the goods of his future life would have occurred.”<sup>139</sup>

For people like you and me, the *time-relative interest account of the badness of death*, as McMahan calls it, generally does not depart significantly from the deprivation account. Unless Alzheimer’s disease strikes or something else drastic happens that causes one’s personality to disintegrate, the psychological connections that hold between mature human beings and the rest of their lives are strong. Hence, an adult’s time-relative interest in continuing to live more or less precisely corresponds to the net prudential value that that person would lose if he or she died. If you and I are living equally happy lives, then both the deprivation account and the time-relative interest account imply that, other things being equal, it would be worse for the younger of the two of us to die than it would be for the older one. The two accounts come apart, however, when applied to immature human beings. While infants, if killed, generally lose more than twenty-year-olds, infants are less invested in their futures than twenty-year-olds are invested in theirs. Insofar infants remember what happened a week ago, for example, these memories are less in number and less rich in content than the memories adults have of the near past. In general, the psychological connections between subsequent future selves are much weaker in the case of infants than in the case of adults. Hence, while the extent to which a twenty-year-old will be psychological related to himself or herself at some later point in life, if he or she continues to live, is significant, it is minimal for an infant. Suppose both an infant and a twenty-year-old would, at forty, experience the joy of attending their daughters’ graduation ceremonies, if they continued to live. The infant has less egoistic reason now to care about that future good than the twenty-year-old. The time-relative

---

<sup>139</sup> McMahan 2013b, p. 10.

interest of an infant not to be deprived of future goods through death is weaker than that of a twenty-year-old. All this also holds, to a slightly lesser extent, for a fetus and an infant. Therefore, on McMahan's account, the death of a fetus is a lesser harm to the fetus than the death of an infant is to the infant, and the death of an infant is a lesser harm to the infant than the death of a twenty-year-old is to the twenty-year-old.

The combination of the time-relative interest account of the badness of death and the claim that, other things being equal, the wrongness of killing is a monotonically increasing function of the badness of death more readily lends itself as a basis of a defense of abortion than other harm-to-the-victim accounts, which have the implausible implication that the killing of a fetus is normally more seriously wrong than the killing of a child or a mature human being.<sup>140</sup> The *time-relative interest account of the wrongness of killing*, as McMahan calls that combination, implies that, other things being equal, it is more seriously wrong to kill a twenty-year-old than an infant, and more seriously wrong to kill an infant than a fetus, making it intuitively superior to other harm-to-the-victim accounts.

McMahan thinks that the time-relative interest account adequately accounts for the wrongness of killing those who fall below the threshold of personhood, but he rejects it as an adequate account of the wrongness of killing persons, which he vaguely defines as entities "with a mental life of a certain order of complexity and sophistication."<sup>141</sup> While the strength of prudential unity relations between present and future self does not vary widely among persons, prospects for future goods do, and young and cheerful persons have a stronger time-relative interest in continuing to live than old and melancholic

---

<sup>140</sup> Arguments for the permissibility of abortion based on the time-relative interest account can be found in McMahan 2002, pp. 267-362, DeGrazia 2003, pp. 413-442, and DeGrazia 2005, pp. 279-294.

<sup>141</sup> McMahan 2002, p. 6.

persons. If applied to persons, the time-relative interest account hence implies that, other things being equal, it is more seriously wrong to kill young persons than old persons, cheerful persons than melancholic persons, etc. These implications clash with our sense of the moral equality of persons, and McMahan instead endorses the commonly accepted view that “the wrongness of killing persons does not vary with such factors as the degree of harm caused to the victim, the age, intelligence, temperament, or social circumstances of the victim, whether the victim is well liked or generally despised, and so on.”<sup>142</sup> McMahan calls this profoundly intuitive thesis the *equal wrongness thesis*, and he thinks that it is best understood and defended as following from what he calls the *intrinsic worth account*, according to which all persons have equal worth, and killing a person is normally wrong because it is a failure to show due respect for the person and his or her worth, and more seriously morally wrong than killing a non-person.<sup>143</sup> For now, let us put this, the second tier of McMahan’s theory aside, and see how the time-relative interest account fares when applied to those for which McMahan proposes it is adequate.

McMahan never gives a precise answer to the question of how complex and sophisticated the mental life of an animal has to be for that animal to qualify as a person. But he thinks that the answer should be guided by the innocent-looking assertion that “[i]t is uncontroversial that the killing of an [non-human] animal is normally less seriously

---

<sup>142</sup> McMahan 2002, p. 235. “Endorsement” might be too strong a word here. McMahan is well aware that his two-tiered account of the wrongness of killing has problems, some of which we will discuss. As a consequence, he is careful to present it as a tentative proposal, to be refined and improved upon, and stops short of all-out endorsing it in its current form (cf. McMahan 2008, p. 82).

<sup>143</sup> Note that, despite its name, the equal wrongness thesis does not imply that all wrongful killings of persons are equally wrong. McMahan allows for the wrongness of killing persons to vary in ways that he thinks are consistent with the moral equality of persons. Factors that may affect the wrongness of killing persons include the killer’s mode of agency, the presence of defeaters, the number of persons killed, and the presence of a special relation (cf. McMahan 2002, p. 236).

wrong than the killing of a person.”<sup>144</sup> While McMahan leaves open the possibility that some (very few) non-human animals are persons in the relevant sense, it seems clear to him that the vast majority are not, and hence do not fall into the domain of the equal wrongness thesis. Considering the remarkable cognitive abilities of many non-human animals who McMahan believes are not our moral equals, that implies that there are some human beings with mental lives not sufficiently complex and sophisticated for them to qualify as persons. McMahan specifically mentions fetuses, infants, and human beings with certain severe congenital cognitive impairments.<sup>145</sup> These human beings fall below the threshold of respect and the morality of killing them hence is supposed to be governed by the time-relative interest account. Accordingly, infanticide, for example, is wrong because it is bad for an infant to be killed. That is, unless others have an interest in the death of the infant that outweighs the infant’s time-relative interest in continuing to live. McMahan considers such a case, a variation on Philippa Foot’s transplant case, and concludes that, according to his view, “it would be permissible (and perhaps even morally required, if other things are equal) to kill the healthy, orphaned newborn in order to use its organs to save [...] three other children.”<sup>146</sup> Many people will regard this as a *reductio ad absurdum*, and McMahan admits that he cannot bring himself to embrace that implication of his theory “without significant misgivings and considerable unease.”<sup>147</sup> I too find it hard to accept that infants may be killed if there are sufficiently strong interests that would be served by their death, just like I find it hard to accept that infanticide and

---

<sup>144</sup> McMahan 2002, p. 190.

<sup>145</sup> Cf., e.g., McMahan 2008, pp. 83 f.

<sup>146</sup> McMahan 2002, p. 360.

<sup>147</sup> McMahan 2002, p. 360.

killing other human non-persons is less seriously wrong than killing a normal adult.<sup>148</sup> McMahan defends his position by pointing out that many people would find the alternative equally intolerable, and that we hence will have to give up one deeply held conviction or another. If infants are not sacrificable, then neither are non-human animals with comparable psychological abilities. It would follow that such animals – McMahan mentions baboons as an example – may not be killed to save the lives of human children.<sup>149</sup> That, he says, too starkly contrasts with common intuition.

More counterintuitive implications emerge when we turn our attention to comparisons between the wrongness of killing different non-persons. As infants, unlike persons, do not fall into the domain of the equal wrongness thesis, the extent to which killing them is wrong can vary with their social circumstances. Consider the parts of Syria and Iraq that currently are – and might, for a long time, remain – under the control of the so-called Islamic State of Iraq and the Levant (ISIL). Because of the vicious gender discrimination enforced by ISIL, women there generally live considerably worse lives than men. Infant boys born in that part of the world hence have a stronger time-relative interest in continuing to live than infant girls, and the time-relative interest account implies that it is less seriously wrong to kill infant girls than infant boys in ISIL-controlled territory.<sup>150</sup> This adds insult to injury and, to me, is an unacceptable implication of McMahan's position. Other factors that the equal wrongness thesis holds to be irrelevant to the wrongness of killing persons, such as intelligence or temperament, invite similar

---

<sup>148</sup> McMahan nowhere explains how exactly we are supposed to evaluate the comparative wrongness of killing a person and a non-person. As those two kinds of killing are said to be wrong for fundamentally different reasons, it is not clear to me (or to him, as he told me in a personal e-mail) how to think about the relative force of these reasons.

<sup>149</sup> Cf. McMahan 2002, p. 361.

<sup>150</sup> Kasper Lippert-Rasmussen discusses a similar case in Lippert-Rasmussen 2007, p. 732.



counterexamples, which can not only be constructed for infants and other human non-persons, but also for non-human animals. This should be particularly troubling to McMahan, as it is his belief in the adequacy of the time-relative interest account in the case of non-human animals that led him to grudgingly also accept that account in the case of human non-persons in the first place. Imagine two dogs. One of them, Max, has been badly abused as a puppy. As a result, he is physically and psychologically scarred. These scars will likely remain for a very long time. Ever since his rescue, however, his condition has been steadily improving. Max sometimes still suffers from sudden attacks of fear and anxiety, in absence of any real danger, but his days now generally contain significantly more good than bad, and he has learned to enjoy life again. The quality of his life will never rise to the level of that of Daisy's though. Daisy has been with a loving and caring family since birth. She was a happy puppy, and is now a happy bitch. If Max and Daisy are about the same age and have about the same life expectancy, and all other things are equal, then the time-relative interest account implies that it would be less seriously wrong to kill Max, as he would lose less through death. Yet, it seems that, if there is a moral reason that justifies, or requires, killing either one of the two dogs, Max should be given at least an equal chance that he will not be picked. Some might even feel that the choice should fall on Daisy rather than Max, because Max so far did not get nearly as much out of life as Daisy did, and hence deserves, more than Daisy does, to live out the remainder of his life, and enjoy whatever little joy future has in stock for him. Borrowing Nozick's phrase, we conclude that the time-relevant interest account cannot be "the whole story," neither for animals, nor for human non-persons.

### 2.3. *The distinction between persons and non-persons, and its moral significance*

I have earlier called the line that McMahan draws between persons and non-persons a threshold. I did so, for example, when I referred to sentient non-persons as “those who fall below the threshold of personhood.” Not every line is a threshold though, and I hence owe an explanation. That explanation will lead us to a fundamental problem of McMahan’s two-tiered account, which is independent from questions about the plausibility of its two tiers, considered separately. McMahan recognizes the problem, but the solutions he offers, I will argue, are inadequate.

A person, in the sense relevant to our discussion, is a sentient animal with a sufficiently complex mental life, where complexity is measured in terms of certain psychological capacities. The most important such capacity, according to McMahan, “is probably autonomy (which presupposes self-consciousness and some degree of rationality).”<sup>151</sup> If autonomy was an all-or-nothing affair, one could take this to mean that one thing that distinguishes persons from non-persons is that the former are autonomous and the latter are not. To describe the relevance of autonomy to personhood in such simple terms would be misleading though, as it makes good sense to say that some are more autonomous than others, which suggests that autonomy is a more-or-less rather than an either-or.

Autonomy, very roughly, is the ability to do one’s own thing, a form of self-government, which stands in contrast to being a slave to others, or one’s instincts. There are different proposals on how to make this more precise. The differences between these proposals, however, are not important to us, so we will consider just one example. Gerald Dworkin

---

<sup>151</sup> McMahan 2002, p. 261.

defines autonomy as “a second-order capacity to reflect critically upon one’s first-order preferences and desires, and the ability either to identify with these or to change them in light of higher-order preferences and values.”<sup>152</sup> Autonomous beings must hence be able to distance themselves from their first-order desires, at least sometimes, evaluate them with a cool head, and change them, if they are incompatible with those parts of their psychology that are (commonly thought to be) more indicative of who they are, such as their higher-order desires and values. That requires, among other abilities, abstraction, introspection, and impulse control. Some of us are better at these things than others. There are those who spend a lot of time thinking deeply about their goals and values, and then try to live their lives in accordance with what really matters to them – and they are good at it. They easily abstract from their current reality, know how to effectively keep their impulses and instincts from messing with their rational deliberation process, and are well able to adjust their motivations such that they are more in line with their authentic selves. Others are less able to control their lives. They might wish that they did not have the desire to smoke, for example, but never quit smoking anyway, because they lack the willpower to do so. While they have it in them to change some of their first-order desires, the desire to smoke is just not one of them. They further lack the critical thinking skills that would be required to overcome much of the blind conformity, custom, and tradition that dictate most aspects of their lives, often in a way that keeps them from truly being their own persons. Etc. Plainly, the former group of people is more autonomous than the latter. Autonomy is best understood as a scalar property that varies from one person to another. The same is true for other psychological capacities that are commonly associated

---

<sup>152</sup> Dworkin 1988, p. 108; see also Dworkin 1976.

with personhood. Rationality, moral personality, self-consciousness, and so on, too come in degrees.<sup>153</sup>

Now, if personhood – a matter of either-or – is based on the possession of certain properties that vary in degree, it must be the possession of these properties to a degree that surpasses a certain threshold that is necessary and sufficient for personhood, and hence for falling within the scope of the equal wrongness thesis. Where does that threshold lie? McMahan gives some clues about where he thinks the threshold lies, but he stops short of specifying a precise point. That is entirely understandable, as it is unclear, to say the least, how such a point could be specified, even if one wanted to. Neither is there a comprehensive and universally agreed-upon list of psychological capacities that constitute the basis of personhood, nor are there quantitative measures for these capacities. At this point, where our scientific understanding of the mind is still very limited, assigning numbers to highly complex mental phenomena, such as rationality, would be simplistic and misleading at best, and pseudo-science at worst. That does not mean that the concept of personhood lacks semantic content though. Consider baldness, which similarly is a binary property that supervenes upon a property which comes in degrees, the property of having hair on the head. Some people have more hair, some less. Cases of people who are at any one of the two ends of the spectrum are clear. If your head is entirely covered with hair, you are clearly not bald. If you have three hairs on your head, no matter how they are distributed, I am afraid there is no doubt that you are bald. But there are cases in between, in which it is not clear at all whether or not the concept of baldness applies. Yet, that does not necessarily mean that there is no well-

---

<sup>153</sup> For a detailed and careful argument for the claim that “moral agency [...], like self-awareness and language, admits of both kinds and degrees” (DeGrazia 1996, p. 204), see DeGrazia 1996, Chapter 7.

defined threshold for baldness, nor that the concept of baldness has no content – we just saw it has. Analogously, although we are not able to draw a precise line that separates persons from non-persons, this neither implies that there is no such line, nor that the concept of personhood has no content. Normal adult human beings clearly are persons, ants, worms, and orchids clearly are not. That much is uncontroversial. McMahan though is more specific, arguably as specific as we can expect anybody to be. Fetuses, infants, and human beings with certain severe cognitive impairments, he thinks, too are not persons, as otherwise consistency would require that non-human animals with comparable psychological capacities be recognized as our moral equals as well, which he is convinced they are not. An example for animals about which McMahan remains agnostic are gorillas.<sup>154</sup> This is all fair enough. The problem lies elsewhere.

Personhood as a threshold concept with a well-defined, yet unknown threshold, is meaningful, and there are enough cases where it clearly applies for it to be useful, but it is not suited to carry the moral weight McMahan wants it to carry. While baldness is morally innocent, and hence unproblematic, personhood is neither. “Morally, the gap between those above the threshold and those below it is immense.”<sup>155</sup> According to McMahan, to kill an animal below the threshold of personhood is morally wrong because, and to the extent that, death is bad for that animal. In contrast, to kill an animal above the threshold is wrong, and much more seriously so than killing a non-person, because it fails to show the respect that he or she is owed due to his or her being a person. This stark difference in moral status contrasts with a seemingly insignificant difference in relevant psychological capacities between those just below the threshold, and those just

---

<sup>154</sup> Cf. McMahan 2002, p. 260.

<sup>155</sup> McMahan 2002, p. 261.

above. Where there is supposed to be a canyon in terms of morality, development psychology does not even allow for a fissure. The psychological capacities relevant to personhood develop gradually, without any abrupt discontinuities. Both you and I, sometime in the past, had a mental life no more complex than that of an average, grown-up dog or cat. Today, both of us are persons, fully capable of governing our own lives. In between then and now, there must have been a moment where we crossed the threshold of personhood, wherever exactly that threshold may be located on the spectrum of psychological capacities. Psychologically, there is nothing about this point in time that makes it stand out. Morally, however, it is momentous. According to McMahan, it is the point at which human beings suddenly transform from non-persons, whom it may or may not be wrong to kill, depending on how much harm they would suffer if killed, into persons, whom it is normally seriously morally wrong to kill, regardless of the harm they would suffer if killed.<sup>156</sup> In just an instant, the time-relative interest account becomes irrelevant, and the intrinsic worth account, which appeals to a wholly distinct set of moral considerations, takes over. This, I think, is too weird to believe. It is implausible that a minor, seemingly insignificant difference in capacity can make a fundamental difference in moral status.<sup>157</sup> The threshold concept of personhood, it seems, is not suited to bear the moral weight McMahan assigns to it, just like baldness would not be either. “McMahan wants to combine a naturalistic, broadly Humean, picture of the world where continuous properties come in degrees, with a set of Kantian intuitions that clearly require sharp

---

<sup>156</sup> Many normal adult human beings also lose their personhood at some point before their death and spend the last part of their life as non-persons. My argument equally applies to that transition.

<sup>157</sup> Also, why is it that here a small difference in capacity is morally crucial, whereas the degree to which a person’s relevant psychological capacities exceed the threshold is wholly irrelevant?

boundaries between persons and non-persons.”<sup>158</sup> This, Tim Mulgan correctly observes, “is an essentially unstable combination [...]”<sup>159</sup>

McMahan acknowledges the problem,<sup>160</sup> which he admits makes him “profoundly uncomfortable.”<sup>161</sup> He writes that it “is hard to avoid the sense that our egalitarian commitments rest on distressingly insecure foundations.”<sup>162</sup> In a first attempt to build a more secure foundation, he proposes the introduction of an intermediate moral status for those who are in the process of becoming persons, but are not quite persons yet. “During that period, we are neither altogether unworthy of respect nor worthy of full respect; we are neither freely violable in the service of the greater good nor fully inviolable.”<sup>163</sup> The threshold, which in the simpler model was a point, is now a line. Young children who are capable of governing some aspects of their lives, but still are not yet the moral equals of you and me, are somewhere on that line, and maybe so are gorillas and other cognitively advanced non-human animals. They are, so to say, “within the threshold itself.”<sup>164</sup> McMahan’s two-tiered account effectively transforms into a three-tiered account. Nowhere, however, does McMahan fully develop the middle tier. Doing so is one of the problems his 2002 book leaves unresolved. One possibility, the only one McMahan explicitly mentions in that book, is to describe the wrongful killing of those with intermediate moral status as a failure to show adequate respect, too, with the caveat that they have a lower worth than persons, and assuming that the wrongness of killing varies

---

<sup>158</sup> Mulgan 2004, p. 458.

<sup>159</sup> Mulgan 2004, p. 458.

<sup>160</sup> Cf. McMahan 2002, pp. 261 ff., and McMahan 2008, pp. 93 ff.

<sup>161</sup> McMahan 2008, p. 104.

<sup>162</sup> McMahan 2008, p. 104.

<sup>163</sup> McMahan 2002, p. 265.

<sup>164</sup> McMahan 2002, p. 265.

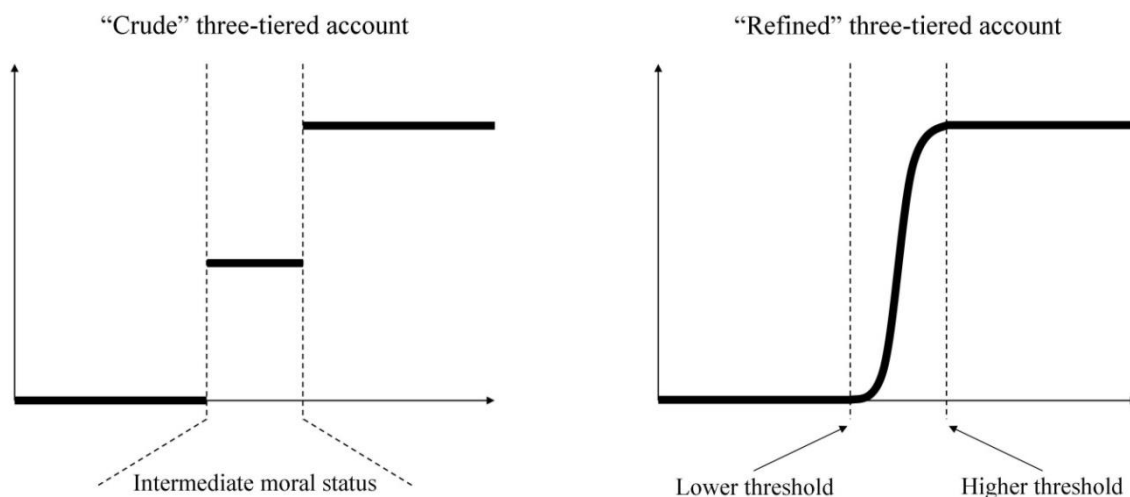
with the worth of the one being killed.<sup>165</sup> To illustrate this, imagine a diagram where the horizontal axis signifies the degree of psychological complexity, and the vertical axis signifies moral worth. The function relating the two characteristics starts at the origin and moves along the horizontal axis until it reaches the lower threshold, which separates non-persons from those with intermediate moral status. There, the function jumps to some positive value, and continues horizontally until it reaches the higher threshold, which separates those with intermediate moral status from persons. Another jump follows, to some greater positive value, and the function again continues horizontally. In one regard, this makes things better, as the transition from non-person to person is no longer instantaneous but instead occurs in stages. There is now a stage in between, which differs less from both its predecessor and its successor than they differ from each other. Young children, while still not our moral equals, command at least some respect in this picture. That is closer to common intuition than the suggestion that they are wholly sacrificable. In another regard, however, the three-tiered account makes things worse. The diagram we described is a step function, with two discontinuities. The height of its steps is less than the height of the single step in the analogous diagram corresponding to the two-tiered account. The core objection to the two-tiered account, however, is not that the step to personhood is too high, but that there is a step at all. A discontinuous progression of moral status sits unwell with the smoothly gradual spectrum of psychological capacities on which it is superimposed, and that objection applies equally, or to an even greater extent, to the three-tiered account, which has not just one but two abrupt discontinuities. Around both of these discontinuities, a seemingly insignificant difference in capacity is supposed to make a fundamental difference in moral status. That is implausible, as we

---

<sup>165</sup> Cf. McMahan 2002, p. 265, and McMahan 2008, pp. 97 ff.



have observed before, and the reason why McMahan further refines his account in a 2008 paper, where he moves from the view that all those with intermediate moral status have the *same* moral status to a more nuanced view, according to which the treatment of animals whose psychological capacities lie between the two thresholds “is governed by constraints [whose] [...] strength [...] varies with the level of psychological capacity of the individual to whom they apply.”<sup>166</sup> The straight line which represents intermediate moral status in the diagram illustrating the crude three-tiered account is replaced by a monotonically increasing, continuous function that connects the line representing the moral status of those who are wholly sacrificable with that representing the moral status of those who are maximally inviolable:



According to the refined three-tiered account, you and I did not undergo a radical change in moral status when we achieved personhood earlier on in our lives. We merely changed from being almost-maximally inviolable to being maximally inviolable, and we did so gradually. That stands in contrast to both the two-tiered account and the crude three-tiered account, according to which we abruptly changed from being wholly sacrificable to

<sup>166</sup> McMahan 2008, p. 98. McMahan writes that the former view is “cruder” (McMahan 2008, p. 98, n. 15) than the latter.

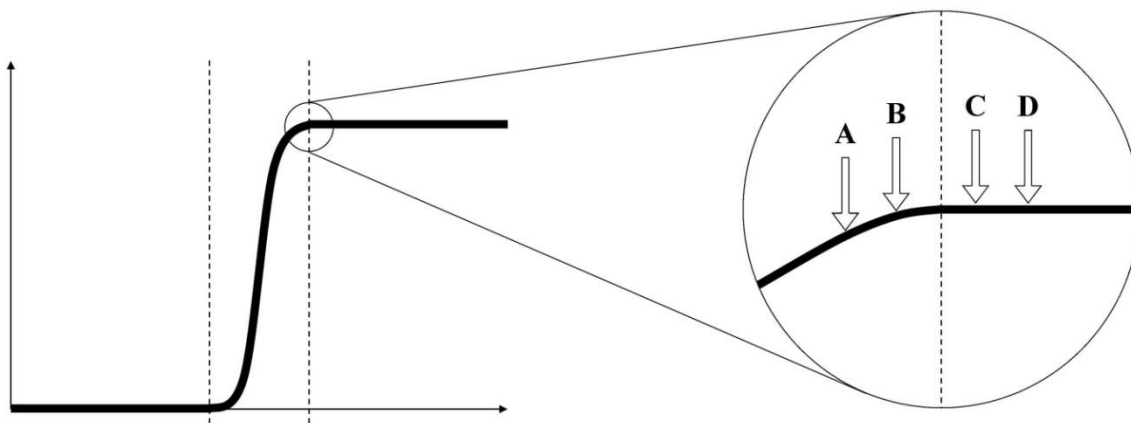
being maximally inviolable, and from being somewhat inviolable to being maximally inviolable, respectively. The refined three-tiered account is hence an improvement, in that, unlike its predecessors, it does not implausibly associate a jump in moral status with an insignificant change in psychological capacity. However, while an improvement, it does not fully solve the problem. The change in moral status around the thresholds might be small, but it can still make all the difference. The thresholds retain a special moral significance which has no analogue in the spectrum of psychological capacities, and the implausible mismatch between moral status and psychological capacity that led us to reject McMahan's earlier accounts hence persists in the refined three-tiered account.

Animals with intermediate moral status are protected by constraints with varying levels of stringency. The more developed their relevant psychological capacities, the harder it is to morally justify killing them. If, in a given situation, harm to others can be prevented by killing an animal with intermediate moral status, and that harm is just barely great enough to morally justify killing that animal,<sup>167</sup> then another animal who also falls below the higher threshold but has significantly more developed psychological capacities may not be killed in a relevantly similar situation. "In order for the sacrifice of [...] [the latter animal] to be permissible, the harm that would thereby be prevented would have to be [...] greater,"<sup>168</sup> and it would have to be even greater for persons, if they can be sacrificed at all.

---

<sup>167</sup> Note that there might be other situations in which the magnitude of harm to others that can be prevented is the same, yet killing would not be justified. That is, because there might be other conditions, not concerned with the magnitude of harm, that must be met. These other conditions, for example, might be concerned with whether or not those who stand to be harmed are innocent, or with promises that were made.

<sup>168</sup> McMahan 2008, p. 99.



Now consider four children who rank very close to one another in terms of psychological capacity: A, B, C, and D. A and B are yet to become persons. Their levels of psychological capacity are slightly below the higher threshold, with B being marginally ahead of A. C and D are your and my moral equals already. Their levels of psychological capacity are slightly above the higher threshold, which indicates that it was only very recently that they transitioned into personhood. D is psychologically more mature than C, but the difference is as insignificant as it is in the case of A and B. McMahan's refined three-tiered account implies that there is a certain amount of harm to others, for the prevention of which it could be permissible to sacrifice A but not B. If there are no special reasons not to kill A and no special reasons not to kill B, such as reasons arising from promises or special relationships, then whether or not either one of them may permissibly be killed in order to prevent harm to others only depends on the magnitude of harm to be prevented, and there will be harm that is just barely great enough to morally justify killing A and hence not quite great enough to morally justify killing B. The same cannot be said about C and D. If the harm to be prevented is great enough to make it permissible to sacrifice C, then, all other things being equal, the same harm is sufficient to morally justify killing D, and vice versa. If the harm to others that could be prevented

by killing C is not sufficiently great to justify killing C, then the same harm would also be insufficient to justify killing D if D was in C's situation, and vice versa. I leave it open whether the harm to be prevented can ever be great enough to make it permissible to sacrifice a person. Either way, the difference between C and D in terms of psychological development makes no moral difference. They enjoy the same protection as anybody else in the exclusive club of persons. Above the higher threshold, differences in psychological capacity stop mattering and everybody is equal. While the ethics of killing is governed by the same kind of moral considerations just below and just above the higher threshold, and there is no abrupt discontinuity in moral status, something dramatic still happens at the higher threshold: It separates unequals from equals. As we have seen, in some rare cases, it can make all the difference whether one is an unequal and slightly less mature than another unequal, or an equal and slightly less mature than another equal. What is so special about the higher threshold that just below it a small difference in psychological capacity can be morally crucial, whereas just above it an equally small difference in psychological capacity is wholly irrelevant? Scientifically, there is nothing about this point in the spectrum of psychological capacities that makes it stand out, and mapping it onto a point in the scale of moral worth that does have a special moral significance hence is arbitrary. There is no conceivable reason not to make any other nearby point in the spectrum of psychological capacities the higher threshold instead. The arbitrariness in choosing the higher threshold is not the innocent kind that is at work, say, when we define baldness. Nothing morally important depends on whether we define "bald" to mean "having less than 123 hairs on one's head" or "having less than 122 hairs on one's head." In McMahan's ethical framework, however, whether or not the threshold of

personhood is placed below or above the levels of capacity of two given young children can potentially make the difference between it being morally permissible to sacrifice the less mature child yet impermissible to sacrifice the other child on the one hand, and it being morally permissible to sacrifice any one of the two children on the other hand. In the latter case, each child has an equal chance to continue living, if a third party chooses to kill one of them, as he or she may. For the less mature child, this can mean the difference between life and death. Hence, there better be good reasons to set the higher threshold one way or another. I have argued that there are no such reasons, which suggests that there should be no arbitrary higher threshold.<sup>169</sup> Besides, to repeat a point from the previous section, that there could be a situation in which it would be morally permissible to kill A but not B, and a situation in which it would be permissible to kill B but not C, just because A is slightly less mature than B and B slightly less mature than C, deeply offends our sense that all human beings, or at least all post-natal human beings, are equal. Two healthy young children, maybe even of the same age, but one slightly ahead in terms of psychological development, surely should enjoy the same protection. McMahan mentions another way to map the spectrum of psychological capacity onto the scale of moral status that we should briefly discuss, too, as it differs from the accounts we have talked about so far in that it does not have any points that stand out. The idea is rather straightforward: Take the function at the center of the diagram, corresponding to the refined three-tiered account, and stretch it horizontally as to make it cover the entire spectrum of psychological capacities, getting rid of the plateaus on the left and right in the process. That is, we could “embrace a fully gradualist account of moral status, one

---

<sup>169</sup> The same is true for the lower threshold. Whichever point in the spectrum of psychological capacities is chosen to correspond to the lower threshold, that choice is necessarily arbitrary, which is problematic insofar distinctions of moral import should not be based on arbitrary choices.

without significant thresholds.”<sup>170</sup> The advantage of this account is clear: Unlike its competitors, it does not make matters of great moral import depend on whether arbitrary choices are made one way or another. The downside, however, is that a fully gradualist account of moral status moves us even further away from common sense morality.

“Such a view might hold that the killing of an individual is more seriously wrong the higher that individual’s relevant psychological capacities are. Or it might hold instead that the degree to which it is wrong to kill an individual is determined by weighting that individual’s interest in continuing to live by the level of his or her psychological capacities.”<sup>171</sup>

Either way, it would be more seriously wrong to kill an adult human being with an extraordinarily high level of psychological capacity than an average adult human being. I agree with McMahan that “[t]his sort of elitist view is profoundly counterintuitive.”<sup>172</sup> Of course, this does not conclusively show that it is false. If there really was no plausible way to account for our egalitarian intuitions, we might just have to abandon the thesis that it is equally wrong to kill you or me. Whether there is such a way will be the central question of the two subsequent chapters.

#### **2.4. *The second tier: the intrinsic worth account of the wrongness of killing***

In the previous chapter, we noted that consequentialist explanations of the wrongness of killing do not contain any essential reference to the victim. That, we said, is a reason to

---

<sup>170</sup> McMahan 2008, p. 101.

<sup>171</sup> McMahan 2008, p. 101.

<sup>172</sup> McMahan 2008, p. 102.

reject such explanations. Intuitively, the answer to the question of why it is wrong to kill should more prominently feature the one being killed. Harm-to-the-victim accounts fare better in that regard, but also fail to satisfy. Unlike consequentialist explanations of the wrongness of killing, which ultimately appeal to the loss of good to the world and can be stated without any mention of the victim, explanations in terms of the future valuable experiences that killing takes away from the victim necessarily mention the one being killed – it is the harm *to him or her* that does the explaining. This focus on what killing takes away from the victim, however, is misplaced, as we have seen, for example, when we discussed the case of involuntary euthanasia. The focus, we have argued, should instead be on the victim himself or herself, as someone who is morally important and valuable as an individual, and the fact that he or she is annihilated.

The first tier of McMahan's theory, the time-relative interest account of the wrongness of killing, is more sophisticated than the harm-to-the-victim accounts we have considered in the previous chapter, but it is a harm-to-the-victim account nevertheless, and hence subject to the same wrong-explanation objection. In contrast, McMahan's intrinsic worth account, which is intended to explain the wrongness of killing persons, seems to be a significant improvement:

“To kill a person, in contravention of that person's own will, is an egregious failure of respect for the person and his worth. It is to annihilate that which is irreplaceable, to show contempt for that which demands reverence, to assert a spurious authority over one who alone has proper authority over his own life, and to assume a superior position vis-à-vis one who is in reality one's moral equal. Killing is, in short, an offence against what might be called a requirement of

respect for persons and their worth. Indeed, because killing inflicts the ultimate loss – the obliteration of the person himself – and is both irreversible and uncompensable, it is no exaggeration to say that it constitutes the ultimate violation of the requirement of respect.”<sup>173</sup>

Persons are intrinsically valuable, and equally so, regardless of the prudential value of their lives, and regardless of any instrumental value they might have, simply in virtue of being persons. Killing a person is normally wrong, according to McMahan, because it constitutes a failure to respect the victim and his or her worth.

This explanation affords the victim a place at its very center, and hence – one might think – the kind of prominence common sense demands. Timothy Chappell suggests, however, that the intrinsic worth account only ostensibly awards the desired importance to the one being killed. At second glance, he insists, we must accept that McMahan’s account, no less so than consequentialism, is what he calls a *receptacle view*, and should therefore be rejected. The most straightforward example of a receptacle view is utilitarianism, which portrays us as mere “bank accounts for utility,”<sup>174</sup> not valuable in ourselves, but only as receptacles in which impersonal value can be realized. Similarly, Chappell argues, it is not really us to whom McMahan’s account accords intrinsic value, but the higher psychological capacities that constitute personhood and set us apart from most or all non-human animals. As we only become valuable “when certain favored properties (the

---

<sup>173</sup> McMahan 2002, p. 242. At the center of this explanation is the concept of worth, which we must carefully distinguish from the concept of the prudential value of a life that plays a central role in the time-relative interest account. The prudential value of a life is the value of that life for the one who lives it, and hence determined by what happens within that life. As some people benefit more from life than others, the prudential value of people’s lives varies. In contrast, the worth of a person is independent of the contents of that person’s life. Worth does not vary from person to person. Therefore, when we say that the lives of all people have equal value, we do not mean to make a claim about the prudential value of their lives, but are instead saying that they all have equal worth, thereby commanding equal respect.

<sup>174</sup> Chappell 2004, p. 108.



psychological capacities) are instantiated [in us],”<sup>175</sup> and retain our value only as long as we continue to be persons (in the technical sense used by McMahan), we are in fact “merely possible receptacles for intrinsically valuable properties.”<sup>176</sup> If so, McMahan’s explanation of the wrongness of killing persons merely *appears* to focus on the fact that an individual is being destroyed, whereas the *actual* moral concern is for the destruction of certain desirable properties, and the one being killed as such does not really matter at all.

I think Chappell is on to something important, which we will try to bring out in the remainder of this chapter, but he is too quick to jump from the observation that the intrinsic worth account affords value to individuals only as long as they have the properties that constitute personhood to the claim that that account therefore is a thinly disguised version of the receptacle view, which is really a proposal to accord intrinsic worth to properties, rather than people. The problem with this argument is that it fails to distinguish between something being valuable in virtue of having certain properties on the one hand, and those properties being valuable themselves on the other hand. The analogous distinction in the practice of valuing is that between valuing *for* properties and valuing *of* properties. And, just as to value one’s bicycle for its grey color and its comfort is not the same as to value greyness and comfortableness, McMahan is hardly committed to the view that personhood is intrinsically valuable, rather than the individuals who instantiate personhood. McMahan is therefore quite right to point out that, “although [the] [...] psychological capacities [that constitute personhood] are ultimately the *basis* for respect, or that which makes persons worthy of respect, they are not themselves *objects*

---

<sup>175</sup> Chappell 2004, pp. 111 f., n. 27.

<sup>176</sup> Chappell 2004, p. 107.

of respect.”<sup>177</sup> Respect is for individuals, not for their properties. It is the person who is intrinsically valuable, not personhood.

I nevertheless share Chappell’s discomfort vis-à-vis McMahan’s proposal, and I think Chappell’s mistake is that he has misidentified that discomfort’s source. The real source, I believe, is not the fact that persons are said to be intrinsically valuable in virtue of having certain properties – after all, it seems plain that value always supervenes on non-evaluative properties of that which has value –, but rather the particular *kind* of properties McMahan proposes as a basis of worth, and thereby of membership in the community of moral equals. Given McMahan’s account of personal identity, according to which you and I are embodied minds that came into existence with the onset of consciousness, the basis of worth, personhood, is an accidental property.<sup>178</sup> To say that personhood is an accidental property is to say that it is possible for us to exist and not be persons. In fact, for both you and me, there was a time in the past when this possibility was realized. During the first phase of our existence, before we reached a certain age, we lacked the higher psychological capacities that constitute personhood. Consequently, while still immoral on account of the harm it would have caused, it would not have been a failure of respect to kill us back then, as we were not yet intrinsically valuable persons – according to McMahan. That, I think, is not how we like to think, and typically do think, about intrinsic value and respect. We want to be valued and respected for what we really are, rather than for what we sometimes happen to be. Respect, in that regard, has a certain similarity with love, and it will be instructive to compare the two concepts. “For Anne Gregory,” the well-known 1933 poem by William Butler Yeats starts like this:

---

<sup>177</sup> McMahan 2002, p. 243.

<sup>178</sup> For a defense of McMahan’s embodied mind account of personal identity, see McMahan 2002, Chapter 1.

“‘Never shall a young man  
 Thrown into despair  
 By those great honey-coloured  
 Ramparts at your ear,  
 Love you for yourself alone  
 And not your yellow hair.’  
 ‘But I can get a hair-dye  
 And set such colour there,  
 Brown, or black, or carrot,  
 That young men in despair  
 May love me for myself alone  
 And not my yellow hair.’”<sup>179</sup>

The blonde woman in this poem is dismayed at the thought of being loved just for her yellow hair. Her anxiety does not stem from the fear that she will not be loved at all, if she is loved for her yellow hair. After all, if one is loved for the color of one’s hair, one *is* loved. Rather, her concern is that she might not find the particular kind of love she is longing for. She does not want to be loved for something as trivial as the color of her hair – such love seems no more desirable than having somebody love one’s hair. She wants to be loved for the *right reasons*, as the common expression goes.<sup>180</sup> She wants to be loved “for herself alone.”<sup>181</sup> The color of one’s hair, to her, and to most of us, I presume, does not seem to be an appropriate basis for true love, and neither do other properties that are located at the periphery of the self of the one who is loved. We want to be loved for properties that are fundamental or essential to who we are, or “properties [...] [we take]

---

<sup>179</sup> Yeats 1962, p. 55.

<sup>180</sup> Cf. Delaney 1996, p. 343.

<sup>181</sup> For discussions of love for oneself alone, cf. Delaney 1996, Lamb 1997, Velleman 1999, and Bicknell 2010.

to be central to [...] [our] self-conception [...],”<sup>182</sup> which guarantees that love is not fleeting, but survives superficial changes in appearance, or character. True love, we like to believe, is deeply personal, rather than shallow, and has our innermost selves as a basis.

Similarly, I think, not every kind of property is equally appropriate as a basis for worth and, in turn, moral equality. Worth is not something we expect to disappear if our periphery changes. In fact, while we may accept that even the truest love can lose its basis and hence dissolve when lovers change in profound ways, we commonly think of worth, or dignity, as something that stays with us as long as we are. Respect, intuitively, is the appropriate response to the incomparable value of the essential selves of individuals like you and me. While there is no logical contradiction in holding that we are intrinsically valuable in virtue of accidental properties, an account of worth that grounds worth in this way seems superficial. It backgrounds you and me, and foregrounds our psychological capacities. It emphasizes aspects of ourselves that are not essential to who we are, and fails to focus our attention on what really matters: us in our own right. An account of equal moral worth that accommodates these intuitions, which I believe are widely shared, and are reflected in human rights legislation globally, and recognizes not only that we all are the same, but that we are *fundamentally* the same, bases our equality not on the superficial fact that you and I both happen to have certain accidental properties, but on an essence you and I share, and holds that you and I are to be respected for ourselves alone, just for being us, is preferable in that regard to an account that makes dignity conditional in the way McMahan’s account does.

---

<sup>182</sup> Delaney 1996, p. 343; also compare Nozick 1989, p. 75.

An account that accords intrinsic value to us solely by virtue of who we are essentially is superior to McMahan's intrinsic worth account which holds that people are intrinsically valuable in virtue of having certain accidental properties not only because it is plausible in itself that dignity is an essential property, but also because it does not seem to share the shortcomings of the two-tiered account that we discussed earlier. According to McMahan, what makes persons each other's moral equals is the fact that they possess equal intrinsic value. As we have noted in the previous section, the difference in the way we treat persons on the one hand and those who are not our equals on the other hand is momentous. Such a difference in treatment, it would seem, should be based on a similarly momentous difference in non-evaluative properties. Yet, the accidental property McMahan proposes as a basis for the intrinsic value of persons – and hence as a basis for equality – is the possession of certain psychological capacities to a degree that surpasses a certain threshold, and the acquisition of that property in the normal development of human beings does not mark a momentous change in how they are non-evaluatively.<sup>183</sup> Further, there seems to be no principled way to set the threshold. Why choose this degree of development of some set of psychological capacities, rather than that degree? Any choice is necessarily arbitrary, making the property of personhood, as defined by McMahan, unsuitable as a basis for the intrinsic value and equality of you and me. If, however, intrinsic value is part of our essence, the line that needs to be drawn is that between you, me, and others like us on the one hand, and everything else on the other hand, and hence – as I will argue in subsequent chapters – both momentous and non-arbitrary.

---

<sup>183</sup> Patrick Lee calls this argument the *discontinuity-continuity argument*. He uses it, among other places, in Lee 2011.

Let us briefly summarize this chapter's discussion. We have pointed out that McMahan's account of the wrongness of killing implausibly implies that infants and other human non-persons may be sacrificed for the greater good, and that it fails to give an adequate account of the comparative wrongness of killing human and non-human non-persons. We have objected to the distinction between persons and non-persons on the ground that it lacks a credible empirical basis and is arbitrary. Finally, we argued that McMahan's morality of respect is superficial, and that it would be preferable to have an account which grounds our worth not in properties you and I happen to have, but in who we are essentially.

### Chapter 3: Are human beings more equal than other animals?

The belief that each human being has equal intrinsic worth or dignity, and hence ought to be treated with equal respect, is one of the most widely held moral beliefs of our time – even though, regrettably, it is not always practiced. It informs the moral consciousness of most members of most contemporary societies, and has found frequent expression in the law, particularly since the end of the Second World War. Human dignity occupies a central place in both the *Charter of the United Nations* (1945), whose preamble affirms faith in the “dignity and worth of the human person,” and the *Universal Declaration of Human Rights* (1948), which opens with the observation that the “recognition of the inherent dignity and of the equal and inalienable rights of all members of the human family is the foundation of freedom, justice and peace in the world” (Preamble), and proceeds to declare that “[a]ll human beings are born [...] equal in dignity” (Article 1). Subsequent to its inclusion in these seminal international documents, the concept of human dignity has made its way into numerous national constitutions. While the constitutions of only five countries appealed to the concept of human dignity before 1945, the constitutions of 162 countries, out of the 193 sovereign countries that are members of the United Nations, did so in 2012.<sup>184</sup> In the *Basic Law for the Federal Republic of Germany* (*Grundgesetz*, 1949), for example, the concept of human dignity has been given a most prominent place, in Article 1, Section 1: “Die Würde des Menschen ist unantastbar. Sie zu achten und zu schützen ist Verpflichtung aller

---

<sup>184</sup> Cf. Shultziner & Carmi 2014.

staatlicher Gewalt. [Human dignity shall be inviolable. To respect and protect it shall be the duty of all state authority.]”<sup>185</sup>

One of the main reasons why McMahan’s account of the wrongness of killing – which divides humanity into two parts, persons and non-persons, and proposes one set of moral principles for one group, and a different set of moral principles for the other – is hard to accept for many, and plainly repugnant to some, is the fact that it clashes with the belief, which is inconsistent with line-drawing within humanity, that all human beings have equal intrinsic worth and hence an equal right not to be killed.<sup>186</sup> The equal wrongness thesis, which we introduced and discussed in the previous chapter, captures part of, and gets much of its intuitive appeal from, the belief that, as Ronald M. Dworkin has put it, “the life of a human organism has intrinsic value in any form it takes,”<sup>187</sup> but it is too narrow, in that it does not capture widely held intuitions about the morality of killing human non-persons, whom it excludes. McMahan himself is uneasy with the implications of his view with regard to infants and other human non-persons, but he nevertheless feels compelled to reject the idea that all human beings are equal. That is, because he believes that the equal dignity of all human beings, to the exclusion of most or all other animals, would have to be based on an intrinsic difference in immediately or nearly immediately exercisable cognitive capacity between all human beings and all or almost all non-human animals. Yet, as Jeremy Bentham has already noted in the eighteenth century, “a full-grown horse or dog is beyond comparison a more rational, as well as a more conversable

---

<sup>185</sup> The *Basic Law* is the constitutional law of Germany. Since it was regarded as provisional when drafted, the German word for “constitution,” *Verfassung*, was not used. For a historical, legal, and philosophical analysis of the concept of human dignity in German and Kenyan constitutional law, see Ebert & Oduor 2012.

<sup>186</sup> The term “person” here is used in the technical sense employed by McMahan, and refers “to any entity with a mental life of a certain order of complexity and sophistication” (McMahan 2002, p. 6).

<sup>187</sup> Dworkin 1994, p. 69.



animal, than an infant of a day or a week or even a month old.”<sup>188</sup> For every psychological capacity that could plausibly serve as a basis for full moral worth, be it rationality, self-consciousness, autonomy, or moral agency, there are human beings who possess that capacity to a lesser degree than some non-human animals whom we typically do not regard as our moral equals. Instead of giving up universal human equality, McMahan could have opted to allow all non-human animals who have a mind no less complex and sophisticated than that of a newborn or the most retarded human being into the community of moral equals, but he thinks an egalitarianism that inclusive would be absurd, as to him it “is uncontroversial that the killing of an animal is normally less seriously wrong than the killing of [you or me].”<sup>189</sup>

The fact that a philosopher of McMahan’s stature feels compelled to reject the idea of equal human worth, for which he has considerable sympathy, despite its simplicity and almost universal acceptance, is remarkable, and maybe surprising. Other philosophers do not feel so compelled, and instead assert that human equality can be justified.<sup>190</sup> They defend the position that, if other things are equal, the killing of one human being is as seriously wrong as the killing of any other human being, because every human being possesses full and equal human worth or dignity. This position comes in many forms, and different arguments have been offered in its defense.

Sometimes, the case for equal human worth takes a decidedly religious form. According to traditional Christian theology, for example, human beings are special among god’s

---

<sup>188</sup> Bentham 1789b, p. 143.

<sup>189</sup> McMahan 2002, p. 190. Also, shifting the threshold of personhood does not solve the problems, discussed in the previous chapter, of that threshold necessarily being arbitrary, and of basing a momentous difference in moral status on a seemingly insignificant difference in empirical reality.

<sup>190</sup> Cf. Lee & George 2008a & 2008b, George & Gómez-Lobo 2002, Gómez-Lobo 2002, Kass 2002 & 2008, Lee 2013, Rolston 2008, Sulmasy 2008, and Kumar 2008. The idea of equal human worth also plays an important role in many theories of human rights, cf., for example, Dworkin 1977, Finnis 1980, Kymlicka 1990, Nozick 1974, and Rawls 1972.

creation as only they have souls and are created in god's image – which, it is interesting and important to note, contrasts with Indian/Dharmic religious traditions that tend to take a more holistic approach and emphasize our continuity, community, and connectedness with other animals.<sup>191</sup> If “[f]aith is belief without reason,”<sup>192</sup> as the American essayist Roger Rosenblatt once wrote, there is not much to say about these claims by way of rational argument. But more than two thousand years of fierce scholarly debates about matters of religion are a testament to the fact that faith can also be, and often has been, understood as belief that is reasonable, or at least potentially reasonable, and hence neither indifferent to evidence, nor immune to rational criticism. Religious arguments for equal human worth have been and will continue to be examined with great skill.<sup>193</sup> Instead of joining that debate, however, we shall restrict our discussion to secular arguments, which are more likely to have a wide appeal and I believe have a better chance of being sound.

Secular arguments that attempt to make equal human worth plausible generally tend to follow one of two strategies, both of which, I will argue, have much to recommend them, but are also marred by considerable difficulties.

### ***3.1. Two strategies to justify equal human worth***

All human beings are equally members of the human species. One might hence be tempted to explain the supposed fact that all human beings have an equal moral status,

---

<sup>191</sup> For a recent collection of articles about the status of non-human animals in Asian religion and philosophy, see Dalal & Taylor 2014.

<sup>192</sup> Rosenblatt 1984, p. 112.

<sup>193</sup> A recent example for a rational examination of the idea that you and I are morally equal in virtue of having or being souls can be found in McMahan 2002, cf. pp. 7-19 & 209 f.

superior to that of all other animals, by appeal to our common humanity. That explanation, however, is incomplete, as bare membership in the human species is not necessary for membership in the community of moral equals, even though it might be sufficient. To see why that is so, consider the many ongoing scientific projects, most notably SETI, whose aim it is to detect intelligent extraterrestrial life, and suppose that, one day, one of these projects will be successful.<sup>194</sup> We will not only detect but be able to communicate with life on another planet. We will learn that our extraterrestrial counterparts are rational, maybe even more so than we are, shape their lives in an autonomous manner, and engage in complex behavior which they regulate according to self-imposed moral laws that are a frequent topic of sophisticated and critical debates on their planet. Surely, we would be wrong not to treat these extraterrestrials with the same respect with which we treat human beings, regardless of the fact that they are not members of our species (and maybe not even carbon-based life forms). If so, membership in the human species is not a necessary condition for full moral status. That is also why the mere observation that non-human hominidae – orangutans, gorillas, chimpanzees, and bonobos – are not human does not settle the on-going debate over whether they should be granted “human rights.”<sup>195</sup> A full explanation of why human beings and their intelligent extraterrestrial counterparts (and maybe also non-human hominidae) have full moral status, and the privileges that come with it, must specify what it is about all of them in virtue of which they have that status. They must have a property or set of properties in common that confers a special moral status, and entitles them to inclusion in the community of moral equals. Only if it can be established that membership in the human

---

<sup>194</sup> Cf. SETI Institute, URL = <<http://www.seti.org/>> [accessed on October 22, 2015].

<sup>195</sup> Cf. Cavalieri & Singer 1993.

species *guarantees* the possession of that property or set of properties can we truly say that all human beings have equal moral worth *because* they are human.

If it is not a soul or god-likeness that sets human beings apart from other animals, defenders of human dignity must propose another special-status-conferring property or set of properties that all human beings and no, or only few, other animals have, and that is what they have in fact been doing. As I see it, almost all such secular attempts to justify human dignity and equality can be understood as following one of two general strategies:

*Strategy One.* An individual has full moral status, if, and only if, he or she has a certain intrinsic property, or set of intrinsic properties, X. X is necessary for membership in the human species. Therefore, all human beings have full moral status, in virtue of having X.

*Strategy Two.* An individual has full moral status, if, and only if, he or she is essentially a member of a class of individuals, C, defined by one or more properties that are individually necessary and jointly sufficient for membership and have no direct moral significance, whose normal or paradigm members have a certain intrinsic property, or set of intrinsic properties, Y, which has direct moral significance.<sup>196</sup> Humanity is a class defined by the essential properties of human beings. Normal or paradigm members of the human species have Y. Therefore, all human beings have full moral status.

---

<sup>196</sup> If an individual is essentially a member of C, then the properties that are necessary and sufficient for membership in C are essential properties of that individual. An essential property of an entity is a property that makes that entity what it is: If Z is an essential property of A, A could not lack Z. If A loses Z, A necessarily ceases to exist. For example, being even is an essential property of the number two, as it is impossible for the number two not to be even. Properties that are not essential are accidental. Hence, if Z is an accidental property of A, it is possible for A to exist and not have Z. Cf. Robertson & Atkins 2013, which also contains a useful discussion of nuances in the meaning of the terms “essential” and “accidental,” and of doubts about the significance and meaningfulness of the distinction.

The first strategy affirms a principle that James Rachels has termed *moral individualism*. According to moral individualism, “[h]ow an individual should be treated depends on his or her own particular characteristics, rather than on whether he or she is a member of some preferred group [...]”<sup>197</sup> This distinction needs some clarification though, as being member of a group, of course, is itself a property. Hence, if how an individual should be treated depends on whether he or she is a member of some group, how that individual should be treated depends on his or her own particular characteristics, one of which is being a member of that group.<sup>198</sup> Rachels must mean something more specific than just any properties an individual might have when he talks of “own particular characteristics.” I think the most plausible interpretation takes the distinction to be that between having a moral status in virtue of having not just any but certain *intrinsic* properties, and hence independently of any group memberships, and having a moral status in virtue of being a member of some group.<sup>199</sup> Given this interpretation, the second strategy rejects moral individualism. Philosophers who adopt it acknowledge that some human beings do not have any intrinsic properties that are of direct moral significance, such as certain psychological capacities, but nevertheless grant them full moral status because they belong to a kind that is characterized by the fact that its members typically have intrinsic properties that are directly relevant for the attribution of full moral status. In this chapter, we will critically examine both strategies, starting with the first.

---

<sup>197</sup> Rachels 1990, p. 5.

<sup>198</sup> Further, he or she is a member of that group in virtue of those of his or her own particular characteristics that are the basis of his or her membership in that group.

<sup>199</sup> Somebody might object that these are mutually exclusive alternatives only if membership in the relevant group is a relational property, as is the case with national or ethnic groups, but not if the relevant group is humanity, and if membership in the human species is based on having a human genome, or some other intrinsic property. In response, we note that being a member of some group defined by certain intrinsic properties must be distinguished from having these intrinsic properties, even though the latter implies the former. Group membership is always a relational property, even in cases in which it is based on an intrinsic property.

### 3.2. *Strategy One: the argument from substantial identity*

Since the early 1990s, Patrick Lee and Robert P. George have been putting forward an argument that has gained significant traction in bioethics, and is exemplary for Strategy One.<sup>200</sup> They call it the *argument from substantial identity*. It was originally conceived as an argument for the impermissibility of abortion, and it goes like this:

- “1. You and I are intrinsically valuable (in the sense that makes us subjects of rights).
2. We are intrinsically valuable in virtue of what we are (what we are essentially), instead of in virtue of accidental characteristics.
3. What we are is each a human, physical organism. (We are human, physical organisms essentially.)
4. Human, physical organisms come to be at conception. [...]
5. Therefore, what is intrinsically valuable (as a subject of rights) comes to be at conception.”<sup>201</sup>

If sound, though, this argument shows not only that human beings in the early stages of their lives have rights, in particular the right not to be killed, but that all human beings have full moral status.<sup>202</sup> If you and I, and the early-stage human beings we once were, are intrinsically valuable in virtue of being human organisms, then all human beings are intrinsically valuable, for the same reason.<sup>203</sup>

---

<sup>200</sup> Cf., e.g., Lee 2004, and Lee & George 2007.

<sup>201</sup> Lee & George 2007, p. 134.

<sup>202</sup> In fact, in recent years, Lee and George have presented versions of this argument for the expressed purpose of establishing equal human worth, cf. Lee & George 2008a & 2008b, and Lee 2011 & 2013.

<sup>203</sup> It speaks in favor of the position of Lee and George that it avoids the uncomfortable underinclusiveness of McMahan’s morality of respect. However, we should mention that their position, to a lesser extent, too is at odds with common sense. While there is widespread agreement that all postnatal human beings have full

Let us go through the four premises of the argument, one by one. We start with the last premise, which asserts that human organisms come to be at conception. Until about fourteen days after conception, monozygotic twinning can occur. In monozygotic twinning, a single zygote divides into two zygotes. If that happens, and if everything goes well, both zygotes grow into fetuses, and – eventually – two healthy children are born. None of these children can be numerically identical to the freshly-fertilized ovum which is their common origin.<sup>204</sup> If both children were numerically identical to the original, pre-twinning zygote, then they would be numerically identical to each other, as identity is transitive. But, plainly, the twins are two distinct human organisms. The original zygote can also not be numerically identical to only one of the twins, as there is no principled way to pick one twin rather than the other. There is no relevant difference in quality that would allow us to single out one of the two post-twinning zygotes. Therefore, the fourth premise is false. Not all human organisms come to be at conception. At least some human organisms come to be when monozygotic twinning occurs. One way to respond is to say that human organisms do not come to be at conception but when a human zygote passes the point in its development at which monozygotic twinning becomes impossible. If so, zygotes during their first fourteen-or-so days are not complete organisms but mere collections of cells.<sup>205</sup> The argument from substantial identity still stands, except that it now no longer attributes full moral status to very early zygotes, which I think would be easy enough to accept. Another way to respond is to replace the fourth premise with the claim that there are not one but two kinds of events at which human organisms come to

---

moral status, there is no such agreement with regard to fetuses. The moral status of fetuses remains a matter of great controversy. Hence, some will find the position of Lee and George overinclusive.

<sup>204</sup> This point is also made in McMahan 2002, cf. pp. 25 f., and in DeGrazia 2005, cf. pp. 247 ff.

<sup>205</sup> Peter Singer and Helga Kuhse go this way in Singer & Kuhse 1990, cf. p. 67.

be – conception and monozygotic twinning. If so, a human organism with full moral status, the original zygote, ceases to exist whenever monozygotic twinning occurs, making way for two new human organisms. This has the rather unsettling implication that it is the moral equivalent of killing you or me to intentionally induce monozygotic twinning, and that something terrible happens whenever monozygotic twinning naturally occurs, for the prevention of which we are typically willing to spend significant resources: One of us ceases to exist. Also, it is odd to think that some of us started out as freshly-fertilized ova, while others – those who are identical twins – started out as zygotes of one or two weeks, even though we are all of the same kind.<sup>206</sup> Fortunately, we do not need to make a choice between these two options. It is sufficient to note that there *are* options that leave the substance of the argument from substantial identity intact.

The third premise asserts that you and I are essentially human organisms. Even though I suspect this seems obvious to most people, I think it is false, for reasons I will provide in the next chapter. For the purpose of our discussion of the joint view of Lee and George, however, I will assume that each one of us is in fact numerically identical with a human organism, as I think that their argument fails independently of that assumption. The first premise we shall grant as well. If anybody is intrinsically valuable, you and I are.

This leaves us with the second premise, which asserts that you and I are intrinsically valuable in virtue of what we are essentially, rather than in virtue of accidental properties. Philosophical accounts of moral worth that accommodate this assertion resonate with the common intuition that our worth is not a fleeting attribute you and I just happen to have, but a permanent fixture of our existence that is intimately bound up with who we are –

---

<sup>206</sup> I suppose there is also a third option, which is to say that a zygote that will divide into two separate zygotes somehow constitutes two human organisms.



which speaks in their favor, as we have argued in the last section of the previous chapter. But we have also noted that it is far from self-evident that our intrinsic worth supervenes on essential properties. There is no logical contradiction in holding, as McMahan and others in fact do, that we are intrinsically valuable in virtue of accidental properties. The second premise can therefore not just be assumed; it has to be made plausible, beyond merely being intuitive in itself. If we are essentially human organisms, and if, as we have urged above, and as Lee and George readily acknowledge, “membership in the species *Homo sapiens* [...] is not in any direct sense the criterion for moral worth [...]”<sup>207</sup> then for it to be true that we are intrinsically valuable in virtue of essentially being human organisms there must be a property or set of properties that is necessary for being a human organism and, at the same time, sufficient for full moral worth. A full explanation of our moral worth – of which the argument from substantial identity, in the above-stated form, is only a part – needs to specify that property or set of properties.

According to Lee and George, the empirical property that grounds full moral worth, and distinguishes us from other animals, is the possession of “natural capacities for conceptual thought, deliberation, and free choice [...]”<sup>208</sup> These capacities, Lee and George maintain, are “possessed by human beings in all developmental stages, including the embryonic, fetal, and infant stages, and in all conditions, including severely cognitively impaired conditions [...]”<sup>209</sup> Similar claims have been made by a number of other philosophers. Alfonso Gómez-Lobo, for example, writes in a joint statement with George that

---

<sup>207</sup> Lee & George 2008b, p. 176.

<sup>208</sup> Lee & George 2008a, p. 410.

<sup>209</sup> Lee & George 2008b, p. 175.

“[o]f course, human beings in the embryonic, fetal, and early infant stages lack immediately exercisable capacities for mental functions characteristically carried out [...] by most [...] human beings at later stages of maturity. Still, they possess in radical [...] form these very capacities. Precisely by virtue of *the kind of entity they are*, they are from the beginning actively developing themselves to the stages at which these capacities will (if all goes well) be immediately exercisable [...]. Each new human being comes into existence possessing the internal resources to develop immediately exercisable characteristically human mental capacities – and only the adverse effects on them *of other causes* will prevent their full development. In this sense, even human beings in the embryonic, fetal, and infant stages have the *basic natural* capacity for characteristically human mental functions.”<sup>210</sup>

Lee, George, and Gómez-Lobo agree with McMahan in that they too tie full moral worth to certain higher mental functions. But, unlike McMahan, they do not require that those who possess full moral status be capable of now performing these higher mental functions.<sup>211</sup> Instead, their proposed criterion for full moral worth is the possession of *basic natural capacities* for such higher mental functions as conceptual thought and deliberation. But what precisely is meant by “basic natural capacity”? The implicit and explicit answers to this question that can be found in a number of places in the work of Lee, George, and Gómez-Lobo are mostly vague, and at times even seem to be

---

<sup>210</sup> George & Gómez-Lobo 2002, p. 260. Also, see Gómez-Lobo 2005, p. 109. Another example is John Finnis who claims that all human beings possess “the radical capacity to deliberate and choose” (Finnis 1995, p. 31). For further examples, see Ray 1985, and Reichlin 1997.

<sup>211</sup> I will use the term “higher mental functions” as a placeholder for the various mental functions that proponents of the argument at hand consider relevant, such as conceptual thought, deliberation, and rationality.

inconsistent with each other. That is unfortunate, as precision here is sorely needed. We will consider separately a number of different ways in which the concept of basic natural capacity can be narrowed down. In doing so, our aim is not exegetical. We will draw from the published work of Lee, George, and Gómez-Lobo, as well as others, not to find a definite answer to the question of which view either one or all of them in fact hold, but to identify possible versions of the argument from substantial identity, with the aim to evaluate the plausibility of each one of them.

At times, it seems clear that Lee, George, and Gómez-Lobo regard the basic natural capacity for higher mental functions as a potential. In a 2002 article, for example, George and Gómez-Lobo refer to the basic natural capacity for higher mental functions as a *“potentiality that the human being possesses simply by virtue of the kind of entity it is.”*<sup>212</sup> In another article, published three years later, Gómez-Lobo attributes to human embryos *“the remote potentiality to later exercise the typically human activities [...]”*<sup>213</sup> and identifies this as the feature that qualifies them for full moral status and entitles them to respectful treatment, and Lee explicitly likens the distinction between an immediately exercisable capacity and a basic natural capacity to the more familiar distinction, popularized by Michael Tooley, between a “capacity” and a mere “potentiality,”<sup>214</sup> and describes the basic natural capacity for higher mental functions as *“a capacity to develop oneself to the point where one does perform [...] [higher mental functions],”*<sup>215</sup> that is, as a potential.

---

<sup>212</sup> George & Gómez-Lobo 2002, p. 261.

<sup>213</sup> Gómez-Lobo 2005, p. 105.

<sup>214</sup> Cf. Lee 2003, p. 80 n. 9.

<sup>215</sup> Lee 2004, p. 253.

If the basic natural capacity for higher mental functions is understood to be some kind of potential for higher mental functions, then the argument from substantial identity is an argument from potential. Arguments of that kind are a familiar staple of the abortion debate. Importantly, though, the argument at hand avoids the charge, often levelled against arguments from potential, that there is a logical mistake in inferring full moral status from the merely potential possession of properties that are the basis for full moral status, as full moral status requires actual possession of these properties. To use an illustration by Stanley I. Benn, “[a] potential president of the United States is not on that account Commander-in-Chief.”<sup>216</sup> What is claimed here, however, is not that the immediate capacity for higher mental functions is the basis for full moral worth, and that infants and other human beings who do not currently have that capacity nevertheless have full moral status because they have the potential for higher mental functions. Rather, the potential for higher mental functions *itself* is proposed as a basis for full moral status, and the standard charge against arguments from potential hence does not apply.

Another possible objection too can be disposed of rather quickly, and will help us further clarify the position. If a zygote deserves special protection because of its potentiality, one might ask, does a set of an ovum and a sperm that will combine and then grow into an adult human being not deserve the same protection, for the same reason? After all, such a set also has, as Lee says is true of the human embryo, “within [...] [itself] all of the positive reality needed to actively develop [...] [itself] to the point where [...] [it] will perform higher mental functions, given only a suitable environment and nutrition [...].”<sup>217</sup> If so, it seems to follow that it is as seriously morally wrong to destroy a sperm

---

<sup>216</sup> Benn 1973, p. 102.

<sup>217</sup> Lee 2004, p. 253.

and an ovum that are about to combine as it is to kill an innocent adult human being, if other things are equal. This conclusion, I think, would be absurd, but there is a very plausible way to avoid it. Notice that the potential of sperm and ovum is of a fundamentally different kind than the potential of the human embryo. When sperm and ovum combine to form a zygote, they both cease to exist. Hence, their potential to become a human being capable of performing higher mental functions is like the potential of a tree to become a pile of ashes. In contrast, the human embryo has the potential to become a human being capable of performing higher mental functions, to whom the human embryo will be numerically identical. The embryo will become and continue to exist as an adult human being. The embryo and the adult human being are one and the same human organism. Borrowing a term from McMahan, the embryo's potential is *identity-preserving*, the potential of sperm and ovum is not.<sup>218</sup> We shall hence assume, as is sensible, that the potential relevant to membership in the community of moral equals is of the identity-preserving kind.

The position we are considering holds that all human beings have full moral status in virtue of having an identity-preserving potential for higher mental functions. Is it true that all human beings have such a potential, and, if so, what is it about each and every human being that makes it true? The only category of human beings for which it seems obviously true that they have a potential for higher mental functions are normal embryos, fetuses, and infants. Given a suitable environment and nutrition, they will perform higher mental functions in later stages of their lives, and – plainly – one has the potential to

---

<sup>218</sup> Cf. McMahan 2002, pp. 302 ff. In response to a similar challenge, Jim Stone draws a similar distinction, between *weak* and *strong potentiality*, and argues that strong but not weak potentiality is sufficient to establish a right to life. Strong potentiality requires that for A to be a potential B, A will be identical to B if the potential is realized. Cf. Stone 1987, pp. 816-820.

become what one in fact will become. Let us quickly move on then, and turn our attention to the case of adult human beings, which Lee, George, and Gómez-Lobo seem to believe is equally clear. But is it?

On what basis can it be claimed that normal adult human beings have a potential for higher mental functions? According to Tooley,

“[t]o attribute a certain potentiality to an entity is to say at least that there is a change it could undergo, involving more than the mere elimination of factors blocking the exercise of a capacity, that would result in its having the property it now potentially has. It may also be to say that there are now factors within the entity itself that will, if not interfered with, cause it to undergo the relevant change.”<sup>219</sup>

Given this minimal definition, as well as the common sense meaning of the term, potential involves a possible change of what is now into what may become in the future. To say that somebody potentially has a property implies that he or she does *not yet have* that property but might acquire it later on. It is for that reason that it would sound very strange were Barack Obama to proclaim that he has the potential to be the President of the United States. He already is the President of the United States. This suggests that normal adult human beings do not have a potential for higher mental functions, as they already have the capacity for higher mental functions.<sup>220</sup> They do not need to undergo any change to be able to engage in conceptual thought or rational deliberation. The embarrassing implication would be that normal adult human beings do not have full moral status. Confronted with a related objection by Don Marquis, Lee concedes in a

---

<sup>219</sup> Tooley, p. 150.

<sup>220</sup> Carson Strong makes the same point in Strong 2006, cf. pp. 443 ff.

2011 article that he has “expressed [...] [his] argument confusedly”<sup>221</sup> in previous works. While much of his writings make it seem like he regards the immediately exercisable capacity for higher mental functions and the basic natural capacity for higher mental functions as two strictly distinct properties of human organisms, he in fact has always held, he says, that the former “is not actually a new and different capacity [...] [but] only a *degree of actualization or development*”<sup>222</sup> of the latter.<sup>223</sup> Along the same line, George and Gómez-Lobo write that “the difference between these two types of capacity is merely a difference between stages along a continuum.”<sup>224</sup> It is hard to say what to make of this. There seems to be a clear distinction between an immediately exercisable capacity for higher mental functions and a potential for higher mental functions that is categorical rather than gradual. The difference between a normal adult and a zygote is not the difference between one who is more and another who is less capable of writing poetry, but the difference between having a fully developed brain that supports poetry-writing and merely having the internal resources that are necessary for such a brain to develop in the future. It is neither true that all human beings possess the property of having an immediate capacity for higher mental functions, nor that all human beings possess the property of having a potential for higher mental functions in the standard sense of a potential to acquire the not-yet-present immediately exercisable capacity for higher mental functions. You and I do not possess the latter property, and a human fetus does not possess the former, which shows that none of the two properties is necessary for

---

<sup>221</sup> Lee 2011, p. 37 n. 12.

<sup>222</sup> Lee 2011, p. 27.

<sup>223</sup> To me, this looks more like a revision than a clarification. There are a number of places in Lee’s writings in which he unambiguously asserts that the immediately exercisable capacity for higher mental functions and the basic natural capacity for higher mental functions are two distinct sorts of capacity, cf., e.g., Lee 2004, p. 253.

<sup>224</sup> George & Gómez-Lobo 2002, p. 261.

membership in the human species. Lee, George, and Gómez-Lobo, however, need the property of having the basic natural capacity for higher mental function to be a property that *all* human beings have essentially. Therefore, if the basic natural capacity for higher mental functions is to be a potential, we need to move away from the standard understanding of potentiality that is expressed in the above-quoted definition by Tooley, and ask whether there is another sense in which both normal adult human beings and normal fetuses have the potential for higher mental functions. I think there is.

It is typically true for both adult human beings and human fetuses that, if normal circumstances continue to obtain, there will come a point in the future at which, if they were then in a certain condition, they would perform higher mental functions.<sup>225</sup> For fetuses, that point is years away. For you and me, it is usually the very next moment, unless we are under the influence of a drug, or there is some other external factor, that blocks the immediate exercise of our capacity for higher mental functions. Another claim, which is or is not weaker, depending on which notion of possibility is applied, also seems to be true: It is typically possible for both adult human beings and human fetuses that they will perform higher mental functions. Both claims attribute a common property to adult human beings and human fetuses that is recognizably a potential insofar attribution of that property says something about the properties an individual may have in the future. If all human beings are to be intrinsically valuable in virtue of the essential property of having the potential for higher mental functions, that potential cannot plausibly be said to be the kind of potential that Tooley calls “potentiality” but instead needs to be broadly

---

<sup>225</sup> In the case of early fetuses, the term “normal circumstances” refers, e.g., to the environment of the womb. The “certain conditions” are things like being awake, having made a choice or decision to perform some higher mental function, etc.



construed in one of the two ways just outlined, or some other way that is in the same ballpark.

With some effort, we have shown that the claim that normal adult human beings and normal fetuses both possess the potential for higher mental functions can be made plausible. Can we similarly find something in the makeup of all other human beings that would justify us in making the same claim about them? There are (at least) two groups of human beings for which this seems doubtful: those who never had and never will have the immediately exercisable capacity for higher mental functions, and those who used to have but have permanently lost that capacity. The first group contains, for example, anencephalic infants who are born without the cerebrum. As the cerebrum is necessary for cognition and consciousness, the immediate mental capacity of an anencephalic infant will never surpass that of many non-human animals, and will certainly never include a capacity for rationality, conceptual thought, and the like. The second group contains, for example, those who have suffered severe brain damage that has left them in an irreversible vegetative state, and human beings with late-stage Alzheimer's disease. For them, as for anencephalic infants, it is not true that, if normal circumstances continue to obtain, there will come a point in the future at which they will be able to perform higher mental functions. Is there still a sense in which these individuals currently have the potential for higher mental functions?

There is no known medical treatment or cure for anencephaly. Almost all babies with anencephaly die shortly after birth. Similarly, certain physical injuries to the brain irreparably destroy its capacity to support higher mental functioning, and so does Alzheimer's disease, given the current state of medical science. Yet, it is conceivable that

sometime in the future it will become practically possible to treat all three conditions, which would mean that treatment is now physically possible. Just like the loss of a kidney was irreversible seventy years ago, whereas today kidney transplantations are frequently successfully performed, the day might come when doctors will be able to get anencephalic and brain-damaged human beings, and human beings suffering from late-stage Alzheimer's disease, to the point where they will perform higher mental functions. In the case of anencephalic fetuses, that might take the form of a genetic therapy that stimulates cerebral growth, and the form of partial brain transplants in the case of anencephalic babies and those who have suffered severe brain damage or are suffering from Alzheimer's disease. If done with a view to make the argument from substantial identity work, however, relying on a notion of potential built on the supposed fact that it is physically possible for every human being to acquire the immediate capacity for higher mental functions in an identity-preserving way does not seem promising.<sup>226</sup> If human beings who have lost those parts of the brain that are necessary for higher mental functioning have the relevant potential in virtue of it being physically possible to restore their previous cognitive capacities through partial brain transplantation, then surely many non-human animals too have that potential. While we may have reasonable doubts about insects and other animals to whom we have little structural similarity, I can think of no convincing reason why there should be a principled obstacle that will prevent doctors from ever supplying, say, a lion with additional brain matter, through transplantation, so as to enable the lion to perform higher mental functions. If there is in fact no such

---

<sup>226</sup> Alan R. White seems to suggest this unpromising strategy. He says that "infants, children, the feeble-minded, the comatose, the dead, or generations yet unborn [...] may be for various reasons empirically unable to fulfil the full role of a right-holder. But so long as they are persons [...] they are logically possible subjects of rights to whom the full language of rights can significantly, however falsely, be used" (White 1984, p. 90).

obstacle, one would have to concede that lions are on an equal moral footing with human beings. That, however, is exactly what proponents of human dignity vehemently deny. It is true that anencephalic fetuses and infants and human beings with certain neurological diseases or injuries have a potential for higher mental functions only if one assumes an overly inclusive notion of potential which brings animals under the scope of the argument from substantial identity whom it was designed to exclude.

But maybe I am wrong and there is a reasonable conception of potentiality, according to which those with congenital defects that practically preclude the development of an immediate capacity for higher mental functions and those who have permanently lost that capacity are potentially capable of performing higher mental functions, whereas all or almost all non-human animals are not. If so, there remains the question of why we should accept a notion of potential for higher mental functions broad enough to encompass these human beings, who are far removed from any such functions, as a basis for moral worth. Why should how we treat somebody *now* be determined by his or her more or less remote potential to develop higher mental capacities? There is a plausible answer to this question in some cases, for example, when a woman uses drugs during pregnancy and thereby risks significant harm to the adult human being into whom the fetus will grow and who will have an immediate capacity for rationality. But, if using drugs is prohibited in such a case, that need not be because the fetus has full moral status, or any moral status at all for that matter. A prohibition against harming fetuses can as well be explained indirectly by appealing to the harmful effects of harming fetuses on the adults they grow into. If the fetus will not grow into a being immediately capable of conceptual thought and free choice, say, because the fetus is killed before that can happen, then why should the

morality of respect apply just because the fetus otherwise would or could have developed into our cognitive equal?<sup>227</sup> The relevance of potential is even less persuasive in the cases of severely congenitally mentally retarded infants and those who have permanently lost the capacity for higher mental functions, where the potential for higher mental functions is merely theoretical. I can think of no good reason to believe that these human beings have the same moral status as you and I *just because* they have a very remote potential for higher mental functions that will never be realized. This is not to say that human beings who never had and never will have the immediately exercisable capacity for higher mental functions, or have permanently lost that capacity, are unequal in moral worth, but that the attempt to establish their equality drawing on the supposed moral importance of higher mental functions is unconvincing, and reeks of pro-human bias.

### 3.3. *The genetic basis for moral agency account of rightholding*

It seems unlikely that there is a strong case for human equality which bases dignity on the potential for higher mental functions, because, given any reasonable conception of potentiality, either not all human beings have the potential for higher mental functions, or many non-human animals do. Therefore, if one is to make it plausible that anencephalic and severely brain-damaged human beings, and those suffering from late-stage Alzheimer's disease, are closer – in a morally relevant sense – to the capacities people commonly associate with the heightened moral status of individuals like you and me, such as rationality, autonomy, or self-consciousness, than all or almost all non-human

---

<sup>227</sup> And why should rationality be relevant to moral worth in the first place? I think it is not, and I will suggest an alternative factual basis for moral worth in the final chapter of my dissertation.

animals, one would be well-advised to find some empirical difference other than a difference in potential. It is not immediately obvious what that difference could be, but we would be too quick to reject human exceptionalism just yet. The earlier-quoted passage taken from a joint statement by Gómez-Lobo and George, which contains the claim that

“[e]ach new human being comes into existence possessing the internal resources to develop immediately exercisable characteristically human mental capacities – and only the adverse effects on them *of other causes* will prevent their full development [...],”<sup>228</sup>

suggests another possible defense of human dignity that does not rely on some esoteric concept of potential but instead appeals to an actual attribute that they say all human beings share, and that more readily and plausibly lends itself to empirical assessment. Human dignity, one might object to our discussion so far, is not grounded in the potential for higher mental functions *itself* but in its biological or physical basis. In the case of most human beings, that distinction does not matter and hence is easily overlooked, as most human beings clearly are potential conceptual thinkers.<sup>229</sup> But in the case of certain atypical human beings, some kinds of which we have explicitly discussed, the distinction is crucial. While a previously normal human being in an irreversible vegetative state might no longer have the potential for higher mental functions, it is still true that he or

---

<sup>228</sup> George & Gómez-Lobo 2002, p. 260. Also, see Gómez-Lobo 2005, p. 109. Lee makes a similar claim: “[M]embers of [...] [the human] species come to be with whatever it takes to develop [...] [the] immediately exercisable capacity [for higher mental functions], given a suitable environment and nutrition. [...] The human embryo has within herself all of the positive reality needed to actively develop herself to the point where she will perform higher mental functions, given only a suitable environment and nutrition, and so she now has the natural capacity for such mental functions” (Lee 2004, p. 253).

<sup>229</sup> As we have argued above, it is less clear than is commonly assumed though, and only true if one assumes a non-standard meaning for the word “potential,” because only then normal adult human beings can be said to have a potential for conceptual thought.

she came “into existence possessing the internal resources to develop immediately exercisable characteristically human mental capacities [...]”<sup>230</sup> The most plausible explanation of what makes that true, I think, must appeal to the human genome. It is in virtue of having a certain genetic makeup that human beings at the earliest stages of their existence have the potential for higher mental functions. That potential might be irreversibly thwarted at a later stage, by disease or injury, but its genetic basis remains. It is because men and women in irreversible vegetative states still have the genetic basis for higher mental functions that their gametes can be artificially united with gametes of the opposite sex to form zygotes which can grow into normal adults who will perform higher mental functions. There are a few indicators in the work of Lee, George, and Gómez-Lobo which suggest that this might actually be the account of human dignity they mean to defend, but I think their frequent reference to potential and capacity makes that an unlikely interpretation.<sup>231</sup> However, it has recently been developed by another philosopher, Oxford University’s S. Matthew Liao, who calls it the *genetic basis for moral agency account of rightholding*, and it well deserves consideration.<sup>232</sup> Liao proposes that “all human beings are rightholders because they all have the genetic basis for moral agency [...]”<sup>233</sup> As he makes clear with the following analogy, his proposal does not rely on any claim about cognitive potential and hence is not an argument from potential.

---

<sup>230</sup> George & Gómez-Lobo 2002, p. 260.

<sup>231</sup> Here, for example, are Lee and George writing about human beings: “[T]hey are structured [...] in such a way that they are oriented to maturing to [the stage at which they will perform acts of conceptual thought]. The genetic structure orients them toward developing a complex brain that is suitable to be the substrate for conceptual thought” (Lee & George 2008, p. 185). From this, they conclude that “every human being [...] has [...] [the] basic natural capacity for conceptual thought” (Lee & George 2008, p. 185).

<sup>232</sup> Cf. Liao 2010, 2011 & 2015, Chapter 1.

<sup>233</sup> Liao 2010, p. 164.

“[S]uppose Joe’s hands were accidentally sawed off in a wood factory. We would say that Joe no longer actually has hands and that Joe does not have the potential to have hands. However, we can still say that Joe has the genetic basis for the development of hands, because Joe still has the genes for the development of hands.”<sup>234</sup>

Just as Joe has not lost the genetic basis for the development of hands by losing his hands, and thereby his potential to have hands, given any reasonable sense of the word “potential,” he would not lose the genetic basis for moral agency by permanently losing those parts of his brain which are necessary for moral agency, even though we would want to say that he would no longer have the potential for moral agency. If all human beings in fact have the genetic basis for moral agency, Liao’s account supports the widely held belief that all human beings have the same moral status. But do they? McMahan has serious doubts. He thinks that there are human beings who lack the genetic basis for moral agency.<sup>235</sup> I am not so sure. It is hard to tell. According to a guide published by the United States National Library of Medicine, anencephaly, for example, “is likely caused by the interaction of multiple genetic and environmental factors [...],”<sup>236</sup> many of which remain unknown. Vitamin B9 appears to be one of the factors that play a role. “Studies have shown that women who take supplements containing [Vitamin B9] [...] before they get pregnant and very early in their pregnancy are significantly less likely to have a baby with [...] anencephaly.”<sup>237</sup> That means that at least some cases of anencephaly are the

---

<sup>234</sup> Liao 2010, p. 171.

<sup>235</sup> “There are, it seems, human beings [...] in whom the genes that direct the development of those regions of the brain necessary for the capacities for deliberation and choice are either absent or defective” (McMahan 2002, p. 469).

<sup>236</sup> United States National Library of Medicine 2015.

<sup>237</sup> United States National Library of Medicine 2015.

result of Vitamin B9 deficiency and do not involve the absence of a genetic basis for moral agency. In cases where anencephaly is caused by genetic factors, if there are any such cases, it might still be plausible to maintain that the affected individuals have the genetic basis for moral agency.<sup>238</sup> Often, genetic defects that impede the normal development of the brain consist not in a lack of, or defect in, one or more of the genes that contain the information needed for the normal development of the brain, but in the mutation of an unrelated gene needed for the production of a certain protein or enzyme, which results in inadequate prenatal nutrition and, in turn, negatively affects the development of the brain. Examples for conditions where this is how mental retardation comes about are Phenylketonuria (PKU), Tay-Sachs disease, and Sandhoff disease.<sup>239</sup> Whether there any human beings who do not have the genetic basis for moral agency, maybe in virtue of lacking certain genes that direct the development of the cognitive capacities necessary for moral agency, or a lack of proper coordination between such genes, seems to be an open question. If there are in fact no human beings who lack the genetic basis for moral agency, one can still ask whether there *could* be human beings who lack the genetic basis for moral agency. It is conceivable that one could genetically engineer an embryo to be relevantly genetically defective. McMahan thinks that the “resulting radically impaired individual would certainly be a human being by any reasonable criterion of species membership [...],”<sup>240</sup> which would imply that human beings do not essentially have a genetic basis for moral agency. Again, I do not share

---

<sup>238</sup> It might be more appropriate to talk about *a* genetic basis for moral agency rather than *the* genetic basis for moral agency, as “it could be the case that in certain environments, certain genes, A, B, C, might be the genetic basis for moral agency for a particular being, while in certain other environments, genes D, E, F would be the genetic basis for moral agency in the same being” (Liao 2010, p. 165).

<sup>239</sup> That is, according to Liao, cf. Liao 2010, p. 167. I take his word for it.

<sup>240</sup> McMahan 2008, p. 90.



McMahan's confidence. Thankfully, we can avoid taking a position on what makes an organism a human being, as there are sufficiently strong independent reasons to reject Liao's proposal. For the sake of argument, we shall hence assume that all human beings essentially have the genetic basis for moral agency.

As we have noted above, it speaks in favor of Liao's account that it supports the immensely popular belief in human dignity and equality. Is that reason enough to believe that it is true? Animal rights advocates and others who are deeply skeptical about human exceptionalism will likely perceive Liao's account as moral gerrymandering no more persuasive than efforts to ground human dignity in some obscure potential. To them, it will seem like a rather desperate attempt to uphold the status quo, and consequently further strengthen the beliefs they came with. If they are to forgo their skepticism about human exceptionalism, Liao must provide independent reasons why the genetic basis for moral agency endows those who have it with a moral status superior to that of those who do not – which he does not, or only to a very limited extent.<sup>241</sup> Unless you are already convinced that all human beings have a moral status superior to that of all or almost all other animals, the claim that mere genetics makes all the difference between, say, a normal adult lion and a human being who happens to have, always had, and always will have cognitive capacities inferior to those of the lion will hardly have intuitive appeal, and just seem like speciesism in disguise. In contrast, if human dignity is one of the bedrocks of your moral worldview, as is probably true of most people, even though not always in practice, Liao's account might have considerable appeal.

---

<sup>241</sup> Liao mentions, e.g., that his account can handle McMahan's case of the Superchimp, cf. Liao 2010, p. 168.

Independently of its limited intuitive appeal, and more importantly, Liao's account faces a line-drawing problem that is similar to the one we have identified for McMahan's two-tiered account in the previous chapter. Let us start with the observation that there are different senses of moral agency. According to Liao, moral agency is "the capacity to act in light of moral reasons."<sup>242</sup> Other philosophers use a more restrictive definition, further requiring that a moral agent be capable of justifying his or her actions with reference to moral reasons.<sup>243</sup> Depending on which one of the many possible definitions is adopted, more or fewer non-human beings will turn out to be moral agents. David DeGrazia has gathered an impressive range of anecdotes and other evidence which suggest that chimpanzees and cetaceans might well meet plausible and defensible criteria for being a moral agent.<sup>244</sup> Dolphins, for example, are known to sometimes come to the help of drowning sailors, which some have interpreted as intentional assistance, for moral reasons. Others suspect that this interpretation is clouded by anthropomorphism, and suggest that the dolphins' behavior is nothing more than "cetacean volleyball," to which it is in turn objected that there are disproportionately fewer reports about dolphins pushing around unendangered swimmers than people in danger.<sup>245</sup> Whatever is the most accurate description of what dolphins in fact do when they save people in danger at sea, there likely is a plausible sense of moral agency according to which dolphins, as well as all other non-human animals, are not moral agents. Dolphins lack the language skills that seem required to engage in sophisticated moral deliberation, be it in their own minds or in a group with their conspecifics. However, while there are different senses of moral

---

<sup>242</sup> Liao 2010, p. 164.

<sup>243</sup> Cf., e.g., Sapontzis 1987, p. 35.

<sup>244</sup> Cf. DeGrazia 1996, pp. 199-204.

<sup>245</sup> Cf. DeGrazia 1996, p. 201.

agency, some of which effectively exclude all or – at the very least – almost all non-human animals, all senses have in common that they refer to a rather complex capacity, which is grounded in rationality and other psychological capacities that are not all-or-nothing matters. Like other higher mental functions that are commonly associated with the traditional notion of the person and assumed to be the direct or indirect empirical ground of human dignity, moral agency admits of degrees. Accordingly, any definition of moral agency that purports to distinguish human beings from other animals must be the definition of a threshold concept that separates a continuous multidimensional spectrum of cognitive capacity into two complementary parts. Children gradually develop the capacities that make up moral agency. While they grow older, the range of moral reasons to which they are responsive continuously expands, and they more and more engage in more and more robust and critical moral deliberation. Since there are no jumps in the mental development of children, their becoming moral agents, regardless of which sense of the term is deemed relevant, is a rather insignificant matter. They move from one place in a continuum to another, which is only infinitesimally different from it. This proved problematic for McMahan, who directly grounds full moral status in cognitive capacity, as such grounding requires drawing a line that is necessarily arbitrary, hence leaving us with an undesirable gray area, as there is no principled way to decide when exactly a child acquires full moral status, and bases a momentous moral difference on an entirely unremarkable difference in empirical reality. At first glance, it might seem like Liao's proposal avoids this difficulty. Among presently-existing animals, there need not be any gray areas, as all human beings might well unambiguously have the genetic basis for a high degree of moral agency which all or almost all other animals unambiguously lack.

Also, if it is the genome that determines moral status and not capacity, there is no implausibly abrupt discontinuity in moral status in the development of children. Human beings have the genetic basis for moral agency and hence full moral status throughout their existence. The problem remains, however, as all this does not change the fact that the property of having the genetic basis for moral agency too comes in degrees.



Illustration: Carl Buell, obtained from Wong 2013

The picture above is artist Carl Buell's impression of the last common ancestor of all placental mammals, which is based on the findings of a study that reconstructed the anatomy of the animal by comparing and analyzing 4,541 physical traits in 86 fossil and living species.<sup>246</sup> We may assume that this tree-climbing, small insect-eating animal, a common ancestor of animals as diverse as rats, whales, and human beings, had no significant degree of moral agency. In contrast, you and I have a high degree of moral agency. In between lie dozens of millions of years and a very large number of generations. Now imagine the line of your ancestors, going back to the animal in the picture. Due to the gradual and mostly slow nature of evolution, any two individuals

---

<sup>246</sup> Cf. O'Leary et al. 2013.

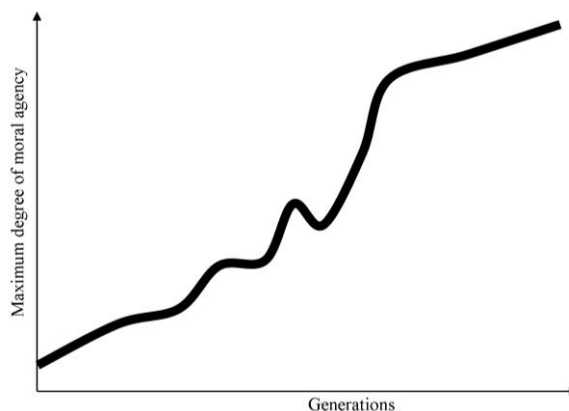
placed next to each other on that line will be practically indistinguishable in terms of cognitive capacity. Yet, as we just noted, the cognitive difference between the two individuals at the ends of the line is immense. While the development of mental functions in individual animals also significantly depends on environmental factors, the differences in cognitive capacity among your ancestors cannot be explained without reference to their respective genetic makeups. There is no environment in which one of your ancestors who lived, say, forty million years ago would have acquired a capacity for sophisticated moral reasoning, as the genetic information necessary for the development of a brain capable of supporting such a complex mental function was simply not there. The best way to describe the line of animals we are looking at, taking into account the influence of both environment and genetics on the development of cognitive capacities in individuals, is to say that, for each animal in the line, there is a maximum degree of moral agency he or she can achieve.<sup>247</sup> To illustrate this, let us imagine there was a linear scale which allows us to put a number to an animal's degree of moral agency, like the IQ scale for intelligence. Strictly speaking, this is probably too simple, as moral agency is a complex capacity that can vary along multiple dimensions, but for the point we are about to make we can ignore that complication. The greater an animal's capacity for moral agency, the higher he or she scores on the scale of moral agency. Suppose your genome would have enabled you to reach the one hundred mark on that scale, if ideal conditions had continuously obtained since you came into existence.<sup>248</sup> Your great-grandmother's score was probably in the same neighborhood, maybe at 95, maybe at 105. Evolution is not strictly progressive with

---

<sup>247</sup> As it is possible for every animal to lose the whole of his or her cognitive potential before the brain develops to the point at which mental functions start to emerge, there is no (greater-than-zero) minimum degree of moral agency he or she is guaranteed to achieve.

<sup>248</sup> In fact, conditions most likely were not ideal throughout, which is why your actual score is smaller than one hundred.

regard to cognitive capacity, or any other trait for that matter. The maximum degree of moral agency as a function of the number of generations hence is unlikely to be monotonically increasing. However, moving along the line of your ancestors from the animal in the picture to your grandparents and parents, we will recognize an upwards trend, as schematically depicted in this graph:



Proponents of the kind of human exceptionalism that finds its expression in the belief in human dignity want to say that you, as well as your parents and grandparents, have intrinsic value, and have it equally, while your remote ancestor, the one depicted in Buell's illustration, is not part of the community of moral equals – whom it is consequently much less seriously wrong to kill than any of your immediate ancestors or other family members. If they are to defend that position in line with Liao's proposal, they must draw a distinction between genomes that contain the basis for moral agency and genomes that do not. As we have argued, that distinction must be the distinction between genomes that contain and do not contain the basis for a *certain degree* of moral agency. That is, to draw the distinction which is essential to Liao's proposal, a threshold of moral agency must be specified, which an individual's genome has to be able to support for the individual to qualify for full moral status. In the long line of your

ancestors, there must have been one individual, let us call her Eve, who was the first to have full moral worth. The genome of Eve differed from that of her parents in that it contained the genetic basis for a degree of moral agency that is slightly higher than the degree of moral agency her parents could at best have achieved. That difference in genetics seems entirely unremarkable. If Eve and her parents were to stand in front of us, we would most likely not be able to tell who among them initially had the greatest genetic potential for moral agency, and we might well find that Eve's parents now have a higher degree of moral agency than Eve, say, as they grew up under more favorable circumstances. In stark contrast to these minor differences in biological reality, the way Liao would have us treat Eve, who has full moral status, is radically different from the way he would have us treat her parents, who do not have full moral status. That is implausible. It is implausible that a small difference in genetics, which may or may not make a difference in terms of actual cognitive capacity, makes a momentous difference in terms of moral status.<sup>249</sup> That is especially true if one is motivated, as proponents of human dignity usually are, by the idea that what makes you and me stand out from those who purportedly lack human dignity are things like

---

<sup>249</sup> McMahan uses a similar line of argument against the idea that human beings have a special moral status in virtue of having a characteristically human genome. He asks us to imagine a spectrum of human-chimpanzee chimeras, which he calls the *transgenic spectrum*: "The Spectrum begins with a chimpanzee with an unaltered genome. In the next case, a single human gene is inserted in a chimpanzee zygote. In the third case, two human genes are inserted. In each case further along in the spectrum, one more human gene is inserted while the corresponding chimpanzee gene is deleted, if necessary. Thus at the other end of the spectrum is a case in which *all* of the chimpanzee genes are replaced by corresponding genes from a human source. In all cases the genetically altered zygote is implanted in a natural or artificial uterus and thereafter allowed to grow to adulthood" (McMahan 2007, p. 16). The Transgenic Spectrum has some similarity with your line of ancestors, with the notable difference that your ancestors are actual whereas most of the animals in the Transgenic Spectrum are at best possible. The point McMahan hopes to make is that "it would be absurd to suppose that the moral status of any individual in the spectrum would be determined by how many, or what proportion, of its genes were human or were taken from a human being" (McMahan 2002, p. 213).

“syntactical language, art, architecture, variety in social groupings and in other customs, burying the dead, making tools, religion, fear of death (and elaborate defense mechanisms to ease living with that fear), wearing clothes, true courting of the opposite sex [...], free choice, and morality [...]”<sup>250</sup>

If Eve’s parents in fact grew up under circumstances that are more favorable to cognitive development than the circumstances under which Eve grew up, the sense in which they are closer to these characteristically human abilities and activities is much more tangible than the sense in which they are not, which is tied to a rather abstract and impotent possibility buried in their genes.

We have seen in the previous chapter that McMahan’s two-tiered account of the wrongness of killing has an analogous feature, which is similarly unattractive. Interestingly, proponents of human dignity have pointed to exactly that feature as a reason to reject McMahan’s account. Lee, for example, contends that “[a] mere *quantitative* difference (having more or less of the same feature [...]) cannot by itself be the basis for why we should treat different entities in *radically* different ways.”<sup>251</sup> He further objects to the arbitrary line-drawing that McMahan’s account makes necessary.<sup>252</sup>

Liao’s proposal faces a similar problem, in addition to the distinct problem of it being implausible that a small difference in terms of genetic makeup makes a momentous moral difference. There seems to be no principled way to separate the line of your ancestors into two parts. Any line we could draw would seem arbitrary, and hence unjust. If we draw the line at having the genetic resources to score  $x$  on our imaginary scale of moral agency, the question immediately arises why we did not choose  $x$  minus  $0.1$ , or  $x$  plus  $0.1$ ,

---

<sup>250</sup> Lee & George 2008, pp. 184 f.

<sup>251</sup> Lee 2004, p. 254.

<sup>252</sup> Cf., e.g., Lee 2004, Abstract, and Lee 2011, p. 26.



to which there can hardly be a principled answer. That is not an innocent case of vagueness, but a serious shortcoming, as much depends on where we draw the line in terms of morality.

Dean Stretton has argued that both objections also apply to Lee's and George's argument from substantial identity. Chimpanzees, he claims, too have a basic natural capacity for higher mental functions. Hence, if human beings but not chimpanzees are to have full moral status, that must be because "human beings have a *greater* natural capacity for higher mental functions."<sup>253</sup> That, however, appears to be a mere quantitative difference, and as such – according to Lee, as just quoted – cannot ground a radical difference in terms of moral status. Also, as natural capacities come in degrees just as much as immediately exercisable capacities do, "Lee equally faces the problem of specifying a non-arbitrary cut-off point within the range of possible capacities."<sup>254</sup> Lee's response to Stretton is puzzling. He claims that Stretton has misconstrued the criterion he and George, as well as other philosophers, have proposed for inclusion in the community of moral equals:

“[T]he conclusion of my argument was not that the criterion for the right to life is natural capacities, but [...] being a certain type of substance. [...] [T]he genuine criterion for having a right to life is *being* [...] a distinct substance of a rational nature [...]. I hold that human beings *do* differ in kind, and not just in degree, from other animals [...].”<sup>255</sup>

---

<sup>253</sup> Stretton 2004, p. 270.

<sup>254</sup> Stretton 2004, p. 271.

<sup>255</sup> Lee 2007, p. 97. Just one year later, Lee wrote in a joint chapter with George that “the criterion is: having a rational nature, that is, having the natural capacity to reason and make free choices” (Lee & George 2008a, p. 412), which clearly seems to support Stretton's interpretation of Lee's position, as it was also presented in earlier writings. The phrase “that is” indicates an equation.

Lee argues that all human beings are different in kind from all other animals in that they each are a substance of a rational nature. If that is true, in virtue of what is it true? If it is not the possession of a basic natural capacity for higher mental functions that makes an individual a distinct substance of a rational nature, as Lee seems to implicitly concede in his response to Stretton, what could it be? What is it that a normal adult chimpanzee lacks but an anencephalic infant has, which justifies denying a rational nature to the former while attributing it to the latter? Lee does not say, or say clearly enough, but maybe we can take his reference to the human kind as a hint. It is true that all human beings are of a kind whose members typically have the potential or capacity for higher mental functions, and one could speculate that Lee takes that to imply that all human beings have a rational nature. If that in fact is what Lee thinks, he is plainly making a mistake. As McMahan's correctly observes,

“[t]he morally significant properties characteristic of a kind do not get to be a part of an individual's nature simply because that individual possesses the closely but contingently correlated properties that are essential to membership in the kind. Properties that are inessential to membership in the kind do not define the nature of the kind, even if they are characteristic or typical.”<sup>256</sup>

We have argued earlier that it is doubtful that there is a property related to rational thought, moral agency, or any other characteristically human activity, be it having the corresponding higher mental capacities, or the potential or genetic basis for these capacities, that is lacked by all or almost all non-human animals yet essential to membership in the human kind. Further, even if there is such a property, it is doubtful

---

<sup>256</sup> McMahan 2005, p. 358.

whether its moral relevance can be made plausible, and there are the problems of line-drawing and basing a radical moral difference on a mere difference in degree.

However, there may also be another way to interpret Lee's reference to the human kind. One could read Lee's response to Stratton as a retraction from moral individualism. Non-essential properties do not define the nature of a kind but they can nevertheless be common within a kind, and it can be argued that the criterion for full moral status is not the possession of a rational nature in the strict sense which would exclude some human beings, but membership in a kind whose members characteristically or typically have the potential or capacity for higher mental functions as such. That would mean a move away from Strategy One, and towards Strategy Two, which we shall discuss in the remainder of this chapter.

#### **3.4. *Strategy Two: the nature-of-a-kind argument***

Strategy One is to identify an intrinsic property that has direct moral significance in that it confers upon everybody who has it a special moral status, and then to show that all human beings yet no other animals, with possibly very few exceptions, possess that property. We have argued that Strategy One holds little promise. Proponents of that strategy either have to build their position on a notion of potentiality which, even if it can be made intelligible, will likely be obscure and open them to the charge of moral gerrymandering, or have to endorse the genetic basis account which is based on the rather shaky empirical assumption that all human beings in fact have the genetic basis for higher

mental functions and runs into the earlier-discussed problems that arise at the line separating those who have and do not have the relevant genetic basis.

Proponents of Strategy Two recognize that there likely is no morally relevant intrinsic difference between all human beings and all or almost all other animals, and argue that it is misguided to look for such a difference, as moral status is not a matter of the intrinsic properties of individuals, but a matter of belonging to a certain kind. Carl Cohen, for example, says that human beings

“who are unable, because of some disability, to perform the full moral functions natural to human beings are certainly not for that reason ejected from the moral community. The issue is one of kind.”<sup>257</sup>

Cohen is not alone. Similar suggestions have been made by philosophers time and again. Michael Cox argues that “what counts in establishing rights are the characteristics that a certain class of beings share in general, even if not universally.”<sup>258</sup> Similarly, Michael Allen Fox writes that “human beings have basic moral rights *because* they are beings of the requisite *kind* [...]”<sup>259</sup> Thomas M. Scanlon agrees: “[T]he class of being whom it is possible to wrong will include at least all those beings who are of a kind that is normally capable of judgment-sensitive attitudes [...],”<sup>260</sup> and so does Rahul Kumar:

“[T]o fall under the protection of the morality of respect an individual need not in fact have the capacities which make rational self-governance possible. It may well

---

<sup>257</sup> Cohen 1986, p. 866.

<sup>258</sup> Cox 1978, p. 110.

<sup>259</sup> Fox 1986, p. 56.

<sup>260</sup> Scanlon 1998, p. 186.

be enough that one belongs to a kind of species that normally develops the requisite capacities for rational self-governance.”<sup>261</sup>

The common idea is that there is no test, to be applied one by one, on the basis of which full moral status is awarded or denied to individuals solely in view of their intrinsic properties. Rather, it is sufficient that certain desirable properties be a normal or characteristic occurrence among the members of a kind for all members of the kind to enjoy full moral status, regardless of their individual intrinsic properties.

The virtue of Strategy Two is obvious. If successful, it provides a justification for human equality which, unlike the proposals we discussed in this chapter so far, appeals to a property that all human beings *uncontroversially* possess. This, however comes at a significant cost. One of the features that made Strategy One appealing is its grounding of human dignity in an essential intrinsic property. That, we have argued, resonates with the common intuition that worth is a deep fact about us that obtains in virtue of what each one of us is, as an individual. Strategy Two grounds worth in the property of belonging to a kind whose members normally or characteristically have certain status-relevant mental capacities. That property is one of group membership, and group membership is always a relation between an individual and a set of individuals. Strategy Two hence bases our worth on a relational rather than an intrinsic property, which does not fit well with the idea that human beings should be respected in virtue of who they are as individuals. Worse, the current proposal also makes worth an accidental property. As whatever makes us human does not guarantee possession of the status-relevant mental capacities, there is a possible world, with a human population, in which no human being ever had or ever

---

<sup>261</sup> Kumar 2008, p. 72.

will have any capacity for higher mental functions.<sup>262</sup> In that world, human beings would not be of a kind whose members normally or characteristically have the relevant higher mental capacities and consequently lack full moral status. In contrast, a lone chimpanzee who through gene therapy or mutation acquired the capacity for higher mental functions, being the sole representative of his or her kind, would have full moral status. The idea of such a chimpanzee is taken from McMahan who refers to him or her as the *Superchimp*.<sup>263</sup> In arguing against the moral relevance of species membership, McMahan asks his readers to consider our world just as it is, with the only difference that there is a Superchimp. Here, then, the scenario is reversed. While all human beings, including those who are severely mentally retarded, have full moral status, the Superchimp does not, as he or she does not belong to a kind whose members normally or characteristically have the relevant higher mental capacities. That is implausible, McMahan argues, as the Superchimp is our cognitive equal. How could it be just to treat somebody who can think, speak, and act much like us in radically different ways just because he or she is not a member of our species? What is even less plausible, however, is the claim that the moral status of the Superchimp and any given human being who does not possess the relevant higher mental capacities depends on their circumstances. In the first scenario we considered, the former but not the latter has full moral status, whereas the latter but not the former has full moral status in the second (McMahan's) scenario, despite the fact that the two individuals as they figure in one scenario are physically indiscernible from themselves as they figure in the other. If we are essentially human organisms, we are not

---

<sup>262</sup> You could imagine a world that, across all time, contains only one severely mentally retarded human being, without the capacity for the relevant higher mental functions, making him or her the sole representative of his or her kind.

<sup>263</sup> Cf. McMahan 2002, pp. 147 ff. McMahan himself seems to have borrowed the idea from James Rachels, cf. Rachels 1990, p. 187.

essentially of a kind whose members normally or characteristically have higher mental capacities, which is why Strategy Two renders our worth accidental. That is counterintuitive in itself, as we have argued at length at the end of the previous chapter, and has implausible implications for the Superchimp and human beings without directly status-relevant properties.

Some may object at this point that my interpretation of the argument from kind membership is uncharitable. I have assumed that what is normal or characteristic for members of a species is a matter of mere statistics, and it is obviously implausible, Kumar says with regard to McMahan's Superchimp, "to think that the morality of how it is appropriate to treat an individual should be sensitive to changes in mere *statistical* facts."<sup>264</sup> However, he continues, claims about what is normal or characteristic for members of a species

"are not statistical generalizations. Rather, what they concern is the essential nature of a living kind, revealing facts about the *normal life-cycle* of that kind of living thing. The use of 'normal' here is unashamedly normative. Claims about the life-cycle of a particular kind of living thing, or species, are just constitutive of what it is to be a member of that species. Certain events may, of course, prevent a particular individual member of a species from living out the life cycle that is normal for species of that kind."<sup>265</sup>

Kumar illustrates what it means, according to this interpretation, to say that something is normal or characteristic for members of the human species by means of an analogy with plants. Just like some human beings never achieve the higher mental capacities that are

---

<sup>264</sup> Kumar 2008, p. 72.

<sup>265</sup> Kumar 2008, p. 73.

normal or characteristic for human beings, some seeds never grow. But others do, and when they do, they thereby do not transform into another kind of thing. Rather,

“[b]eing a seed is just a stage in the normal development path of a plant. A complete understanding of a particular kind of plant will include understanding that it is the nature of plants of that kind that they start life as seeds, and if the appropriate conditions obtain, they will realize their nature as full-grown plants.”<sup>266</sup>

Seeds that never grow are of the same kind as full-grown plants of the same species in that they share the same essential nature. The only difference between the two is the external fact that more favorable conditions obtained when the now full-grown plants were seeds. All seeds, when they come into existence, are set to grow, and they will in fact do so if appropriate conditions obtain. That is what is meant when one says that their kind is characterized by transformation into full-grown plants. Analogously, to say that certain higher mental capacities are normal or characteristic for the human species is to say that these capacities are “constitutive of what it is to be a member of that species.” Every particular individual member of the human species will at some point acquire the relevant higher mental capacities, unless that is prevented by “certain events,” which implies that all human beings come into being with an innate potential to go through the life-cycle that is normal or characteristic for members of their kind. But, if that is how to interpret the argument from kind membership, then the appeal to kind membership just needlessly complicates the issue and we are really facing an instance of Strategy One, against which we have already listed a number of objections. If what is normal or characteristic for a species indicates the essential nature of every individual member of

---

<sup>266</sup> Kumar 2008, p. 73.



the species, and if the moral status of the members of the species is determined by what is normal or characteristic for the species, then every individual member of the species possesses an intrinsic property that is directly status-conferring, and reference to the species is inessential.

At the end of the last section, we have noted that Lee, when pressed, seems to switch from Strategy One to Strategy Two. Kumar similarly seems to switch strategies, but in the opposite direction, when confronted with the example of the Superchimp, which casts significant doubt on the relevance of mere group membership to moral status. Ultimately, I think both strategies hold little promise. Before we draw that conclusion, however, we should say a bit more about Strategy Two. Just as there is no logical contradiction in grounding worth in an accidental intrinsic property, as McMahan does, there is no logical contradiction in grounding worth in an accidental relational property. We have shown that making membership in a kind whose members normally or characteristically have certain higher mental capacities, which according to our original interpretation is an accidental relational property, the criterion for full moral status has implications that are unintuitive. So basing worth, however, has the advantage, over basing it on the possession of higher mental capacities, of bringing all human beings into the community of moral equals. For that reason alone, we should give some further consideration to the nature-of-a-kind argument.

Proponents of Strategy Two argue that the moral status of individuals is based not on their capacities or potential, but on the kind to which they belong. If certain intrinsic properties are normal or characteristic for the kind an individual belongs to, that individual has full moral status, regardless of whether he or she possesses these intrinsic

properties. Notice, however, that each human being belongs to a multitude of kinds. We all are animals, mammals, and hominidae, some of us are male mammals, some of us are female mammals, some of us are red-haired Caucasians, some of us are red-haired Caucasians born to blonde parents, and so on. Proponents of Strategy Two assume that, among all those kinds, the human species is the relevant kind, and it is unclear how that assumption could be motivated other than by the mere desire to find a way to bring all human beings within the scope of the equal wrongness thesis. There are numerous biological definitions of the term “species.”<sup>267</sup> Among the more common ones are those that appeal to the ability to interbreed, lineage, and the possession of a specifically human genome. What reason could there be for whom one can procreate with, who one’s parents are, or having a certain kind of DNA to be more important, morally speaking, than the biological properties which determine our membership in, say, the mammalian kind? I cannot think of a good reason. Marking out species as the relevant kind seems arbitrary, yet it is crucially important for proponents of Strategy Two that animals be grouped at the species level rather than some higher taxonomic rank. Suppose moral status was instead assigned at the class level. Dogs and human beings are of the same kind insofar they both belong to the mammalian class. Some members of that kind, such as you and me, possess a wide range of higher mental capacities. Is that sufficient for the mammalian kind to be “characterized” by higher mental capacities, or for higher mental capacities to be the “norm” among mammals? If so, both dogs and human beings belong to a kind whose members normally or characteristically have the relevant higher mental capacities, and hence have full moral status. If not, both dogs and human beings do not belong to a kind whose members normally or characteristically have the relevant higher mental capacities,

---

<sup>267</sup> Cf., e.g., Hey 2001.

and hence lack full moral status. Both conclusions contradict the claim, which proponents of human dignity aim to establish, that human beings and no other animals – with a possible exception for dolphins, hominidae, and other cognitively advanced non-human animals, but certainly not for dogs – have full moral status. We hence see that the nature-of-a-kind argument has its very own problem of arbitrariness.

In addition to having its very own problem of arbitrariness, the nature-of-a-kind argument also has more or less the same problem of arbitrariness that we have encountered before. Recall our common evolutionary ancestor depicted in Buell's illustration, and the long line of individuals connecting him or her and us. Who among those individuals was the first human being? There is no definite answer to that question, as all definitions of the term "species" are vague to one degree or another. If the criterion of species membership is genetic, we must specify the degree of genetic match between an individual and today's human beings that is necessary for that individual to qualify as a human being, and there seems to be no principled way to do so.<sup>268</sup> If the criterion of species membership has to do with lineage or the ability to interbreed, we similarly must set an arbitrary cutoff point. Any individual in our long line of ancestors was able to interbreed with his or her immediate predecessor and successor, yet you and I would not be able to interbreed with any of our remote evolutionary ancestors. That shows that, unlike the relation "is a member of the same species as," the ability to interbreed is not transitive, which is why the ability to interbreed as such cannot be the criterion of membership in

---

<sup>268</sup> The same problem arises for the transgenic spectrum that has a chimpanzee at one of its ends and a human being at the other. "Is there a point along this spectrum at which the individuals cease to be chimpanzees and become human beings? Is there, in other words, a point at which there is an individual with just enough human genetic material to count as a member of our species?" (McMahan 2002, p. 213.)

the same species.<sup>269</sup> If species membership is somehow based on the ability to interbreed, the relationship between species membership and the ability to interbreed must be more complicated, and very likely involves an arbitrary line. But suppose we had an entirely non-arbitrary criterion of membership in the same species. That would not settle the relevant question of who was the first to be of a kind characterized by higher mental capacities. Were higher mental capacities sufficiently common among the first generation of human beings for them to be of a kind whose members normally or characteristically possess higher mental capacities? Or were higher mental capacities already sufficiently common within the species that preceded ours? With the birth of the first of our ancestors who met some threshold of mental capacity, did all of his or her conspecifics suddenly acquire full moral worth? Or did a kind for which a certain degree of mental capacity is normal or characteristic only emerge later? To answer these questions we must first decide which capacity determines moral status. Is it moral agency (Cohen), the capacity of having judgment-sensitive attitudes (Scanlon), or rational self-governance (Kumar)? Once we have made that choice, we need to specify the degree to which the capacity must be developed to be relevant for full moral status, and we must then say how common the possession of the specified degree of the relevant capacity must be within a species for that species to qualify as a kind for which the specified degree of the relevant capacity is normal or characteristic. I cannot imagine how this could be done in a principled manner. There is no good reason to choose one degree of capacity over another, especially if the degrees are very close to each other, and making the vague notion of what is normal or characteristic precise necessarily involves a choice that, to

---

<sup>269</sup> So-called *ring species* pose a similar problem for definitions of the term “species” that are based on the ability to interbreed. Cf. Dawkins 1993.

some extent, will always be arbitrary. The nature-of-a-kind argument hence also relies on arbitrary line-drawing, where the line is the boundary of our kind. In that, it is similar to the arguments put forward by McMahan, Liao, and – as Stretton has argued – Lee and George, and the arbitrariness is similarly problematic. It makes a great difference in the way your (great)<sup>n+1</sup>-grandmother would have deserved to be treated whether we draw the line such that she is the first among our ancestors to fall into our kind, or such that instead your (great)<sup>n</sup>-grandmother is the first to fall into our kind, and yet there seems to be no more or less reason to go one way rather than the other. Further, wherever the line is eventually drawn, the intrinsic difference between the individuals just below and above it will be unnoticeable, whereas the difference between them in terms of moral status is momentous – which, as we have noted more than once before, is implausible.

This brings us to the end of our discussion of human dignity. We have argued that there is no intrinsic difference between all human beings and all other animals for which a plausible and convincing case can be made that it is relevant for the attribution of full moral status. We have further argued that all attempts to morally distinguish between human beings and other animals that we have considered in this chapter necessitate drawing a line which is arbitrary, and implausibly opens a moral gulf between individuals whose difference from one another in terms of empirical reality is entirely unremarkable. If human beings whose cognitive impairment is both severe and permanent are to be included in the community of moral equals, it seems like we must lay to rest the idea that we can do so without also including a wide range of non-human animals.

## **Chapter 4: Being a world unto one's self: a new perspective on the wrongness of killing, and some of its implications**

Traditional morality maintains that human beings are special among the animals, in that only they have full moral status, whereas non-human animals have a lower moral status. The religious doctrine that God has created all and only human beings in his image has for a long time been an important foundation of that belief.<sup>270</sup> Whereas religion, and Abrahamic theism in particular, has lost much of its appeal, if not among the general population, certainly among Western philosophers, the belief that each human being has equal intrinsic worth or dignity has maintained its popularity, and pervades our laws, institutions, and practices.<sup>271</sup>

In the previous two chapters, we have discussed a number of secular arguments for human dignity and its approximation, the dignity of persons. We have argued that all of these arguments fail to convince, in one way or another. There seems to be no good reason to believe that there is an important moral difference between all human beings and all other animals, and even the less ambitious project of drawing a line between most human beings and most other animals, by way of appeal to the philosophical notion of a

---

<sup>270</sup> The declaration that “all men are created equal,” which is part of the United States Declaration of Independence, for example, implicitly refers to a creator, and is anchored in the idea that human beings are all equally created in the image of God, cf., e.g., Fletcher 2001. Jeremy Waldron has recently argued that John Locke, who is commonly known as the “Father of Liberalism,” too has defended equal human worth on the basis of our supposed equal God-likeness, cf. Waldron 2002.

<sup>271</sup> In 2014, David Bourget and David J. Chalmers surveyed 931 faculty members in departments of Philosophy in the United States, Canada, Europe, and Australasia, in order to determine their views on central philosophical issues. 72.8% of respondents stated that they accept or lean towards atheism, while only 14.6% stated that they accept or lean towards theism, cf. Bourget & Chalmers 2014, p. 494. It should be noted that belief in the existence of *a* god does not guarantee belief in the Abrahamic god, and even less belief that all human beings are created in God's image.

person, is marred by considerable difficulties. The fundamental problem, roughly, is one of mapping. Before the emergence of Darwinism, the ancient Aristotelian picture of a hierarchy of nature, made up of neatly distinguishable categories of beings, made a good deal of sense. Species were believed to possess unchanging essences, and insofar these essences were thought to carry moral significance, it was easy and intuitive to draw a line between “us” – that is, humanity – and the “others,” who do not get to be part of the community of moral equals.<sup>272</sup> With the advance of evolutionary theory, the boundaries between the old categories of being have become ever more blurry. Biological characteristics vary widely among individuals of a particular species, as well as across species boundaries, and the evolutionary forces of selection, mutation, migration, and random drift guarantee that there is a continuous and mostly gradual change of biological characteristics from one generation to another. If we could look at all organisms that have ever existed at once, we would see in front of us a continuous spectrum of the biological characteristics that philosophers commonly associate with our special moral status. Not only would we not be able to make out any discontinuities between all human beings and all other animals, and especially their immediate predecessors, we would find that there is no recognizable discontinuity in the spectra of degrees of capacity or potentiality for autonomy, rationality, moral personality, self-consciousness, and so on *at all* that would recommend itself as a line of moral demarcation, which is a serious problem not only for defenders of human dignity, but also for McMahan and other philosophers who similarly seek to superimpose a binary morality onto these spectra. In order to do so, they need to define a map between these continuous spectra on the empirical side, which is messy as all of nature is, and the neatly compartmentalized world of traditional or otherwise

---

<sup>272</sup> For a good introductory discussion of the death of essentialism in biology, see Ereshefsky 2010.

dichotomous morality, and we have argued that there is little hope that this can be done in a non-arbitrary and intuitively plausible way.

I am hardly the first one to notice that our modern scientific understanding of our place in nature puts pressure on traditional morality. Tim Mulgan, for example, writes that

“the [...] claim that all persons have equal worth, and thus that the killing of any person is equally wrong [...] sits uneasily with any naturalistic picture of what gives rise to worth, as the underlying properties necessarily come in degrees. This is a perennial problem for naturalistic liberals, the dominant party in contemporary moral and political philosophy. [...] Unsurprisingly, the search for a coherent naturalistic foundation inevitably fails.”<sup>273</sup>

James Rachels has dedicated a whole book, *Created from Animals: The Moral Implications of Darwinism*, to the tension between a Darwinist view of the world and traditional morality. He argues that “Darwinism undermines both the idea that man is made in the image of God and the idea that man is a uniquely rational being [...], [and that] it is unlikely that any other support for the idea of human dignity will be found.”<sup>274</sup> From this, Rachels concludes that the idea of human dignity is no more than “the moral effluvium of a discredited metaphysics”<sup>275</sup> and should be rejected, and with it the belief that each human being has equal intrinsic worth or dignity. In lieu of traditional morality, he advocates a version of consequentialism, and it is no coincidence that Mulgan too is a consequentialist. Consequentialism is *one* way to avoid the mapping problem we just described, but it denies or at least makes questionable what is commonly seen as one of humanity’s greatest moral achievements, the almost universal acceptance of the idea that

---

<sup>273</sup> Mulgan 2004, p. 459.

<sup>274</sup> Rachels 1990, p. 5.

<sup>275</sup> Rachels 1990, p. 5.



every human being has equal worth and hence an equal right not be killed.<sup>276</sup> We look back with horror at the dark episodes in history where that idea was systematically denied in brutal and disturbing ways.

I think that the idea of equal human worth is in fact a great achievement and should hence not easily be surrendered. In this chapter, I will propose what I believe to be a tenable and attractive alternative response to the tension between our modern naturalistic view of the world and traditional morality, which is less radical than consequentialism in that it retains the idea that we are all equal in terms of moral worth, but also more radical in that it rejects human exceptionalism in a way most versions of consequentialism do not. It will be expedient to briefly recapitulate what we have argued thus far, as close attention to the different ways in which the accounts we have discussed in previous chapters have left us unconvinced will help us develop an account of the wrongness of killing that avoids these shortcomings.

#### ***4.1. What we have argued thus far, and an outline of what lies ahead***

Consequentialists explain the wrongness of killing in terms of the harm it does to society in general, or the victim in particular. As the value of the consequences of death varies from person to person, and from circumstance to circumstance, it did not come as a surprise that we were able to demonstrate that the implications of consequentialism for the comparative wrongness of killing are often gravely at odds with our considered intuitions. Also, while it is, of course, true that it is often wrong to deprive others of

---

<sup>276</sup> In the earlier chapter on consequentialism, we have also listed a number of other reasons that cast doubt on the ability of consequentialism to give an adequate account of the wrongness of killing.

future value or otherwise harm them, we have suggested that the harm that killing typically brings about is not its primary wrong-making feature. This became particularly clear in our discussion of cases where killing involves little or no harm. It seems seriously morally wrong to kill us against our will, *regardless* of how miserable or unproductive our lives may be, and *regardless* of how little good our lives might do for ourselves or others. Further, even in the case of killings that involve a great deal of harm, which consequentialism correctly condemns, we have argued that the consequentialist explanation of the wrongness of such killings in terms of the consequent harm is problematic, as it appears to be of the wrong kind. Killing is deeply personal, in the sense that we expect the explanation of its wrongness to have the one who is killed herself at the center, rather than contingent facts about the effects on the world, or her well-being. Consequentialism focuses our attention on what killing takes away from the victim, and diverts our attention from the seemingly more important fact that the victim is annihilated. In the words of Timothy Chappell, who I think has put this point perfectly, “[i]n killing, the main point is not that something is taken away from someone, but that *the someone* is taken away.”<sup>277</sup>

McMahan proposes an account along these lines, and offers an explanation of the wrongness of killing that, unlike consequentialism, *does* make essential reference to the victim. He argues that killing one of us is wrong, when it is wrong, because it involves a failure to respect the victim and his or her moral worth. Insofar as worth is equal for all of us, and hence independent of the value of our futures, and our value to others, all killings of individuals like you and me are equally wrong, other things being equal. This – to me, at any rate – is a more natural and promising explanation of the wrongness of killing than

---

<sup>277</sup> Chappell 2004, p. 111.

the one offered by consequentialists. McMahan, however, fails to provide a convincing factual basis for moral worth, and counterintuitively excludes human beings who do not meet his criterion for moral worth – which, we recall, is personhood in the Lockean sense – from the community of those whom it is normally equally wrong to kill. Very small children are not yet persons and therefore, according to McMahan, not yet one of “us,” and severely congenitally mentally retarded infants never will be members of the community of moral equals.<sup>278</sup> Much of the unease generated by McMahan’s account of the wrongness of killing arises from the underlying account of our moral worth, which Lee and George argue, I think correctly, is defective. Instead of superficially grounding our moral worth in a set of accidental properties you and I happen to have, as McMahan does, Lee and George base our equality on the rational nature they say we have essentially in common with all living human beings, and thereby reinstate as a tenable position, without reference to God, the traditional view that the killing of one human being is as seriously wrong as the killing of any other human being. However, Lee’s and George’s argument form substantial identity and other human dignity accounts have significant problems of their own.

In the previous chapter, we have argued that the property of having a rational nature, regardless of whether it means having the potential or the genetic basis for rational thought, is not an adequate basis for equal moral worth, for three reasons. First, there is no good reason to believe that all human beings, or even all of those whom most consider to be our moral equals, have a rational nature. There are members of our species who are constituted such that, under anything like normal circumstances, their brains will never

---

<sup>278</sup> In Kittay 2005, Eva Feder Kittay argues passionately and forcefully that the exclusion of severely congenitally mentally retarded infants from the moral consideration of persons is “as morally repugnant as earlier exclusions based on sex, race, and physical ability have been” (p. 100).

develop to the point at which immediate capacities for rationality, or any other higher mental functions for that matter, can arise. Second, we have argued that the distinction between animals who have a rational nature and animals who do not have a rational nature is no less arbitrary than McMahan's distinction between persons and non-persons. Finally, that distinction, no matter how it is made, seems ill-suited to carry the moral weight of the distinction between those who have and those who do not have full moral worth and status, as it is entirely unremarkable in terms of empirical reality.

In order to avoid these objections against the accounts we have discussed in the previous chapters, an alternative account of the wrongness of killing would have to explain the wrongness of killing not in terms of what killing does to the victim, or the world, but in terms of its being a grievous failure to respect the victim and his or her moral worth. It would have to reject the superficiality of McMahan's morality of respect, and instead affirm that moral worth is not transient but the kind of thing we have during our entire existence. It would consequently ground our worth not in a property that you and I merely happen to have, but in a property that we have essentially. It would retain the idea that we are all equal in terms of moral worth, and that killing any one of us hence is as seriously wrong as killing another, and carve out of the natural world of organisms the community of moral equals along a non-arbitrary line that carries empirical significance.

If there is to be an account that has these features, what is the empirical property at its heart that confers equal moral worth on individuals like us? I think there is a rather straightforward answer to this question, and there is a hint of it in Chappell's contention that what typically makes killing wrong is the fact that a "someone is taken away." I will argue that what makes you, me, other normal adults, and almost all human beings who

are not normal adults fundamentally equal in the relevant sense is not that we are all somehow equally distinct from other animals, in virtue of having a rational nature, but that we all are equally *somebodies*, rather than mere *somethings*.<sup>279</sup> If this is to work, we need to show that the property of being a somebody meets the conditions set out above. We will begin that task by specifying what makes a somebody a somebody. We will then argue that you and I not merely happen to be somebodies, but that we are *essentially* somebodies, and that being a somebody is a binary property. From this we will conclude that the line that separates somebodies from somethings is anything but arbitrary and signifies a remarkable distinction in nature. Finally, we will put the pieces together and state, in more detail and explicitly, a respect-based account of the wrongness of killing that makes the property of being a somebody the basis of our moral equality, and explore some of its implications.

#### **4.2. *Being a somebody as a matter of phenomenal consciousness***

Roughly, the crucial difference between a somebody and a something – that is, a mere thing – is stated easily enough. You and I are conscious, whereas mere things such as stones and rocks are not. The word “consciousness,” however, has more than one meaning, and can refer to a great many different phenomena. We will hence need to determine which meaning of the word is the one that is relevant to the distinction between somebodies and mere things. It is beyond the scope of this dissertation to provide an exhaustive survey of the many concepts of consciousness that have been

---

<sup>279</sup> George Sher makes a similar proposal in Sher 2014, Chapter 5.

discussed in the philosophical literature.<sup>280</sup> Instead, we will restrict ourselves, as will turn out to be sufficient for our purposes, to the most important distinction, which to me seems to be that between phenomenal and access consciousness.

*Phenomenal consciousness* is notoriously hard to define. The British psychologist Stuart Sutherland has even gone as far as to say that it “is impossible to define except in terms that are unintelligible without the grasp of what consciousness means. [...] [I]t is impossible to specify what it is, what it does, or why it evolved.”<sup>281</sup> Yet, at the same time, phenomenal consciousness is the phenomenon we are most intimately acquainted with.<sup>282</sup> In an important sense, there is nothing we know better than what it means to be phenomenally conscious as we smell freshly-brewed coffee, observe a sunset, taste a bowl of pasta, or feel the pain of a headache. René Descartes made this point forcefully in his *Meditations on First Philosophy*, which is the one thing that everybody who has read the treatise is most likely to remember.<sup>283</sup> Maybe the best we can do is to describe phenomenal consciousness using terms that are no more basic, hence not suitable for definition, but perhaps clearer, and point to instances of it, hoping that anybody capable of phenomenal consciousness will recognize what is meant. Take the smelling of coffee or the feeling of pain. Each involves a plethora of highly complex neural processes that take place in one’s brain, but there is more to it than these processes, and that “more” is what we are trying to get at. There is *something it is like* to smell coffee or feel pain. Each

---

<sup>280</sup> For a useful discussion of the history of the study of consciousness, the many meanings of “consciousness,” and the relation of consciousness to other mental phenomena, see van Gulick 2004, and Siewert 2011.

<sup>281</sup> Sutherland 1995, p. 95.

<sup>282</sup> In David J. Chalmers’ words, phenomenal consciousness is “at once the most familiar thing in the world and the most mysterious. [...] We know consciousness far more intimately than we know the rest of the world, but we understand the rest of the world far better than we understand consciousness” (Chalmers 1996, p. 3).

<sup>283</sup> Descartes 1641, Second Meditation.

of these experiences has a distinctive subjective quality to it, a “feel,” and it is that internal aspect of experience which we mean when we talk of phenomenal consciousness. Phenomenal consciousness, then, is the kind of consciousness for which it is the case that there is something it is like for the one who is conscious to be conscious.<sup>284</sup> *Access consciousness*, in contrast, can occur while “all is dark within,” as it is often put. Mental states that are access conscious may be, yet are not necessarily, associated with a subjective feel. A state is access conscious, according to a popular definition by Ned Block, “if, in virtue of one’s having the state, a representation of its content is [...] poised for use as a premise in reasoning, [...] poised for rational control of action, and [...] poised for rational control of speech.”<sup>285</sup> Access consciousness hence is a matter of the availability of information. A state is access conscious if the information it contains is generally or globally available for use by the one whose state it is.

That phenomenal and access consciousness can, at least in principal, come apart follows, for example, from the logical possibility of phenomenal zombies.<sup>286</sup> Phenomenal zombies are entities that are identical with human beings in terms of function, in the computational sense, yet lack phenomenally conscious experiences altogether. A phenomenal zombie can verbally report its inner states, and detects, processes, and responds to stimuli in the same way you and I do, except that there is no internal aspect to these processes. Despite their lack of phenomenal consciousness, it seems clear that phenomenal zombies are access conscious. Also, as Neil Levy has pointed out, there is no reason why a computer could not be access conscious as well. “[T]here is no reason why

---

<sup>284</sup> Cf. Nagel 1974.

<sup>285</sup> Block 1995, p. 231.

<sup>286</sup> For an argument for the logical possibility of zombies, see, e.g., Chalmers 1996, pp. 94 ff. It should be noted, however, that not everybody agrees that phenomenal zombies are possible. For a dissenting view, see Shoemaker 1999. For a helpful introduction to the philosophical debate over zombies, see Kirk 2015.

a computer should not be set up so that its central processing mechanisms have access to the contents of any of its other systems.”<sup>287</sup> If so, this is another instance of access consciousness without phenomenal consciousness. Both zombies and computers seem to be mere things, rather than somebodies, and we shall hence assume that the notion of consciousness that is relevant to the distinction between somebodies and mere things is phenomenal consciousness, which is also what people usually mean by “consciousness.” Yet, being a somebody cannot simply be the same thing as being phenomenally conscious in the sense of currently having phenomenally conscious experiences. Plainly, there often are times when we are unconscious yet do not cease to be somebodies, for example, during dreamless sleep.<sup>288</sup> We say “Somebody is asleep,” not “Something is asleep.” Therefore, we will tentatively equate being a somebody with having the *capacity* for phenomenal consciousness, which persists through periods of temporary unconsciousness, and show, moving forward, how this ties in well with our overall project.

The next step in our argument is to show that you and I have the capacity for phenomenal consciousness essentially, rather than accidentally.

#### **4.3. *What sort of things are we essentially?***

One of the premises of the argument from substantial identity, which we have discussed at length in the previous chapter, is the claim that “[w]hat we are is each a human,

---

<sup>287</sup> Levy 2014, p. 3.

<sup>288</sup> Some philosophers have defended the view that we remain dimly conscious even during dreamless sleep. Edmund Husserl was one of them, cf. Smith 2003. As that view is very speculative, we do better if we do not assume its truth, and instead work on the assumption that the common sense view is correct. For the sake of brevity, I will often shorten the term “phenomenally conscious” to “conscious.”



physical organism. (We are human, physical organisms essentially.)”<sup>289</sup> We have shown that Lee’s and George’s case for human dignity is implausible independently of whether or not that premise is true, which is why I had granted it, promising that I will get back to the issue later. This is the place to make good on that promise. To say that each of us is essentially a human organism is to say, more precisely, that each of us is numerically identical with a human organism. According to Lee and George, and a number of other philosophers who subscribe to this view, you and the human organism with which you currently coexist – “your organism” – are one and the same thing, rather than two distinct things.<sup>290</sup> I suspect that most people without philosophical training, when asked, would say that this is clearly true. It might seem as nothing more than scientific common sense that the infant my parents named “Rainer” and I are one and the same thing, because the infant’s organism is the same organism as my organism and both the infant and I just *are* our respective organisms. Yet, the majority of modern philosophers reject the idea, which is in fact more aptly characterized as metaphysical than scientific, that we are human organisms, and I think there are good reasons for us to follow suit and do the same.

Neurosurgeons are optimistic that human brain transplants, even though still very far away, are possible, and their optimism seems warranted. As early as in the 1960s, the American medical doctor Robert J. White conducted a number of neurosurgical experiments with animal brains, and it is reported that he “was able to obtain electrical activity from intact brains of dogs and monkeys, either transplanted to the neck

---

<sup>289</sup> Lee & George 2007, p. 134.

<sup>290</sup> Perhaps the most notable defense of this view, which is commonly called *animalism*, in recent times is that set out by Eric T. Olson in his 1997 book, *The Human Animal* (Olson 1997).

vasculature or kept alive in jars.”<sup>291</sup> Suppose, then, that transplantation of a functioning brain from one human organism to another becomes a reality, and imagine Maria and her husband get into a terrible car accident. A small piece of the car enters the husband’s skull and shatters his brain. He dies instantly. The rest of his body, however, is intact, including most of his head. Maria’s condition is reversed. She sustains severe injuries to multiple vital organs in her torso, while her brain remains intact. In an effort to save Maria’s life, a team of surgeons transplants her brain into the body of her deceased husband. If everything goes well, it will be difficult to refuse the claim that Maria has survived the procedure, and will go on to live her life with the body that was formerly her husband’s. The person waking up in the hospital after the brain transplantation has been performed will believe that she is Maria, and have all the memories, beliefs, character traits, and other psychological characteristics Maria had before the accident occurred (or, more realistically, most of them, as profound trauma often has significant effects on one’s psychology). But, if Maria in fact survives, she cannot have been numerically identical with the organism from which her brain was extracted during the transplantation procedure, as that organism is now dead. In contrast, the organism of Maria’s husband is still alive after the transplantation, with the only noteworthy difference that it now has a new brain, and yet the husband is gone. As a thing cannot at the same time exist and not exist, Maria’s husband cannot be the same thing as his organism.

Defenders of the claim that we are essentially living human organism could bite the bullet, and insist that, even though the surviving individual thinks he is Maria, he is mistaken and in fact Maria’s husband, who now, because of his new brain, has Maria’s

---

<sup>291</sup> Freed 2000, p. 38. In the same decade, Sydney Shoemaker first used the idea of a brain transplant in the context of the philosophical debate over personal identity, cf. Shoemaker 1963, pp. 23 f. The following thought experiment is ultimately owed to him.

psychology. A more plausible, but I think ultimately also unconvincing, response has been put forward by Peter van Inwagen. His argument, adapted to our thought experiment, suggests an alternative description of what happens during the transplantation process: Maria's brain is not extracted from her organism, but rather – after the extraction has taken place – *is* her organism. What is left behind is not a whole organism but the non-essential parts of Maria's organism, a mere collection of cells, while Maria persists in a “radically maimed”<sup>292</sup> form, as a woman “about as maimed as it is possible for a [wo]man to be.”<sup>293</sup> Maria, who is now a bare brain, is then equipped with the non-essential parts that are left of her husband. This, I think, is a very unnatural description of the transplantation procedure, which I am hence unwilling to adopt. It also raises the question of what makes the brain so special? Van Inwagen's answer is that what makes the brain so special is its special function as the “control center” of an organism. However, in general, if a thing that has a control center is separated into two parts, its control center and whatever remains, it is not true that that thing persists as the control center. To suppose that Maria's organism persists as a brain after having been removed from her skull, McMahan therefore argues, is

“comparable to supposing that, if the control panel is removed from the cockpit of an airplane, the airplane continues to exist as the control panel while the remainder, consisting of the fuselage, wings, and so on, becomes just an assembly of spare parts.”<sup>294</sup>

Another problem for the view that we are essentially human organisms arises from the real-life phenomenon of dicephalus, which can occur when a human zygote divides

---

<sup>292</sup> Van Inwagen 1995, p. 172.

<sup>293</sup> Van Inwagen 1995, p. 172.

<sup>294</sup> McMahan 2002, p. 34.

incompletely. Dicephalus is a rare kind of conjoined twinning, where twins are conjoined below the neck. There often is some duplication of organs below the neck, but there are cases of dicephalus where the systems that carry out the basic life processes which define an organism, such as metabolism or reproduction, each occur only once. In such a case, it seems more accurate to speak of one human organism than of two distinct human organisms. Yet, such an organism is clearly “inhabited” by, or coexists with, two individuals. The two separate functioning brains support two separate mental lives which have no direct access to each other. The experiences of each of the two individuals are private, and each individual has his or her very own memories, beliefs, desires, and personality traits. Twins so conjoined cannot both be numerically identical with the organism they inhabit, as that would make them the same thing, which they are not. Identity is transitive. It also seems unreasonable to suppose that one of them is identical with the organism whereas the other is not, as there is no distinguishing characteristic that marks out either one of them. If it was the case that only one of them is identical with the organism, that would leave us with the other, who would then *not* be identical with an organism. Hence, there are pairs of individuals, essentially just like us, who share a single organism, both or at least one of whom are not identical to that organism. Hence, we are not human organisms.<sup>295</sup>

S. Matthew Liao has objected to this argument that, in all cases of dicephaly, “there are in fact two organisms, although they may not be completely independent organisms.”<sup>296</sup> We

---

<sup>295</sup> This argument reminds us of our discussion of monozygotic twinning in the previous chapter, where we mentioned what could be presented as an additional problem for the view that we are essentially human organisms. If zygotes during their first fourteen-or-so days of existence, when monozygotic twinning can occur, are complete human organisms, then some of us – those who are not identical twins – began their existence as freshly-fertilized ova, while others – those who are identical twins – started out as zygotes of one or two weeks. As we are all of the same kind, that is a rather odd implication.

<sup>296</sup> Liao 2006, p. 340.

can imagine that van Inwagen would take a similar position, as there are two, rather than one, control centers. As we have noted before, however, it is unclear why the brain should be regarded as the only non-essential part of an organism, even if it is in fact accurately described as the control center. A human organism is an integrated biological system with a number of organs, more than one of which is vital. It can no more by itself maintain its basic life processes without a heart than without a brain. But, if external support is provided, a human organism can continue to exist without a heart just as much as it can continue to exist without a brain. The shared body of dicephalic twins with little organ duplication constitutes a single integrated biological system, with one circulatory system, one immune system, etc., and hence is relevantly more like a single organism with an extra pair of lungs or an extra heart than two distinct human organisms.<sup>297</sup> Not every reader will be convinced by this, and it might seem as much a matter of scientific convention as it is a matter of philosophical inquiry what makes a human organism. I hence do not want to withhold from the reader a thought experiment proposed by McMahan, which provides further reason to reject the view that we are human organisms, and which I find intuitively very convincing.

If you are identical to a human organism, you came into existence when that organism came into existence. There is some uncertainty about when a human organism comes into existence, mainly due to the possibility of monozygotic twinning, but it is uncontroversial that a human organism comes into existence no later than two or three weeks after conception, well before the onset of consciousness. At that point, the cells of the embryo have begun to work together as an integrated biological system. Were you really once such a pre-sentient embryo? McMahan suggests that, in order to answer this question, “it

---

<sup>297</sup> Cf. McMahan 2002, pp. 35 ff., and McMahan 2015, pp. 515 f.

is helpful to consider, not whether one *was* a tiny cluster of cells, but whether one could ever *become* such an entity.”<sup>298</sup> The thought experiment he is here proposing very much appears to be inspired – even though he never says it is – by F. Scott Fitzgerald’s 1922 short story, *The Curious Case of Benjamin Button*, which was recently made into a film. Benjamin is born with the body of a 70-year-old man. A few years after his birth, his family realizes that his aging process is reversed. Benjamin ages backwards, in biological terms. “When Benjamin was eighteen he was erect as a man of fifty; he had more hair and it was of a dark gray; his step was firm, his voice had lost its cracked quaver and descended to a healthy baritone.”<sup>299</sup> As the years fly by, he transforms from a grown man into a teenager, and then from a teenager into a child. Still later, he reaches a point, from which onward he needs the care of a nurse, Nana.

“He did not remember clearly whether the milk was warm or cool at his last feeding or how the days passed – there was only his crib and Nana’s familiar presence. And then he remembered nothing. When he was hungry he cried – that was all. Through the noons and nights he breathed and over him there were soft mumblings and murmurings that he scarcely heard, and faintly differentiated smells, and light and darkness.

Then it was all dark, and his white crib and the dim faces that moved above him, and the warm sweet aroma of the milk, faded out altogether from his mind.”<sup>300</sup>

Think yourself into Benjamin’s shoes. Imagine your organism would grow younger rather than older, and ask yourself at what point you would cease to exist. There will come a time when your consciousness will noticeably start to fade away, in that its

---

<sup>298</sup> McMahan 2002, p. 29.

<sup>299</sup> Fitzgerald 1922, p. 11.

<sup>300</sup> Fitzgerald 1922, p. 26.

contents will become less and less sophisticated. It is unlikely that you will be in a crib then, as Fitzgerald writes, more likely that you would have had to be placed in a fetal incubator to be kept alive. Eventually, it will be all dark, and then – shortly afterwards – it will not even be dark anymore: no inner feel, no subjectivity, just a cluster of cells, still a functioning organism, but no consciousness. At that point, and here I concur with McMahan, “I find it impossible to believe that I would still be around [...]”<sup>301</sup> If we accept that intuition, that you cease to exist before your organism does, then you cannot be the same thing as your organism.

All of the foregoing objections against the view that we are essentially human organisms suggest that what makes us what we are has to do with our mental lives. It seems obvious that Maria goes where her brain goes, and that is, it seems natural to think, because her brain takes along all of Maria’s psychological characteristics. Dicephalic twins are two individuals rather than one because they have two separate mental lives, and Benjamin goes out of existence when everything becomes dark because his existence continues only as long as his mental life does. A ready alternative for the view that we are human organisms hence is the view that we are essentially psychologies, which in one form or another has been held by most modern philosophers.<sup>302</sup> Its first substantial defense has been provided by the English philosopher John Locke. In his *Essay Concerning Human Understanding*, he argues that our numerical identity over time consists in facts about memory,

“[f]or should the soul of a prince, carrying with it the consciousness of the prince’s past life, enter and inform the body of a cobbler, as soon as deserted by

---

<sup>301</sup> McMahan 2002, p. 29.

<sup>302</sup> Prominent proponents include Thomas Nagel, Robert Nozick, Derek Parfit, Sydney Shoemaker, and Peter Unger, cf. Nagel 1986, Nozick 1981, Parfit 1971 & 1984, Shoemaker 1970, and Unger 1990.

his own soul, everyone sees he would be the same person with the prince, accountable only for the prince's actions [...].”<sup>303</sup>

After the soul transfer – which is the 17<sup>th</sup>-century equivalent of brain transplantation – has taken place, the person inhabiting the body that was formerly the cobbler's is the prince rather than the cobbler, as he remembers, “is conscious of,” the prince's past life rather than the cobbler's. In the century following the publication of Locke's *Essay*, the Scottish philosopher Thomas Reid and the English theologian John Sergeant object that identity cannot consist in having the same memories, as identity is transitive and having the same memories is not.<sup>304</sup> Both my current thirty-year-old self and my ten-year-old self may remember that I fell off my bicycle and badly scraped my knee at the age of five, and both my current self and my future eighty-year-old self may remember what my first year in Houston was like, but my eighty-year-old self may well no more have any memory of that bicycle accident. Locke's theory of identity would imply, it seems, that my ten-year-old self is identical with my current self, and that my current self is identical with my eighty-year-old self, but that my ten-year-old self is not identical with my eighty-year-old self, which violates the transitivity of identity and hence is contradictory.<sup>305</sup> For this and other reasons, later philosophers sympathetic to Locke's view have given up the requirement that there be direct memory connections between the temporal slices of individuals who persist through time. Instead they appeal to the *continuity* of memory as well as other psychological features such as beliefs, desires, intentions, and so on. One recent and influential example is Derek Parfit. The central

---

<sup>303</sup> Locke 1690, Book II, Chapter 27, p. 229.

<sup>304</sup> Cf. Behan 1979.

<sup>305</sup> Further, it seems plain that we can survive a total or near-total loss of memory. If so, personal identity cannot consist in having the same memories.



concept of Parfit's theory of personal identity is one we are already familiar with, that of psychological connectedness.<sup>306</sup> For instance, if you remember an experience you had five years ago, then there is a direct psychological connection between your current self and yourself five years ago. "Other such direct connections are those which hold when a belief, or a desire, or any other psychological feature, continues to be had."<sup>307</sup> Psychological connectedness is the basis of psychological continuity, which obtains when there are overlapping chains of strong psychological connectedness, where strength consists in there being a sufficiently large number of direct psychological connections.<sup>308</sup> Psychological continuity, then, is said to constitute personal identity. If an individual at one time is psychologically continuous with another individual at another time, they are numerically identical. This account avoids the objection by Reid and Sergeant, as psychological continuity is transitive. For the individual who was associated with my body twenty years ago and the individual who will be associated with my body fifty years from now to be one and the same, there need not be any direct psychological connections between the two, as long as the connections along the line of individuals that trace from the ten-year-old to the eighty-year-old are sufficient in number.

According to Parfit's account, which is commonly known as the *psychological account*, an individual's existence extends in either direction of the timeline until the point where there is no more psychological continuity.<sup>309</sup> As Maria's memories, belief, intentions,

---

<sup>306</sup> We have encountered that concept in our discussion of McMahan's account of the wrongness of killing in Chapter 2, when we talked about egoistic concern, and just now, when we talked about memory connections.

<sup>307</sup> Parfit 1984, p. 205.

<sup>308</sup> "[T]here is enough connectedness if the number of direct connections, over any day, is *at least half* the number that hold, over every day, in the lives of nearly every actual person. When there are enough direct connections, there is what I call *strong* connectedness" (Parfit 1984, p. 206).

<sup>309</sup> The possibility of division complicates things. We shall ignore this complication. Suffice it to note that psychological continuity is sufficient for personal identity only if there is no branching. If an individual is

desires, and other psychological features go wherever Maria's brain goes, the individual waking up in her husband's body after the transplantation procedure is Maria. The psychological account also implies that, once it has become dark, Benjamin is gone, which is similarly in line with our intuitions. That, however, immediately brings us to one of the psychological account's major shortcomings. In the case of Benjamin, it seems to overcorrect. Psychological continuity ceases to obtain well before it becomes dark, which would mean that Benjamin ceases to exist before his stream of consciousness ends. To me, it is equally impossible to believe that Benjamin is already gone well before it becomes dark, as it is to believe that Benjamin is still around when all that is left of his organism is just a cluster of cells. The problem is not restricted to hypothetical cases. The psychological account has similarly counterintuitive implications for normally-aging people. If we look at human beings in early stages of development, we will find considerably fewer psychological connections from one day to another, or from one week to another, than in the case of normal adult human beings. Likely, there are not sufficiently many psychological connections between a very young infant and the individual who inhabited his or her body the previous day for there to be strong connectedness and hence psychological continuity.<sup>310</sup> If so, you and I are not identical to the young infants from whom we evolved. That is in stark contrast to how we commonly think about ourselves. When presented with a picture of the infant whose organism grew into the organism you now inhabit, and asked who that infant is, you would say, without any second thoughts, "That's me!", and who would doubt you? After all, who else could

---

psychologically continuous with more than one individual, then he or she ceases to exist at the moment when the branching occurs. Hence, stated more precisely than above, *non-branching* psychological continuity constitutes personal identity.

<sup>310</sup> Cf. McMahan 2002, pp. 44 f.

it be? Similarly, it seems obvious that people suffering from Alzheimer's disease do not cease to exist when the number of psychological connections between consecutive days drops below a certain threshold, which is why the desire of somebody in an early stage of Alzheimer's disease to be well taken care of when he or she reaches the final stage is accurately described as self-interested rather than other-interested.<sup>311</sup> It would take very strong reasons for us to abandon these deeply seated intuitions about personal identity, and as far as I can tell there are no such reasons.

Parfit's psychological account faces another objection, which should not go unmentioned. It assumes that "the notion of continuity of psychology can be logically prior to the notion of continuity of person."<sup>312</sup> That assumption seems mistaken. Experiences, beliefs, and intentions are logically secondary to, and hence cannot be the building blocks of, the one who experiences, believes, and intends.<sup>313</sup> If you remember an experience you had five years ago, then it was necessarily you who had that experience. Otherwise, you would not actually remember the experience but merely believe, *falsely*, to remember it. As the English philosopher Peter Frederick Strawson has pointed out more than fifty years ago, "[s]tates of consciousness could not be ascribed at all, unless they were ascribed to persons."<sup>314</sup> Hence, when – in some particular case – we say that there is a direct memory or other psychological connection, we have already assumed identity, which makes the psychological account circular. Michael Lockwood seems to be making essentially the same point when he writes that "it's a mistake to try to *define* identity in

---

<sup>311</sup> We are here employing what Baruch Brody has called the *going-out-of-existence test*, which exploits the connection between identity over time and the essentiality of properties. Recall that an "object *a* has property *P* essentially just in case that *a* cannot lose it without going out of existence" (Brody 1974, p. 86). If we can lose the property of being psychologically continuous with our earlier selves without going out of existence, then we are not essentially psychologies of the kind Parfit suggests we are.

<sup>312</sup> Chappell 2004, p. 114.

<sup>313</sup> Cf. Chappell 2004, pp. 114 f.

<sup>314</sup> Strawson 1959, p. 102.

terms of such things as memory, continuity of personality and so forth,”<sup>315</sup> as “one’s identity over time is [...] conceived of as a *deep* fact: something we think of as lying behind these connections and continuities, something of which the latter are merely a manifestation.”<sup>316</sup> Psychological continuity is usually a good indicator that an individual is still around, just as continuity of color and shape is usually a good indicator that the lump of gold on my table is the same thing I saw sitting there a few moments ago. What makes the lump of gold I see now the same lump of gold I saw a few moments ago, however, is the atomic structure that underlies its more superficial properties. In unusual lighting conditions, the lump of gold, while still being one and the same, may not be the same color. Analogously,

“[p]rovided that the deep, underlying continuities, which alone are constitutive of personal identity, are, in sufficient measure, present, the person may be said to persist even if other factors cause the more superficial psychological continuities of memory, personality and so forth (that followers of Locke consider essential to bind the later to the earlier self) to be absent. [...] The essence of the person may, in that sense, survive a physical or emotional trauma that results in amnesia and sudden change of personality.”<sup>317</sup>

The account of individual identity that has recently been put forward by McMahan is recognizably in the psychological tradition, yet avoids the apparent circularity of Parfit’s account, and allows us to say that we were once very young infants. It also explains why it is rational to have self-interested concern for the one you might become if you have the unfortunate fate of receiving an Alzheimer’s diagnosis someday. McMahan takes up

---

<sup>315</sup> Lockwood 1988, p. 205.

<sup>316</sup> Lockwood 1988, p. 204.

<sup>317</sup> Lockwood 1988, p. 205.

Lockwood's thought, and insists that "our account of personal identity must focus [...] not on the superficial continuities of mental life but on that which underlies and sustains them."<sup>318</sup> He argues that the deep, underlying continuity constitutive of individual identity is the "physical and minimal functional continuity of the brain."<sup>319</sup> That continuity is closely related to psychological continuity, but the two can come apart, as is the case with young infants and some people with Alzheimer's and certain other diseases, where psychological connections over time are very weak. If they do come apart, the persistence of the affected individual is explained by the continued presence of the same consciousness, which is guaranteed by the physical and minimal functional continuity of the brain.<sup>320</sup> More precisely, then, McMahan suggests that the "criterion of personal identity is the continued existence and functioning, in nonbranching form, of enough of the same brain to be capable of generating consciousness or mental activity [...],"<sup>321</sup> which corresponds to the view that we are essentially embodied minds.

McMahan's *embodied mind account* of individual identity captures a good deal of widely shared intuitions, but it might seem overly strict. As Frances Kamm has noted, it is unclear why, for the same consciousness "to be present, the material substrate of consciousness must remain the same."<sup>322</sup> Individual identity over time, according to McMahan, requires that enough of the same brain continues to exist – hence his contention that teletransportation means death and the creation of a replica.<sup>323</sup> He does not require that all of the same brain continues to exist, as he is well aware of the fact that

---

<sup>318</sup> McMahan 2002, p. 69.

<sup>319</sup> McMahan 2002, p. 69.

<sup>320</sup> A brain is minimally functioning, according to McMahan, only as long it retains the capacity for consciousness.

<sup>321</sup> McMahan 2002, p. 68.

<sup>322</sup> Kamm 2007, p. 274.

<sup>323</sup> Cf. McMahan 2002, pp. 56 ff.

the molecules that make up the brain are continuously being replaced, due to metabolism. However, if we survive the slow and gradual replacement of the constituent matter of our brains over time, why should we not be able to survive a sudden and more substantive replacement of the constituent matter of our brains?

“[S]uppose it turned out to have always been true of our brains that the seat of consciousness moves, as cells in a previous area die en masse, with a seamless flow of consciousness throughout. Would we really think that no one had ever survived as long as we had previously thought?”<sup>324</sup>

If the identity of an individual is preserved in a hypothetical case like this, why could the same not be true in other cases, such as teletransportation, where consciousness also moves in unusual ways?

Imagine a time even further in the future than brain transplantation. A device has been invented which can transfer the psychological states of one brain to another. If, as we shall suppose, Maria is fortunate enough to have been born into that time, an accident such as the one described earlier no longer necessitates brain transplantation. Instead of moving her brain from her skull to her husband's, doctors at the hospital 3D-print a complete human body and then transfer Maria's memories, beliefs, character traits, and other psychological states to the brain of that just-created body. As in Kamm's example, we suppose that there is a seamless flow of consciousness throughout. That is, because the device not only transfers psychological states but is also a streamal diverter. I borrow the idea of a streamal diverter from Barry Dainton and Tim Bayne, who describe it as a “device [that] can be used to divert the flow of consciousness from one brain to another,

---

<sup>324</sup> Kamm 2007, p. 274.

in an instant.”<sup>325</sup> Whether or not a streamal diverter is physically possible is irrelevant, as what we are after is conceptual clarity. For that purpose, it is sufficient that such a device is conceivable. From Maria’s perspective, moving from one body to another is quite an unusual and possibly disturbing experience, but at no moment is her stream of consciousness interrupted. Each moment of experience flows seamlessly into the next, just as usual. To me, it seems clear that Maria will survive the procedure.

Compare this scenario with the teletransportation scenario discussed by McMahan. McMahan’s teletransportation device “creates, out of new matter, an exact, cell-for-cell duplicate of one’s original body.”<sup>326</sup> Simultaneously, the original body is destroyed. Is the replica the same individual as the one who inhabited the original body before its destruction? McMahan intuitively believes that the replica is not the same individual, and he counts it as an advantage of his account of individual identity that it can explain that intuition.<sup>327</sup> I myself have no clear intuition either way. Before making a judgment, I would like to know how the procedure is experienced from the perspective of the one being teletransported. As I see it, there are two ways to elaborate on McMahan’s thought experiment. One could tell the story such that there is a seamless flow of consciousness throughout, in which case I would intuitively tend to believe that the replica is the same individual. If it is rather like one light going black and another light being sparked, I would say that the replica is a different individual. What is missing in one case but not the other, and therefore seems to matter for individual identity, is a kind of continuity, phenomenal continuity, that we have not yet considered, and that forms the basis of a

---

<sup>325</sup> Dainton & Bayne 2005, p. 556.

<sup>326</sup> McMahan 2002, p. 56.

<sup>327</sup> The replica is not the same individual as the one who inhabited the original body before its destruction, because the replica’s brain is not (at all) physically continuous with the original brain.

recent proposal by Dainton and Bayne. “[E]xperiences at different times [...] are [...] phenomenally *continuous* with one another provided they are linked by an overlapping chain of direct phenomenal connections [...],”<sup>328</sup> and direct phenomenal connections are what make your stream of consciousness a stream rather than a collection of discrete experiential atoms.

“Think of what it is like to suffer a prolonged toothache, or to hear an extended tone played on a flute, or to watch a balloon float slowly across the sky. Each phase of your experience merges seamlessly with the next, and the next – indeed, so seamless is the flow that the division of experiences of this kind into distinct phases is often entirely arbitrary.”<sup>329</sup>

Phenomenal continuity is different from both psychological continuity, in Parfit’s sense, and the physical and minimal functional continuity of the brain, as it neither requires the persistence of any psychological states nor the same brain. The device used on Maria transfers all of Maria’s psychological states and guarantees continuity of experience throughout, and hence maintains both kinds of mental continuity. In contrast, consider a mere brain-state transfer. Suppose all of my psychological states are transferred to your brain, and all of your psychological states are transferred to mine, while our respective brains continue to sustain our respective streams of consciousness. No streamal diverter is being used. Dainton and Bayne suggest that “each person remains with their original body. The envisaged procedure amounts to no more than a form of brainwashing [...].”<sup>330</sup> I agree, and so would McMahan, as the physical and minimal functional continuity of our brains is maintained. Taken together, these two cases support the claim that you always

---

<sup>328</sup> Dainton & Bayne 2005, p. 554.

<sup>329</sup> Dainton & Bayne 2005, p. 554.

<sup>330</sup> Dainton & Bayne 2005, p. 557.



stay with your stream of consciousness, regardless of where your psychology goes, and regardless of whether the physical and minimal functional continuity of your brain is maintained. Dainton and Bayne call this point the *inseparability thesis*: “self and phenomenal continuity cannot come apart: all experiences in a single (non-branching) stream of consciousness are co-personal.”<sup>331</sup>

The inseparability thesis states a sufficient condition for individual identity over time. As long as phenomenal continuity is maintained, we continue to exist. However, this is not yet a full-fledged account of individual identity over time, as phenomenal continuity does not seem to be a necessary condition for our survival. If phenomenal continuity is to be our guide to individual persistence, we must deal with the problem posed by interruptions in the stream of consciousness that do not seem to threaten our survival, such as dreamless sleep. The challenge to explain how we can persist through periods of unconsciousness is called the *bridge problem*, which Dainton and Bayne suggest can be solved by “a shift in perspective. Rather than regarding persons as primarily things that *are* conscious, we regard them as things that *are capable* of being conscious, as beings that possess *capacities* for experience.”<sup>332</sup> Dainton and Bayne then propose, finally, that retaining the capacity for consciousness is both sufficient and necessary for our survival. This proposal is very close to McMahan’s, the embodied mind account, which demands that in order for an individual, on waking, to be the same individual he or she remembers being the day before, his or her consciousness today must be supported by the same brain that supported the consciousness of the individual he or she remembers being the day before. In other words, the brain of the individual who went to sleep yesterday must be

---

<sup>331</sup> Dainton & Bayne 2005, p. 557.

<sup>332</sup> Dainton & Bayne 2005, p. 565.

physically and minimally functionally continuous with the brain of the individual who woke up today for these two individuals to be one and the same. The difference between the two proposals comes down to this: While McMahan requires that, in order for us to persist, the material basis of our capacity for consciousness remains the same, Dainton and Bayne allow for the material basis of our capacity for consciousness to change, as long as the capacity itself persists. Of course, what it means for a capacity for consciousness to persist through such change needs explaining. Recall Maria, whose stream of consciousness is diverted to her husband's brain. She remains conscious throughout, and her capacity for consciousness hence trivially persists. But what if she had fallen into dreamless sleep just prior to the procedure, just to find herself in her husband's body when she wakes up? What would then make the capacity for consciousness before and after one and the same? Briefly, Dainton and Bayne suggest that Maria's capacity for consciousness would persist in virtue of it being the case that, if her brain had been active just prior to the doctors pushing the button, and if the brain that has formerly been her husband's had been active shortly thereafter, then the experiences her brain would have produced the moment before the transfer takes place and the experiences the brain that has formerly been her husband's would have produced the moment after the transfer has taken place would have been "diachronically phenomenally connected (i.e., [would have] occur[ed] within a single specious present)."<sup>333</sup>

In my estimation, the embodied mind account and the account put forward by Dainton and Bayne are the most plausible accounts among the ones we have discussed, and in line

---

<sup>333</sup> Dainton & Bayne 2005, p. 566.

with what I take to be widely shared intuitions.<sup>334</sup> I lean towards the latter, but fortunately we need not make a choice, which is why I did not go into the rather complex and technical details of Dainton's and Bayne's proposal.<sup>335</sup> It is sufficient to note that we have established what we set out to establish, which is the one thing McMahan, Dainton, and Bayne are in agreement about, that we are essentially capable of consciousness, and hence, according to what we have said in the previous section, essentially somebodies.<sup>336</sup>

#### ***4.4. The binary nature and moral significance of the capacity for phenomenal consciousness***

In the previous chapter, we have argued that having the potential for higher mental functions is a matter of degree, and hence a problematic basis for moral worth, in that a

---

<sup>334</sup> Of course, this has been no more than a brief survey of some of the most important arguments in the debate over personal identity, and the reader may well disagree with my evaluation. If the reader still feels drawn to the view that we are essentially organisms, or that we are essentially psychologies, I hope I could nevertheless show why reasonable people might endorse one of my preferred alternatives.

<sup>335</sup> Besides Dainton & Bayne 2005, Dainton 2000 is a good source for anybody who wants to know more.

<sup>336</sup> If we are essentially capable of consciousness, we come into existence when our brains acquire the capacity to generate consciousness, and we cease to exist – science fiction aside – when our brains lose that capacity. This view of when we cease to exist is reflected in the Ad Hoc Committee of the Harvard Medical School's widely accepted brain death criterion for death (cf. Ad Hoc Committee of the Harvard Medical School 1968). In a skeptical note about the brain death criterion, the neurologist D. Alan Shewmon wrote that "the introducers of the concept [of brain death] intended to redefine death in terms of unconsciousness rather than diagnose it as the cessation of biological life of the human organism" (Shewmon 1997, p. 81). With our distinction between us and our organisms in place, we can offer an explanation of what is going on here. We cease to exist ("die") when our brains lose the capacity for consciousness, which corresponds, give or take, to brain death. Our organisms, however, seem to have different persistence conditions. Just like our organisms were around before we were, there is no reason why it should not be possible for them to continue to exist, at least for a while, after we are gone. If a human organism without a heart can live on life support, then why should it not also be possible for a human organism without (most of) its brain to live on life support? We can therefore understand the distinction which Shewmon rightly draws between brain death and the cessation of biological life as the distinction between the death of the individual and the death of the organism. If we in fact cease to exist when we lose capacity for consciousness, regardless of whether or not our organisms continue to exist, then that in turn lends support to our claim that we are essentially capable of consciousness. This is the earlier-mentioned going-out-of-existence test, which Brody takes to support the view that "the foetus [...] [is] a living human being from about six weeks, the time at which we begin to note foetal brain activity" (Brody 1974, p. 88), where the term "living human being" is to be understood not to refer to an organism but the individual "with all the rights" (Brody 1974, p. 69).

threshold must be specified that is necessarily arbitrary, and further marks an entirely unremarkable empirical difference, which seems ill-suited to carry the momentous moral significance defenders of human dignity seek to attach to it. That the potential for higher mental functions is a matter of degree is true in two regards. First, all of the higher mental functions which are traditionally regarded as relevant for the possession of full moral worth themselves admit of degrees. Hence, as not just any minuscule capacity for rationality, moral personality, self-consciousness, and so on will do for those who want to justify a moral gulf between human beings and other animals, defenders of human dignity need to specify the degree to which an animal must potentially possess higher mental functions in order to qualify for full moral worth. Another way of putting this is to say that they must specify *which* potential an animal must possess to gain entrance into the community of moral equals – the potential to develop the relevant higher mental functions to the degree  $x$ , the potential to develop the relevant higher mental functions to the degree  $x+0.1$ , etc.? Now suppose we make a choice and say that the relevant degree is  $x$ . Instead of making the possession of a potential for higher mental functions a condition for full moral worth, we require, more precisely, the potential to develop the relevant higher mental functions to the degree  $x$ . Then, there is still the question of what is meant by “potential,” which brings us to the second regard in which the potential for higher mental functions is a matter of degree. Potential admits of both kind and degree. There are wide senses of “potential,” according to which almost anything has almost any potential.<sup>337</sup> For example, we have considered the proposal that anencephalic fetuses possess human dignity in virtue of having the potential for higher mental functions in the sense of it being physically possible for them to acquire the relevant higher mental

---

<sup>337</sup> Cf. Feinberg 2014.

functions. That, however, we have observed, would imply that many non-human animals too have full moral worth, which runs against the very idea of exclusively human dignity. If human dignity is to be successfully defended, the potential on which it is supposed to supervene must be, to borrow Joel Feinberg's expression, of a less "promiscuous" kind than the one tied to physical possibility. The less promiscuous sense which defenders of human dignity usually have in mind is exemplified by Lee's contention that

"[t]he human embryo has within herself all of the positive reality needed to actively develop herself to the point where she will perform higher mental functions, given only a suitable environment and nutrition, and so she now has the natural capacity for such mental functions."<sup>338</sup>

Unlike physical possibility, which is arguably a matter of either-or, potential in that sense comes in degrees. That point is made by McMahan when he writes that

"[i]n a paradigmatic case of internal directness or intrinsic potential, all that is needed from an external source for the realization of the potential is nutrition, hydration, shelter, and so on. But there is then a spectrum of possible cases in which in each succeeding case just a little more is needed from the outside for an individual to develop a rational nature."<sup>339</sup>

"Potential for higher mental functions" hence is an umbrella term for a large number of distinct properties which vary along two dimensions. One dimension corresponds to the different degrees of higher mental function, and the other to the more-or-less that is needed from the outside for an organism to realize a given degree of higher mental function.

---

<sup>338</sup> Lee 2004, p. 253.

<sup>339</sup> McMahan 2008, p. 92.

We will now argue that, whereas the potential for higher mental functions is doubly a matter of degree, the capacity for phenomenal consciousness is best understood as a binary concept, and hence, in that regard, a less problematic basis for our moral equality and worth. Like the potential for higher mental functions, the capacity for phenomenal consciousness is a compound concept, which combines two more basic concepts: the concept of capacity, and the concept of phenomenal consciousness. It will be helpful to look at these two concepts in turn, starting with the latter. Plainly, there is a sense in which one can be more or less phenomenally conscious, in that the inner lives of individuals can be more or less rich. Bats, for example, but not human beings, have the sense of echolocation, and their sensory experience hence is more rich than that of human beings in that regard. But there is also a sense, and that is the sense we shall use, in which phenomenal consciousness clearly seems to be a matter of either-or, or rather a matter of *on-or-off*. Either the light is on, or the light is off and all is dark within. For the light to be on, it is sufficient that there is *any* experience, as experience implies an experiencer with a distinct point of view, regardless of its richness.<sup>340</sup> That point of view either is occupied, or it is not, and phenomenal consciousness in the relevant sense is present whenever there is a subjectivity, or a center of consciousness.

The term “capacity” is sometimes used synonymously with the term “potential,” which we have just shown to admit of both kind and degree. This rather wide meaning, which is not the one we intend, must be carefully distinguished from a narrower meaning which covers the notions of both an immediately exercisable capacity and a blocked capacity, as defined by Michael Tooley.

---

<sup>340</sup> In contrast, access consciousness is a scalar property, as an entity can be more or less such that information is readily available, which is further reason to spell out being a somebody in terms of phenomenal rather than access consciousness.

“To attribute an immediately exercisable capacity to something is to make a statement about how the thing would be behaving, or what properties it would have, if it were now to be in certain circumstances, or in a certain condition.”<sup>341</sup>

Thus, to say that Abby, who is currently asleep and not dreaming, has the immediately exercisable capacity for consciousness is to say that she would be conscious if she was now awake. In contrast, an entity has a blocked capacity if “all of the ‘positive’ factors required for the immediately exercisable capacity are present, but there are also negative factors that prevent the exercise of the capacity.”<sup>342</sup> For example, suppose Abby is under general anesthesia. Her brain is intact, yet the action of the anesthetic drugs prevents her brain from generating consciousness. While Abby is anesthetized, she still has the capacity for consciousness, even though her current lack of consciousness is not just a matter of her not being in a certain condition. Only once the action of the anesthetic drugs, which currently impedes her capacity for consciousness, has ceased, she will regain the immediately exercisable capacity for consciousness. From now on, the term “capacity for consciousness” shall mean “immediately exercisable or blocked capacity for consciousness.” To say that an entity has the capacity for consciousness in that sense is to say that it has all the internal resources needed to generate consciousness. Abby is capable of consciousness while sleeping and while being under general anesthesia because in both cases the material substrate of her consciousness remains intact. Capacity so understood does not admit of degrees, as it either is the case that the material substrate of consciousness is intact and in principal functional, or it is not. Hence, just as Abby’s existence is a matter of either-or, so is her having the capacity for consciousness, which

---

<sup>341</sup> Tooley 1983, p. 149.

<sup>342</sup> Tooley 1983, p. 150.

we have argued is essential to her existence and the basis of her numerical identity over time.<sup>343</sup>

---

<sup>343</sup> There is a potential worry that I wish to acknowledge, but cannot entirely dispel. We have fixed the meaning of the term “capacity for consciousness” such that it refers to a binary property. We have said that an organism has the capacity for consciousness as long as it is constituted such as to permit phenomenal consciousness. The realization of the capacity for phenomenal consciousness, in that sense, does not involve a constitutive change of the one who has it, but merely requires a change of state, or the removal of negative factors that block the capacity. Now, consider the case, imagined by Tooley, of “an adult human being that has suffered brain damage that makes it impossible for the organism to enjoy any consciousness at all, let alone rational awareness. [...] [T]he damage might involve the complete destruction of those structures that are the positive constitutional basis of consciousness and rational awareness” (Tooley 1983, pp. 152 f.). The organism here described currently is not constituted such as to permit consciousness, in that the part of the brain that used to generate consciousness is no longer intact and in principal functional. The organism hence lacks the capacity for phenomenal consciousness in the sense we have specified. If we are essentially capable of consciousness, as we have argued, then the individual who inhabited the organism before the damage occurred – let us call him Oran – does not exist anymore. That might prove problematic, if we assume that the brain damage is repairable. I do not have the required medical expertise to say whether this is a real possibility, but it is certainly conceivable. Hence, suppose there is in fact a medical intervention, involving a constitutive change, which allows the organism to reacquire the capacity for consciousness. Is the individual who emerges after that intervention has taken place identical with Oran, or is he a mere replica of Oran? If the emerging individual is a mere replica, killing the organism during the time between the damage and its repair would arguably amount to preventing an individual like us from coming into existence, and hence, roughly, be the moral equivalent of contraception. That seems to fit with my account of the wrongness of killing, which I will spell out in more detail below. Insofar full moral status supervenes on the capacity for consciousness, as I propose, it would be less seriously wrong to kill the organism as it would have been to kill Oran while he was still around, or as it would be to kill Oran’s replica once he has come into existence. However, to many people it will seem more natural or plausible to say that Oran persists through the time of brain damage, and has full moral status throughout, which is why killing the brain-damaged organism is as seriously wrong as it is to kill you or me (cf., e.g., Thomas 1978). In order to accommodate this intuition, we would need to bridge the period of time during which Oran lacks the capacity for consciousness. This is as much a challenge for my account of the wrongness of killing as it is a challenge for the accounts of individual identity proposed by McMahan, Dainton, and Bayne, for much the same reason. Dainton and Bayne are aware of that. In a discussion of potential difficulties for their phenomenalist approach to personal identity, they consider the imaginary case of “a race of creatures who outwardly resemble human beings, but whose nervous systems undergo a restorative restructuring once per year [...]. [D]uring this time these creatures no longer possess any experiential capacities” (Dainton & Bayne 2005, pp. 568 f.). This case raises the same questions as the kind of brain damage that Tooley describes. Dainton and Bayne hint at a possible solution: “we can appeal to second-order experimental capacities” (Dainton & Bayne 2005, p. 569). It is unclear, however, whether there is an intelligible conception of second-order capacity that makes having it a binary property, as it should be, insofar existence, like full moral worth, is a matter of either-or. It should be noted that similarly puzzling cases can be constructed for human dignity accounts that are based on animalism, such as the one defended by Lee and George. Reversible organic death seems no less problematic than reversible loss of the capacity for consciousness. Suppose it was medically possible to bring human corpses back to live within some period of time after death. Would it be less seriously wrong to destroy a fresh corpse than it is to kill one of us? If not, how can animalists bridge the time between death and resurrection? One could say that a corpse has full moral status as long as it is resurrectable, just as one could say that loss of the capacity for consciousness does not imply the loss of full moral status as long as the neural substrate of consciousness is repairable. Yet, it is unclear whether resurrectability and repairability can be plausibly understood to be binary rather than continuous properties. For what it is worth, I tend to think that the brain-damaged organism, sans capacity for consciousness, is uninhabited and does not have full moral worth. If that is right, there is no problem. This is not to say that it is morally permissible to kill such an organism, as it



If that is right, then the capacity for consciousness in the sense of an immediately exercisable or blocked capacity for phenomenal consciousness of any degree of richness is a binary property, and making it the factual basis for our moral equality and worth hence avoids the embarrassment of having to draw an arbitrary line across a continuum. Further, unlike the difference between having a more or less remote potential for the development of certain higher mental functions to a lower or higher degree, or the difference between persons and non-persons, the difference between having and not having the capacity for consciousness is remarkable in empirical terms. It is the difference between there being somebody, and there not being anybody, between there being experience, a somewhat mysterious phenomena that continues to puzzle philosophers as well as scientists, and there not being experience.<sup>344</sup> Also, while it is unclear what could warrant the extraordinary moral importance the defenders of human dignity and McMahan attach to their respective demarcation lines, it is evident that the capacity for consciousness plays a central and important role in our moral thinking.

Being phenomenally conscious seems necessary for an entity's having a well-being and interests. That is not immediately obvious, as there are interests that arguably do not require phenomenal consciousness, as we see when we look at some of the most popular theories of well-being, which we have discussed in Chapter 1. While value-hedonism maintains that only experiences can be good or bad for us, the possession and satisfaction of certain desires does not obviously require phenomenal consciousness. Similarly, the

---

might be morally wrong to destroy a potential somebody of the kind at hand (or to prevent a somebody from coming back into existence), for some reason other than it being a failure to respect phenomenally conscious beings and their worth.

<sup>344</sup> The problem of accommodating phenomenal consciousness within a scientific picture of the world is called the *hard problem of consciousness*, cf., e.g., Chalmers 1996. There is no such problem for non-conscious organisms.

prudential goods postulated by objective-list theories of well-being, which typically include things like knowledge and achievement, do not all seem to require phenomenal consciousness. A phenomenal zombie can have desires, and it can have its desires satisfied, and there is no obvious reason why a phenomenal zombie should not be able to obtain knowledge or achieve things. Yet, it seems to be a mistake to attribute well-being and interests, in the sense that is morally significant, to phenomenal zombies. That is, as Guy Kahane and Julian Savulescu have argued, because, “[w]hen interests are promoted or set back, these are not just ways in which things in the world might go impersonally better or worse. It is someone’s *good* that is being promoted or set back.”<sup>345</sup> This echoes Frances Kamm’s discussion of the difference between saving a painting and saving a bird:

“I do not act for its sake when I save a work of art, because I do not think of its good and how continuing existence would be good for it when I save it. [...] Rather, I think of the good *of* the work of art, its worth as an art object, when I save it for no other reason than that it will continue to exist.

By contrast, when I save a bird, I can do it for its sake, because it will get something out of continuing to exist, and it could be a harm to it not to continue. It seems that something must already have or have had the capacity for sentience or consciousness in order for it to be harmed by not continuing on in existence. This is because an entity having such characteristics seems to be necessary for it to be a beneficiary or victim. It must be able to get something out of its continuing

---

<sup>345</sup> Kahane & Savulescu 2009, pp. 12 f.

existence, and capacity for sentience or consciousness seems to be necessary for this.”<sup>346</sup>

The point Kamm is making in this passage is, of course, not limited to continued existence but one about any kind of benefit or harm. Without phenomenal consciousness, there is nobody for whose sake one can act, who can be benefitted or harmed. This suggests that desire-satisfaction and objective goods like knowledge and achievement can be good only for somebody and never for something. If so, an entity can have interests only if it is phenomenally conscious.<sup>347</sup> Phenomenal consciousness also has close conceptual ties with the morally charged concept of empathy. It is only because others are conscious that we can try to empathize with them. If there is “nobody at home” in another being, there are no shoes you could put yourself into.<sup>348</sup> In contrast, if another being has a point of view, just like you do, then you can try to see the world from that being’s perspective.

#### ***4.5. The dignity of subjectivity account of the wrongness of killing***

The capacity for consciousness, in the sense that we have developed throughout the previous sections, is just the right kind of property to serve as a secular substitute for, or analogue of, the God-likeness of traditional morality. We have argued that we cannot exist at any time without having the capacity for phenomenal consciousness at that time. Basing full moral worth on that capacity hence guarantees that we are intrinsically

---

<sup>346</sup> Kamm 2006, pp. 228 f.

<sup>347</sup> For an opposing view, see Levy 2014.

<sup>348</sup> For Bernard Williams, “respecting a person involves putting oneself sympathetically in her shoes, looking inside that person and then looking out at the world from her point of view” (Carter 2011, p. 551).

valuable throughout our existence, from the moment when we come to be until the moment of death.<sup>349</sup> If the capacity for phenomenal consciousness is in fact an essential property of us, and if that property is the basis of our worth and also what makes us one another's moral equals, despite our numerous physical and mental differences, then we have found an account of moral worth according to which you and I are equal not because we are superficially the same insofar we happen to have certain accidental properties in common, but, as we would expect, because we are the same in a *fundamental* respect. On this account, unlike on McMahan's, the basis of respect is at the same time an essential property of the object of respect. We have further argued that having the capacity for phenomenal consciousness is not a matter of degree. Infants, advanced Alzheimer's patients, most human beings who are severely congenitally mentally retarded, and those with extremely low IQs no less have a subjective point of view than normal, healthy adults, and those with extremely high IQs. Hence, the account I propose, unlike McMahan's account and human dignity accounts, does not necessitate the drawing of a line across a continuum, and avoids the theoretical problems that come with such line-drawing.

For the lack of a better name, let us call my proposed answer to the question of why killing is wrong the *dignity of subjectivity account of the wrongness of killing*. According to this account, what makes it true that you and I have equal moral worth – so that, other things being equal, it is equally wrong to kill you and me – is the fact that we share the essence of having the capacity for phenomenal consciousness, on which our worth

---

<sup>349</sup> Note that this "we" excludes not only sperm and ovum but also, e.g., anencephalic infants, and hence does not include quite as many human beings as Lee and George would like to include. To my mind, that is an advantage rather than a disadvantage of my view. An anencephalic infant is not and never will be phenomenally conscious, and hence has no sake, no capacity for well-being, and no interests, as we have argued in the previous section.

supervenies. Killing a phenomenally conscious being, one of us, is normally wrong because it is a failure to show due respect for that being and his or her worth.

#### **4.6. *Some implications***

According to my account, what makes it morally wrong to kill you or me now would also have been present if we had been killed when we were late fetuses or small children. In contrast, if having the capacity for phenomenal consciousness is an essential property of us, then an early human fetus that does not have a point of view yet is not yet one of us, and the killing of such a fetus hence would not be as seriously wrong as the killing of you or me would be. This is not to say that killing sufficiently early fetuses is not wrong. Even though fetuses that do not have the capacity for consciousness do not have the worth of an experiencing subject, there may (or may not) be other reasons why destroying such a fetus is wrong, just like it might be wrong to destroy great pieces of art, or other things we care about. Similarly, on my account, human beings who have irreversibly lost their capacity for consciousness are no longer our moral equals, even though they might still qualify as living human organisms. If the capacity for consciousness is essential to who you are, it is possible that you cease to exist before your organism dies. If that happens, destroying the surviving organism is less seriously wrong than it would have been to kill you, but might nevertheless be wrong, for reasons other than failure to respect the intrinsic worth of an experiencing subject. These implications of my view for early fetuses and human beings who have irreversibly lost their capacity for phenomenal consciousness are more or less in accordance with the widely held belief

– reflected in the law of a substantial number of nations – that there is an important moral difference between early and late abortion, and with the popular brain-death definition of death.<sup>350</sup> It is the case of non-human animals where the implications of my view radically diverge from popular opinion.

If it is morally wrong to kill you or me primarily because it is a failure to show the respect we are due in virtue of being somebodies, then it is equally wrong to kill non-human animals who have the capacity for phenomenal consciousness, for the same reason. The question, however, of which non-human animals have that capacity is not an easy one to answer. The problem is a special case of the traditional philosophical problem of other minds.<sup>351</sup> Phenomenally conscious states are inherently private and directly accessible only to the ones whose states they are, which famously led Thomas Nagel to wonder whether we will ever know exactly what it is like to be a bat.<sup>352</sup> Yet, having said that, in our daily lives, we are very confident that common sense and empathy often provide us with sufficient evidence as to whether or not an animal is phenomenally conscious. If you see a dog bleeding and whining after having been hit by a car, there does not even seem to be a need to infer that he or she is in pain. You just see or know. John Searle has expressed this sentiment as follows:

“I do not infer that my dog is conscious, any more than, when I came into this room, I infer that the people present are conscious. [...] I just treat them as conscious beings and that is that. If somebody says, ‘Yes, but aren’t you ignoring the possibility that other people might be unconscious zombies, and the dog might be, as Descartes thought, a cleverly constructed machine, and that the chairs and

---

<sup>350</sup> Cf. Ad Hoc Committee of the Harvard Medical School 1968, and McMahan 2002, pp. 423 ff.

<sup>351</sup> Cf. Hyslop 2015.

<sup>352</sup> Cf. Nagel 1974.

tables might, for all you know, be conscious? Aren't you simply ignoring these possibilities?' The answer is: Yes. I am simply ignoring all of these possibilities. They are out of the question. I do not take any of them seriously. [...] [I]t does not matter really how I know whether my dog is conscious, or even whether or not I do 'know' that he is conscious. The fact is, he is conscious and epistemology in this area has to start with this fact."<sup>353</sup>

The claim that we do not infer that dogs, or other human beings for that matter, are conscious may well be true in particular situations, such as the one in our example with the injured dog, where we recognize a familiar pattern which we immediately and directly associate with pain, but it is not the whole truth. There *is* an interference in the background. After all, it is not literally true that I *see* that the dog who has been hit by a car is in pain. I see that he or she is injured, and I can hear his or her whining. My knowledge of the dog's pain is justified by the fact that he or she is very similar to me, in both physiological and behavioral terms, and that, were I the one lying on the streets whining after having been injured in an accident, I would be experiencing pain. Sometimes, though, what we take to be self-evidently true about the mental lives of non-human animals turns out to be no more than an inappropriate ascription of human emotion, such as when a cat "owner" is convinced that his or her cat is lovesick, whereas the cat is really suffering from constipation. This rather common tendency to misinterpret the behavior of non-human animals in light of our distinctively human experience is a form of anthropomorphism that we must be careful to avoid. Jane A. Smith, for example, warns that we might incorrectly conclude that other animals

---

<sup>353</sup> Searle 1998, p. 75.

“experience pain simply because they bear a (superficial) resemblance [to us] [...]. Equally, pain might incorrectly be denied in [...] [other animals] simply because they are so different from us and because we cannot imagine pain experienced in anything other than the [...] human sense.”<sup>354</sup>

Human experience, of course, will always be *a part* of any argument for phenomenally conscious states in other animals, as it is the only kind of experience of which we have direct knowledge. The investigation of phenomenal consciousness in non-human animals should hence be guided by a *critical anthropomorphism*, which

“involves the critical use of human experience to recognize [...] animal suffering [and other phenomenally conscious states in animals] by combining one’s perception of a particular animal’s situation with what can be determined by more objective, science-based observations.”<sup>355</sup>

What this approach, which is the one standardly used in animal science, proposes is that we extend the argument from analogy that constitutes the basis of our everyday judgments about the mental states of others to include indicators which are not available to everyday perception. These extra indicators are important to confirm and reinforce our everyday judgements, which might sometimes be false even when they seem obviously true, and even more important in the case of animals who look and behave very different than us.

One of the best-studied emotions in non-human animals is pain, which is partly explained by its immense practical relevance in the context of the use of non-human animals for human purposes. According to the widely-accepted definition by the International

---

<sup>354</sup> Smith 1991, p. 29.

<sup>355</sup> Nuffield Council on Bioethics 2005, p. 64.



Association for the Study of Pain, pain is “an unpleasant sensory and emotional experience associated with actual or potential tissue damage, or described in terms of such damage.”<sup>356</sup> Similarly, Donald M. Broom, an English biologist and professor of animal welfare at the University of Cambridge, defines pain as “an aversive sensation and feeling associated with actual or potential tissue damage.”<sup>357</sup> The first definition talks about an “experience,” the second one about a “sensation and feeling,” and both definitions hence imply phenomenal consciousness. If an animal can feel pain, he or she can feel, and therefore has the capacity for phenomenal consciousness. The reverse, however, is not necessarily true. There is no conceptual reason why phenomenal consciousness should not be possible without the ability to feel pain.<sup>358</sup> In fact, there is some evidence that jumping spiders might be phenomenally conscious yet lack the ability to feel pain:

“[A]lthough the brain is composed of a relatively small number of cells, the level of processing is considerable and sophisticated, if rather slow. Evidence for awareness is greater than in any other invertebrates except cephalopods but we have little evidence of a pain system.”<sup>359</sup>

The assumption that this is in fact a case of phenomenal consciousness without the ability to feel pain, however, would be mere speculation, in absence of further research, as the available evidence is very limited. In general, I am not aware of any scientific study that makes a strong case for the presence of phenomenal consciousness without the ability to feel pain in nature. For our purpose, which is just to get a rough sense of the range of

---

<sup>356</sup> International Association for the Study of Pain 1979, p. 250.

<sup>357</sup> Broom 2001, p. 17.

<sup>358</sup> Angels, for example, are phenomenally conscious, but cannot suffer tissue damage, and hence are not capable of suffering pain in the sense given by the above-quoted definitions.

<sup>359</sup> European Food Safety Authority 2005, p. 16.

animals to which my dignity of subjectivity account of the wrongness of killing applies, we may hence restrict our brief review of the scientific literature to the investigation of pain in animals.<sup>360</sup>

Pain in the just-defined sense – as opposed to, for example, phantom pain – presupposes the presence of receptors that respond to actually or potentially tissue-damaging stimuli. Such stimuli, which include cutting, extreme temperatures, pressure, and the like, are commonly referred to as noxious or *nociceptive stimuli*, and the receptors that respond to nociceptive stimuli are referred to as *nociceptors*. There is evidence that nociception is present in all vertebrates, sea anemones, ringed worms, snails, roundworms, fruit flies, California sea hares, medical leeches, and a number of other animals.<sup>361</sup> We should not be too quick to conclude, however, that all of these animals are capable of experiencing pain. “[A] paraplegic whose foot touches a hot iron will not feel anything, due to his spinal cord’s being severed,”<sup>362</sup> despite the presence of functional nociceptors. Nociception can be an indicator for the ability to feel pain, but the two must be carefully distinguished. A variety of strategies have been developed to make the distinction between pain and mere nociception, and we will mention the most important of them.

Pain is aversive, which is why an animal will normally try to escape it. Hence, if an animal behaves as in pain, that might indicate that he or she in fact is in pain. R. W. Elwood, S. Barr, and L. Patterson exploit that correlation between pain and behaviour in a study of pain in crustaceans. Starting from the premise that “if an animal responds to a

---

<sup>360</sup> To be clear and avoid possible misunderstandings, I am not proposing that killing is wrong because it causes pain. Causing pain and killing are two distinct moral wrongs. Our focus on pain is due to the fact that pain has received more attention from scientists than most other kinds of conscious states and was one of the earliest kinds of conscious states to occur in the evolution of life on earth. We use pain as an *indicator* for phenomenal consciousness, which is the ultimate object of our interest.

<sup>361</sup> Cf. Rose & Adams 1989, Mather 2011, Smith & Lewin 2009, Kavaliers, Hirst & Tesky 1983, Castellucci, Pinsker, Kupfermann & Kandel 1970, and Nicholls & Baylor 1968.

<sup>362</sup> DeGrazia & Rowan 1991, p. 195.

potentially noxious stimulus in a manner similar to that observed to the same stimulus in humans then it is reasonable to assume the animal has had an analogous experience [...],”<sup>363</sup> and having examined the responses of crustaceans to potentially noxious stimuli, they conclude that there is a significant possibility that crustaceans experience pain. J. L. Gould and C. G. Gould similarly use behavioral evidence to show that there is no good reason to assume pain experience in insects.<sup>364</sup> Eisemann et al. concur, observing that

“insects will continue with normal activities even after severe injury or removal of body parts. An insect walking with a crushed tarsus, for example, will continue applying it to the substrate with undiminished force [...]; caterpillars [...] continue to feed whilst tachinid larvae bore into them; many insects [...] go about their normal life whilst being eaten by large internal parasitoids; and male mantids [...] continue to mate as they are eaten by their partners.”<sup>365</sup>

One should keep in mind Smith’s warning, though. The fact that an animal reacts to tissue damage in a way that is very different from the way in which a human being would react does not guarantee the absence of pain.

“Species differ in their responses to painful stimuli: for example, dogs and humans make much noise but sheep do not, because loud vocalizations may elicit help from social group members in dogs and humans but just attract more predator attention to an injured sheep. Hence different responses are adaptive in different species. The feeling of pain may be the same even if the responses are very different.”<sup>366</sup>

---

<sup>363</sup> Elwood, Barr & Patterson 2009, p. 129.

<sup>364</sup> Cf. Gould & Gould 1982.

<sup>365</sup> Eisemann, Jorgensen, Merritt, Rice, Cribb, Webb & Zalucki 1984, p. 166.

<sup>366</sup> Broom 2014, p. 65.

Besides the presence of nociceptors and behavioural evidence, scientists have used several other indicators for pain experience in non-human animals, including the presence of

“a suitable central nervous system [...] avoidance learning, [...] physiological changes, [...] opioid receptors and evidence of reduced pain experience if treated with local anaesthetics or analgesics, and [...] high cognitive ability [...]”<sup>367</sup>

Evolutionary explanations of why the ability to experience pain, in contrast to mere nociception, represents a survival advantage are often used to further reinforce arguments for pain in non-human animals.

The case for phenomenal consciousness in non-human animals might seem incomparably weaker than the case for phenomenal consciousness in human beings, because human beings can talk and tell us about their inner lives.<sup>368</sup> Verbal reports, however, too have limited evidentiary value and constitute only one indicator among others. After all, other people could be philosophical zombies.<sup>369</sup> “[T]he difficulties in diagnosing pain [...] [and phenomenal consciousness] in animals vs. humans are not different in kind, but different in degree.”<sup>370</sup> The overwhelming majority of scientists and philosophers today agree with the common sense belief that some animals other than human beings are phenomenally conscious. In 2012, a prominent international group of scientists put together the *Cambridge Declaration on Consciousness*, in which they proclaimed that

---

<sup>367</sup> Elwood, Barr & Patterson 2009, p. 129.

<sup>368</sup> Some philosophers, such as Daniel C. Dennett and Peter Carruthers, have argued that phenomenal consciousness is uniquely human, even though Carruthers thinks that chimpanzees might be an exception, cf. Dennett 1995, and Carruthers 1998 & 2000. Among contemporary philosophers, however, they are a minute minority.

<sup>369</sup> Cf. Kirk 2015.

<sup>370</sup> Würbel 2009, p. 123.

“[c]onvergent evidence indicates that non-human animals have the neuroanatomical, neurochemical, and neurophysiological substrates of conscious states along with the capacity to exhibit intentional behaviors. Consequently, the weight of evidence indicates that humans are not unique in possessing the neurological substrates that generate consciousness. Non-human animals, including all mammals and birds, and many other creatures, including octopuses, also possess these neurological substrates.”<sup>371</sup>

A thorough, yet not fully comprehensive, review of the relevant scientific literature, and a meta-review of two recent reviews by Broom and Joel P. MacClellan, which I found particularly helpful, suggests the following, more detailed picture:<sup>372</sup>

	Capacity to experience pain/ for phenomenal consciousness
Hominids	<i>Certain/almost certain</i>
Other mammals	<i>Almost certain</i>
Parrots & birds in the crow family	<i>Almost certain</i>
Other birds	<i>Very likely</i>
Amphibians	<i>Likely</i>
Reptiles	<i>Likely</i>
Fish	<i>Likely</i>
Decapod crustaceans (crayfish, crabs, lobsters, prawns, shrimp, ...)	<i>Likely</i>
Cephalopods (octopi, squids, ...)	<i>Likely</i>

<sup>371</sup> The Cambridge Declaration on Consciousness 2012.

<sup>372</sup> Cf. Sarjeant 1969, Wells 1978, Gould & Gould 1982, Eisemann, Jorgensen, Merritt, Rice, Cribb, Webb & Zalucki 1984, Smith 1991, Gentle 1992, Machin 1999, Sneddon, Braithwaite & Gentle 2003, Chandroo, Duncan & Moccia 2004, Elwood, Barr & Patterson 2009, Braithwaite 2010, Mosley 2011, Elwood 2011, MacClellan 2012, pp. 180 ff., and Broom 2014, p. 122.

Gastropod mollusks (snails, slugs, swimming sea slugs, ...)	<i>Maybe</i>
Spiders	<i>Unlikely</i>
Insects	<i>Unlikely</i>
Sponges	<i>Unlikely</i>
Cnidarians	<i>Unlikely</i>
Other invertebrates	<i>Mixed likelihoods</i>

Even though this is only a rough picture, we may assume, with a reasonable degree of certainty, that at least mammals and birds have the capacity for phenomenal consciousness. On my view, that makes relevantly normal, developed members of these taxonomic classes intrinsically valuable subjects of experience whom it is no less seriously wrong to kill as it is to kill you or me, other things being equal. But what does it mean to say that killing you or me is normally no less seriously wrong than killing, say, a squirrel? We have briefly discussed degrees of wrongness in Chapter 1. Let us recall and elaborate on that discussion. We said that there is no room for degrees, if the word “wrong” is used merely to indicate whether or not an act is morally permissible, all things considered. In that sense, wrongness is an all-or-nothing matter. An action either is morally permissible, or it is not. We have noted, however, that there is also another sense, in which an act is more seriously wrong than another, if the reasons why it morally ought not to be done are stronger, and it hence would take weightier countervailing considerations to counterbalance or outweigh these reasons.<sup>373</sup> The degree of wrongness of acts hence indicates how hard or easy it is to morally justify them. An act (X) is *as seriously wrong as* another act (Y), if the reasons why X morally ought not to be done are

<sup>373</sup> Cf. McMahan 2002, p. 190, and Lippert-Rasmussen 2007, p. 717.

as strong as the reasons why Y morally ought not to be done and it would hence take just as much to justify X as it would take to justify Y (if it is possible to justify X and Y at all). Therefore, saying that killing you or me is as seriously wrong as killing a squirrel implies, for example, that, if a sufficiently large benefit to a sufficiently large number of conscious beings can (cannot) morally justify the killing of you or me, then the same benefit can (cannot) justify the killing of a squirrel, and vice versa. This has radical implications for the morality of the many ways in which we routinely use or otherwise affect non-human animals. For example, if the benefits for others that a deadly medical experiment on you or me would have cannot morally justify conducting it, then a deadly experiment on a squirrel that would have comparable benefits for others cannot be justified either, other things being equal. Also, as virtually nobody believes that a gustatory preference for human flesh over plant foods can morally justify the killing of human beings for food, consistency would seem to require that we stop killing phenomenally conscious non-human animals – who make up a substantial portions of the animals we eat – for food. These are just two examples of adjustments to our behavior we ought to make if the argument of this dissertation is sound, and they likely constitute little more than the proverbial tip of the iceberg. My account implies a rethinking of our practices that goes way beyond whom we eat or experiment on. We have to rethink how we build and produce things, how we move around, how we entertain ourselves and others, how we do sports, and how we worship, as all of these aspects of human life frequently involve the intentional or unintentional killing of our non-human equals.<sup>374</sup>

---

<sup>374</sup> To give just one less obvious example of the far-reaching implications of my account, consider the commercial production of crops, which often requires field operations that can be deadly to animals of the field, such as plowing, planting, and harvesting. Steven L. Davis estimates that fifteen animals of the field, including opossums, rats, mice, and rabbits, are killed per hectare of cropland harvested in the United

#### 4.7. *Epilogue*

Mohandas K. Gandhi wrote in his autobiography that, to his mind, “the life of a lamb is no less precious than that of a human being.”<sup>375</sup> When I first read this line, it immediately struck me as true, and I still believe it is. We, and that includes many non-human animals, are all here through no fault of our own, and we all have an equal right to be here and live and enjoy our lives according to our abilities and capacities, however rich or limited they may be.<sup>376</sup> Yet, it is probably safe to assume that most people believe that the life of a human being is more precious than the life of a lamb, and would hence intuitively reject the implications of my view for what we owe to non-human animals. This is not surprising. From Aristotle onward, philosophers and theologians – with very few exceptions – have again and again insisted that human beings are the crown of creation, and that other animals are less significant in the “hierarchy of nature.” Part of what underlies and explains this insistence is the tendency, widespread among people in

---

States per year, cf. Davis 2003, p. 390. This number has proved to be highly controversial, but his conclusion that “[m]illions of animals of the field die every year to provide products used in vegan diets” (Davis 2003, p. 393) nevertheless draws our attention to an important and potentially troubling fact, regardless of whether it really is millions of animals, or “just” thousands. We certainly would not morally tolerate an industry that knowingly kills very large numbers of human beings every year, even if we were convinced that the deaths are not intended. If the animals of the field are our moral equals, how can we then tolerate that they are routinely killed? This is a problem, as we have to eat and it may seem as if there is no practical or viable alternative to modern agriculture, but it might not be as big a problem as it at first appears. Rethinking the way in which we produce crops, we might find a low-cost method to reduce the number of field animal deaths such that it becomes comparable to the number of unintended but foreseeable human deaths in industries such as construction and truck transportation that we find acceptable. We might be able to do this, for example, by equipping tractors with loudspeakers that emit sounds at frequencies that are known to shoo field animals away.

<sup>375</sup> Gandhi 1925-1928, p. 235.

<sup>376</sup> The American naturalist writer Henry Beston expresses a similar sentiment in his book, *The Outermost House*: “We need another and a wiser and perhaps a more mystical concept of animals. [...] We patronize them for their incompleteness, for their tragic fate of having taken form so far below ourselves. And therein we err, and greatly err. For the animal shall not be measured by man. In a world older and more complete than ours they move finished and complete, gifted with extensions of the senses we have lost or never attained, living by voices we shall never hear. They are not brethren, they are not underlings; they are other nations, caught with ourselves in the net of life and time, fellow prisoners of the splendor and travail of the earth” (Beston 1971, p. 19).



general, to identify more with individuals similar to one's self. We see this tendency not only in the relationship between human beings and other animals, but also in how human beings have been relating to, and thinking about, each other throughout history. Aristotle, for example, was convinced that "[t]he relation of male to female is naturally that of the superior to the inferior – of the ruling to the ruled [...],"<sup>377</sup> and bolstered his claim with questionable scientific assertions about the inferior deliberative faculties of women, and David Hume wrote, some twenty-one centuries later: "I am apt to suspect the negroes to be naturally inferior to the whites."<sup>378</sup> Sexism and racism have since been largely overcome, at least in theory. We have learned to see people of the other sex or other races as more similar to us than different, and we now accept that the undeniable variation within humanity, not only in terms of looks and other merely physical characteristics, but also in terms of psychological potential and ability, does not undermine the ideal of moral equality and is, in that sense, superficial. I have argued that we ought to go one step further, and challenged the orthodox view that human life is more precious than the life of other animals. Acceptance of the radically egalitarian account of the wrongness of killing that I have outlined in my dissertation will not come easy. Not only do the outward differences between human beings and other animals seem far greater than those among human beings, human beings also have an enormous interest in maintaining the status quo. Most importantly, we want to eat other animals, and "[i]n order to be able to eat animals, many people feel that they must denigrate and devalue them by saying that they are stupid or are in some way less similar to humans than they really are."<sup>379</sup> I am nevertheless hopeful that this exercise was worthwhile. After all, many ethical ideas were

---

<sup>377</sup> Aristotle, *Politics*, Book I, Part V, 1254b2.

<sup>378</sup> Hume 1777, p. 550.

<sup>379</sup> Broom 2014, p. 4.

thought absurd before they were eventually accepted, and often the force of philosophical argument played a significant role in that.

## Bibliography

- Ad Hoc Committee of the Harvard Medical School, "A Definition of Irreversible Coma: Report of the Ad Hoc Committee of the Harvard Medical School to Examine the Definition of Brain Death," *The Journal of the American Medical Association* 205 (1968), pp. 337-340.
- Adams, Robert, "Motive Utilitarianism," *Journal of Philosophy* 73 (1976), pp. 467-481.
- Anscombe, Elizabeth, "Modern Moral Philosophy," *Philosophy* 33 (1985), pp. 1-19.
- Aristotle, *Politics*, URL = <<http://classics.mit.edu/Aristotle/politics.html>> [accessed on February 8, 2016].
- Behan, David P., "Locke on persons and personal identity," *Canadian Journal of Philosophy* 9 (1979), pp. 53-75.
- Benatar, David, *Better Never to Have Been* (Oxford: Oxford University Press, 2006).
- Benn, Stanley I., "Abortion, Infanticide, and Respect for Persons," in: Feinberg, J. (ed.), *The Problem of Abortion* (Belmont: Wadsworth, 1973), pp. 135-144.
- Bentham, Jeremy, *An Introduction to the Principles of Morals and Legislation* (1789a; reprint, London: T. Pain, 1823).
- \_\_\_\_\_, "An Introduction to the Principles of Morals and Legislation" (1789b), reprinted in: Bowring, John (ed.), *The Works of Jeremy Bentham*, vol. 1 (New York: Russell and Russell, 1962), pp. 1-154.
- Beston, H., *The Outermost House* (New York: Ballantine Books, 1971).

- Bicknell, Jeanette, "Love, Beauty, and Yeats's 'Anne Gregory,'" *Philosophy and Literature* 34 (2010), pp. 348-358.
- Block, Ned, "On a confusion about a function of consciousness," *Behavioral and Brain Science* 18 (1995), pp. 227-287.
- Bourget, David & Chalmers, David J., "What do philosophers believe?," *Philosophical Studies* 170 (2014), pp. 465-500.
- Bradley, Ben, "The Worst Time to Die," *Ethics* 118 (2008), pp. 291-314.
- Braithwaite, V. A., *Do Fish Feel Pain?* (Oxford: Oxford University Press, 2010).
- Brandt, Richard B., *A Theory of the Good and the Right* (Oxford: Clarendon Press, 1979).
- \_\_\_\_\_, *Morality, Utilitarianism, and Rights* (Cambridge: Cambridge University Press, 1992).
- \_\_\_\_\_, "Toward a Credible Form of Utilitarianism," in: Darwall, Stephen (ed.), *Consequentialism* (Oxford: Blackwell Publishing, 2003), pp. 207-235.
- Brody, Baruch, "On the Humanity of the Foetus," in: Perkins, Robert L. (ed.), *Abortion: Pro and Con* (Cambridge: Schenkman Publishing Company, 1974), pp. 69-90.
- Broom, Donald M., "Evolution of pain," in: Soulsby, E. J. L. & Morton, D. (eds.), *Pain: Its Nature and Management in Man and Animals, Royal Society of Medicine International Congress Symposium Series 246* (London: Royal Society of Medicine, 2001), pp. 17-21.
- \_\_\_\_\_, *Sentience and Animal Welfare* (Wallingford/Boston: CABI, 2014).
- Bykvist, Krister, *Utilitarianism* (London/New York: Bloomsbury, 2010).
- Carruthers, Peter, "Animal Subjectivity," *Psyche* 4 (1998), URL = <http://journalpsyche.org/files/0xaa52.pdf> [accessed on March 4, 2016].

- \_\_\_\_\_, *Phenomenal Consciousness: A naturalistic theory* (Cambridge: Cambridge University Press, 2000).
- Carson, Thomas L., "Utilitarianism and the Wrongness of Killing," *Erkenntnis* 20 (1983), pp. 49-60.
- \_\_\_\_\_, "Hare's defense of utilitarianism," *Philosophical Studies* 50 (1986), pp. 97-115.
- \_\_\_\_\_, *Value and the Good Life* (Notre Dame: University of Notre Dame Press, 2000).
- Carter, Ian, "Respect and the Basis of Equality," *Ethics* 121 (2011), pp. 538-571.
- Castellucci, V., Pinsker, V., Kupfermann, I. & Kandel, E. R., "Neuronal mechanisms of habituation and dishabituation of the gill-withdrawal reflex in *Aplysia*," *Science* 167 (1970), pp. 1745-1748.
- Cavalieri, Paola, *The Animal Question: Why Nonhuman Animals Deserve Human Rights* (Oxford/New York: Oxford University Press, 2001).
- \_\_\_\_\_, *The Death of the Animal: A Dialogue* (New York: Columbia University Press, 2008).
- Cavalieri, Paolo & Singer, Peter (eds.), *The Great Ape Project: Equality Beyond Humanity* (New York: St. Martin's Press, 1993).
- Chalmers, David J., *The Conscious Mind: In Search of a Fundamental Theory* (Oxford/New York: Oxford University Press, 1996).
- Chandroo, K. P., Duncan, I. J. H. & Moccia, R. D., "Can fish suffer? Perspectives on sentience, pain, fear and stress," *Applied Animal Behaviour Science* 86 (2004), pp. 225-250.
- Chappell, Timothy, "Absolutes and Particulars," in: O'Hear, Anthony (ed.), *Modern Moral Philosophy* (Cambridge: Cambridge University Press, 2004), pp. 95-117.

- Coetzee, J. M., *Elizabeth Costello* (London: Vintage, 2004).
- Cohen, Carl, "The Case for the Use of Animals in Biomedical Research," *New England Journal of Medicine* 315 (1986), pp. 865-870.
- Cox, Michael, "Animal liberation: A critique," *Ethics* 88, pp. 106-118.
- Craig, Winston J. & Mangels, Ann R., „Position of the American Dietetic Association: Vegetarian Diets," *Journal of the American Dietetic Association* 109 (2009), pp. 1266-1282.
- Dainton, Barry & Bayne, Tim, "Consciousness as a guide to personal persistence," *Australasian Journal of Philosophy* 83 (2005), pp. 549-571.
- Dainton, Barry, *Stream of Consciousness: Unity and Continuity in Conscious Experience* (London: Routledge, 2000).
- Dalal, Neil & Taylor, Chloë (eds.), *Asian Perspectives on Animal Ethics: Rethinking the Nonhuman* (Abingdon/New York: Routledge, 2014).
- Davis, Steven L., "The Least Harm Principle May Require that Humans Consume a Diet Containing Large Herbivores, Not a Vegan Diet," *Journal of Agricultural and Environmental Ethics* 16 (2003), pp. 387-394.
- Dawkins, Richard, "Gaps in the Mind," in: Cavalieri, Paolo & Singer, Peter (eds.), *The Great Ape Project: Equality Beyond Humanity* (New York: St. Martin's Press, 1993), pp. 80-87.
- DeGrazia, David & Rowan, Andrew, "Pain, suffering, and anxiety in animals and humans," *Theoretical Medicine* 12 (1991), pp. 193-211.
- DeGrazia, David, *Taking Animals Seriously* (Cambridge: Cambridge University Press, 1996).

- \_\_\_\_\_, "Identity, Killing, and the Boundaries of Our Existence," *Philosophy and Public Affairs* 31 (2003), pp. 413-442.
- \_\_\_\_\_, *Human Identity and Bioethics* (Cambridge: Cambridge University Press, 2005).
- Delaney, Neil, "Romantic Love and Loving Commitment: Articulating a Modern Ideal," *American Philosophical Quarterly* 33 (1996), pp. 339-356.
- Dennett, Daniel C., "Animal consciousness and why it matters," *Social Research* 62 (1995), pp. 691-710.
- Descartes, René, *Meditations on First Philosophy* (1641; reprint, Cambridge: Cambridge University Press, 1996).
- Devine, Philip E., *The Ethics of Homicide* (Ithaca/London: Cornell University Press, 1978).
- Dworkin, Gerald, "Autonomy and Behavior Control," *The Hastings Center Report* 6 (1976), pp. 23-28.
- \_\_\_\_\_, *The Theory and Practice of Autonomy* (Cambridge: Cambridge University Press, 1988).
- Dworkin, R. M., *Taking Rights Seriously* (Cambridge, MA: Harvard University Press, 1977).
- \_\_\_\_\_, *Life's Dominion: An Argument About Abortion, Euthanasia, and Individual Freedom* (New York: Random House, 1994).
- Ebert, Rainer & Oduor, Reginald M. J., "The Concept of Human Dignity in German and Kenyan Constitutional Law," *Thought and Practice: A Journal of the Philosophical Association of Kenya* 4 (2012), pp. 43-73.
- Ebert, Rainer, "Good to die," *Diacrítica* 27 (2013), pp. 139-156.

- Eisemann, C. H., Jorgensen, W. K., Merritt, D. J., Rice, M. J., Cribb, B. W., Webb, P. D. & Zalucki, M. P., „Do insects feel pain? A biological view,” *Experientia* 40 (1984), pp. 164-167.
- Elwood, R. W., “Pain and Suffering in Invertebrates?,” *ILAR Journal* 52 (2011), pp. 175-184.
- Elwood, R. W., Barr, S. & Patterson, L., “Pain and stress in crustaceans?,” *Applied Animal Behaviour Science* 118 (2009), pp. 128-136.
- Epicurus, “Letter to Menoeceus”, in: Oates, Whitney J. (ed.), *The Stoic and Epicurean Philosophers* (New York: The Modern Library, 1940), pp. 30-31.
- Ereshefsky, Marc, “Species,” in: Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2010 Edition), URL = <http://plato.stanford.edu/archives/spr2010/entries/species/> [accessed on January 28, 2016].
- European Food Safety Authority (EFSA), “Opinion of the Scientific Panel on Animal Health and Welfare on a request from the Commission related to ‘Aspects of the Biology and Welfare of Animals Used for Experimental and Other Scientific Purposes,’” *EFSA Journal* 292 (2005), pp. 1-46.
- Ewin, R. E., “What is wrong with killing people?,” *The Philosophical Quarterly* 22 (1972), pp. 126-139.
- Feinberg, Joel, “The Paradoxes of Potentiality,” in: Lizza, John P. (ed.), *Potentiality: Metaphysical and Bioethical Dimensions* (Baltimore: John Hopkins University Press, 2014), pp. 69-71.
- Feldman, Fred, “Hare’s proof,” *Philosophical Studies* 45 (1984), pp. 269-283.



- \_\_\_\_\_, *Confrontations with the Reaper* (Oxford/New York: Oxford University Press, 1992).
- Finnis, John, *Natural Law and Natural Rights* (Oxford: Clarendon Press, 1980).
- \_\_\_\_\_, "A Philosophical Case Against Euthanasia," in: Keown, John (ed.), *Euthanasia Examined: Ethical, Clinical, and Legal Perspectives* (Cambridge: Cambridge University Press, 1995), pp. 23-35.
- Fitzgerald, F. Scott, *The Curious Case of Benjamin Button* (1922; reprint, Juniper Grove, 2008).
- Fletcher, George P., *Our Secret Constitution: How Lincoln Redefined American Democracy* (Oxford/New York: Oxford University Press, 2001).
- Foot, Philippa, "The Problem of Abortion and the Doctrine of the Double Effect," *Oxford Review* 5 (1967), pp. 5-15.
- Fox, Michael Allen, *The Case for Animal Experimentation: An Evolutionary and Ethical Perspective* (Berkeley: University of California Press, 1986).
- Freed, William J., *Neural Transplantation: An Introduction* (Cambridge/London: MIT Press, 2000).
- Frey, R. G., "Introduction: Utilitarianism and Persons," in: Frey, R. G. (ed.), *Utility and Rights* (Minneapolis: University of Minnesota Press, 1984), pp. 3-19.
- Gandhi, Mohandas K., *The Story of My Experiments with Truth* (1925-1928; reprint, Boston: Beacon Press, 1993).
- Gentle, M. J., "Pain in Birds," *Animal Welfare* 1 (1992), pp. 235-247.
- George, Robert P. & Gómez-Lobo, Alfonso, "Statement of Professor George (joined by Dr. Gómez-Lobo)," in: The President's Council on Bioethics (ed.), *Human*

- Cloning and Human Dignity: An Ethical Inquiry* (Washington, D.C.: U.S. Government Printing Office, 2002), pp. 258-266.
- Giubilini, Alberto & Minerva, Francesca, “After-birth abortion: why should the baby live?”, *Journal of Medical Ethics* 39 (2012), pp. 261-263.
- Glover, Jonathan, *Causing Death and Saving Lives* (Harmondsworth: Penguin, 1977).
- Gómez-Lobo, Alfonso, *Morality and the Human Goods: An Introduction to Natural Law Ethics* (Washington, D.C.: Georgetown University Press, 2002).
- \_\_\_\_\_, “On Potentiality and Respect for Embryos: A Reply to Mary Mahowald,” *Theoretical Medicine and Bioethics* 26 (2005), pp. 105-110.
- Gould, J. L. & Gould, C. G., “The insect mind: physics or metaphysics?”, in: Griffin, D. R. (ed.), *Animal Mind – Human Mind* (Berlin: Springer, 1982), pp. 269-298.
- Griffin, James, *Well-Being: Its Meaning, Measurement, and Moral Importance* (Oxford: Clarendon Press, 1986).
- Hare, R. M., *Moral Thinking* (Oxford: Clarendon Press, 1981).
- Harman, Elizabeth, “Creation Ethics: The Moral Status of Early Fetuses and the Ethics of Abortion,” *Philosophy and Public Affairs* 28 (1999), pp. 310-324.
- Heathwood, Chris, “Desire Satisfactionism and Hedonism,” *Philosophical Studies* 128 (2006), pp. 539-563.
- Henson, Richard G., “Utilitarianism and the wrongness of killing,” *The Philosophical Review* 80 (1971), pp. 320-337.
- Hey, J., “The mind of the species problem,” *Trends in Ecology and Evolution* 16 (2001), pp. 326-329.
- Hooker, Brad, *Ideal Code, Real World* (Oxford: Oxford University Press, 2000).

Hume, David, *Essays and Treatises on Several Subjects, Vol. 1* (London: T. Cadell, and Edinburgh: A. Donaldson & W. Creech, 1777).

Hyslop, Alec, "Other Minds," in: Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2015 Edition), URL = <http://plato.stanford.edu/archives/fall2015/entries/other-minds/> [accessed on February 7, 2016].

International Association for the Study of Pain (IASP), "Subcommittee on taxonomy: pain terms; A list with definitions and notes on usage," *Pain* 6 (1979), pp. 249-252.

Jamieson, Dale, "Utilitarianism and the Morality of Killing," *Philosophical Studies* 45 (1984), pp. 209-221.

Kagan, Shelly, *The Limits of Morality* (Oxford: Clarendon Press, 1989).

\_\_\_\_\_, *Normative Ethics* (Boulder: Westview Press, 1998).

Kahane, Guy & Savulescu, Julian, "Brain Damage and the Moral Significance of Consciousness," *Journal of Medicine and Philosophy* 34 (2009), pp. 6-26.

Kamm, Frances M., *Morality, Mortality, Volume II* (Oxford/New York: Oxford University Press, 1996).

\_\_\_\_\_, *Intricate Ethics: Rights, Responsibilities, and Permissible Harm* (Oxford/New York: Oxford University Press, 2006).

\_\_\_\_\_, "The Ethics of Killing: Problems at the Margins of Life by Jeff McMahan," *The Philosophical Review* 116 (2007), pp. 273-280.

Kass, Leon R., *Life, Liberty, and the Defense of Dignity: The Challenge for Bioethics* (San Francisco: Encounter Books, 2002).

\_\_\_\_\_, “Defending Human Dignity,” in: The President’s Council on Bioethics (ed.), *Human Dignity and Bioethics* (Washington, D.C.: U.S. Government Printing Office, 2008), pp. 297-331.

Kavaliers, M., Hirst, M. & Tesky, G. C., “A functional role for an opiate system in snail thermal behavior,” *Science* 330 (1983), pp. 99-103.

Kirk, Robert, “Zombies,” in: Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* (Summer 2015 Edition), URL = <http://plato.stanford.edu/archives/sum2015/entries/zombies/> [accessed on February 7, 2016].

Kittay, Eva Feder, “At the Margins of Moral Personhood,” *Ethics* 116, pp. 100-131.

Kumar, Rahul, “Permissible Killing and the Irrelevance of Being Human,” *The Journal of Ethics* 12 (2008), pp. 57-80.

Kymlicka, Will, *Contemporary Political Philosophy: An Introduction* (Oxford: Oxford University Press, 1990).

Lamb, Roger E., “Love and Rationality,” in: Lamb, Roger E. (ed.), *Love Analyzed* (Boulder: Westview Press, 1997), pp. 23-47.

Lee, Patrick & George, Robert P., *Body-Self Dualism in Contemporary Ethics and Politics* (Cambridge: Cambridge University Press, 2007).

\_\_\_\_\_, “The Nature and Basis of Human Dignity,” in: The President’s Council on Bioethics (ed.), *Human Dignity and Bioethics* (Washington, D.C.: U.S. Government Printing Office, 2008a), pp. 409-433.

\_\_\_\_\_, “The Nature and Basis of Human Dignity,” *Ratio Juris* 21 (2008b), pp. 173-193.

- Lee, Patrick, "The Moral Status of Human Embryos," in: Lund-Molfese, Nicholas C. & Kelly, Michael L. (eds.), *Human Dignity and Reproductive Technology* (Lanham: University Press of America, 2003), pp. 71-80.
- \_\_\_\_\_, "The Pro-Life Argument from Substantial Identity: A Defense," *Bioethics* 18 (2004), pp. 249-263.
- \_\_\_\_\_, "Substantial identity and the right to life: A rejoinder to Dean Stratton," *Bioethics* 21 (2007), pp. 93-97.
- \_\_\_\_\_, "Substantial Identity, Rational Nature, and the Right to Life," in: Tollefsen, Christopher (ed.), *Bioethics with Liberty and Justice* (Heidelberg: Springer, 2011), pp. 23-40.
- \_\_\_\_\_, "The Basis for Being a Subject of Rights: the Natural Law Position," in: Keown, John & George, Robert P. (eds.), *Reason, Morality, and Law* (Oxford: Oxford University Press, 2013), pp. 236-248.
- Levy, Neil, "The Value of Consciousness," *Journal of Consciousness Studies* 21 (2014), pp. 127-138.
- Liao, S. Matthew, "The Organism View Defended," *The Monist* 89 (2006), pp. 334-350.
- \_\_\_\_\_, "The Basis of Human Moral Status," *Journal of Moral Philosophy* 7 (2010), pp. 159-179.
- \_\_\_\_\_, "The Basis of Human Moral Status," in: Brooks, Thom (ed.), *Ethics and Moral Philosophy* (Leiden: Brill, 2011), pp. 335-356.
- \_\_\_\_\_, *The Right to Be Loved* (Oxford/New York: Oxford University Press, 2015).
- Lippert-Rasmussen, Kasper, "Why Killing Some People Is More Seriously Wrong than Killing Others," *Ethics* 117 (2007), pp. 716-738.

- Locke, John, *An Essay Concerning Human Understanding* (1690; reprint, London: T. Tegg and Son, 1836).
- Lockwood, Michael, "Warnock versus Powell (and Harradine): When does potentiality count?," *Bioethics* 2 (1988), pp. 187-213.
- MacClellan, Joel P., *Minding Nature: A Defense of a Sentio-centric Approach to Environmental Ethics*, PhD dissertation (University of Tennessee, 2012), URL = <[http://trace.tennessee.edu/utk\\_graddiss/1433](http://trace.tennessee.edu/utk_graddiss/1433)> [accessed on February 7, 2016].
- Machin, K. L., "Amphibian pain and analgesia," *Journal of Zoo and Wildlife Medicine* 30 (1999), pp. 2-10.
- Marquis, Don, "Why Abortion Is Immoral," *Journal of Philosophy* 86 (1989), pp. 183-202.
- Mather, J. A., "Philosophical background of attitudes toward and treatment of invertebrates," *ILAR Journal* 52 (2011), pp. 205-212.
- McCloskey, H. J., "An examination of restricted utilitarianism," *Philosophical Review* 66 (1957), pp. 466-485.
- \_\_\_\_\_, "A Non-Utilitarian Approach to Punishment," *Inquiry* 8 (1965), pp. 249-263.
- McMahan, Jeff, *The Ethics of Killing* (Oxford/New York: Oxford University Press, 2002).
- \_\_\_\_\_, "Our Fellow Creatures," *Journal of Ethics* 9 (2005), pp. 353-380.
- \_\_\_\_\_, "Infanticide," *Utilitas* 19 (2007), pp. 1-29.
- \_\_\_\_\_, "Challenges to Human Equality," *Journal of Ethics* 12 (2008), pp. 81-104.

- \_\_\_\_\_, "Moral Intuition," in: LaFollette, Hugh & Persson, Ingmar (eds.), *The Blackwell Guide to Ethical Theory*, Second Edition (Chichester: John Wiley & Sons, 2013a), pp. 103-120.
- \_\_\_\_\_, "Killing and disabling: a comment on Sinnott-Armstrong and Miller," *Journal of Medical Ethics* 39 (2013b), pp. 10 f.
- \_\_\_\_\_, "Killing Embryos for Stem Cell Research," in: Kuhse, Helga, Schüklenk, Udo & Singer, Peter (eds.), *Bioethics: An Anthology* (Chichester: John Wiley & Sons, 2015), pp. 508-520.
- Mill, John Stuart, *Utilitarianism* (1861; reprint, Indianapolis/Cambridge: Hackett Publishing Company, 2001).
- Moore, G. E., *Principia Ethica* (Cambridge: Cambridge University Press, 1903).
- \_\_\_\_\_, *Ethics* (London: Williams and Norgate, 1912).
- Mosley, Craig, "Pain and Nociception in Reptiles," *Exotic Animal Practice* 14 (2011), pp. 45-60.
- Mulgan, Tim, "How Satisficers Get Away with Murder," *International Journal of Philosophical Studies* 9 (2001), pp. 41-46.
- \_\_\_\_\_, "The Ethics of Killing: Problems at the Margins of Life by Jeff McMahan," *Canadian Journal of Philosophy* 34 (2004), pp. 443-459.
- Nagel, Thomas, "Death," *Noûs* 4 (1970), pp. 73-80.
- \_\_\_\_\_, "What is it like to be a bat?," *The Philosophical Review* 83 (1974), pp. 435-450.
- \_\_\_\_\_, *The View from Nowhere* (Oxford/New York: Oxford University Press, 1986).

- Nicholls, J. G. & Baylor, D. A., "Specific modalities and receptive fields of sensory neurons in the CNS of the leech," *Journal of Neurophysiology* 31 (1968), pp. 740-756.
- Nozick, Robert, *Anarchy, State, and Utopia* (New York: Basic Books, 1974).
- \_\_\_\_\_, *Philosophical Explanations* (Cambridge: Harvard University Press, 1981).
- \_\_\_\_\_, "About Mammals and People," *The New York Times Book Review* 88 (November 27, 1983), pp. 11, 29 f.
- \_\_\_\_\_, *The Examined Life: Philosophical Meditations* (New York: Simon & Schuster, 1989).
- Nuffield Council on Bioethics, *The ethics of research involving animals* (London: Nuffield Council on Bioethics, 2005).
- O'Leary, Maureen A. et al., "The Placental Mammal Ancestor and the Post-K-Pg Radiation of Placentals," *Science* 339 (2013), pp. 662-667.
- Olson, Eric T., *The Human Animal: Personal Identity Without Psychology* (Oxford/New York: Oxford University Press, 1997).
- Overvold, M. C., "Self-Interest and the Concept of Self-Sacrifice," *Canadian Journal of Philosophy* 10 (1980), pp. 105-118.
- Parfit, Derek, "Personal Identity," *Philosophical Review* 80 (1971), pp. 3-27.
- \_\_\_\_\_, *Reasons and Persons* (Oxford: Oxford University Press, 1984).
- Pettit, Philip, "The Consequentialist Perspective," in: Baron, Marcia, Pettit, Philip & Slote, Michael (eds.), *Three Methods of Ethics* (Oxford: Blackwell, 1997), pp. 92-174.



- Quinn, Warren, "Abortion: Identity and Loss," *Philosophy and Public Affairs* 13 (1984), pp. 24-54.
- Rachels, James, *The End of Life: Euthanasia and Morality* (Oxford/New York: Oxford University Press, 1986).
- \_\_\_\_\_, *Created from Animals: The Moral Implications of Darwinism* (Oxford/New York: Oxford University Press, 1990).
- \_\_\_\_\_, *The Legacy of Socrates: Essays in Moral Philosophy* (New York: Columbia University Press, 2007).
- Rawls, John, *A Theory of Justice* (Oxford: Clarendon Press, 1972).
- Ray, A. Chadwick, "Humanity, Personhood and Abortion," *International Philosophical Quarterly* 25 (1985), pp. 233-245.
- Regan, Tom, *The Case for Animal Rights* (Berkeley/Los Angeles: University of California Press, 1983).
- \_\_\_\_\_, *The Case for Animal Rights*, Second Edition (Berkeley/Los Angeles: University of California Press, 2004).
- Reichlin, Massimo, "The Argument from Potential: A Reappraisal," *Bioethics* 11 (1997), pp. 1-23.
- Robertson, Teresa & Atkins, Philip, "Essential vs. Accidental Properties," in: Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* (Winter 2013 Edition), URL = <<http://plato.stanford.edu/archives/win2013/entries/essential-accidental/>> [accessed on October 22, 2015].
- Rolston, Holmes, III, "Human Uniqueness and Human Dignity: Persons in Nature and the Nature of Persons," in: The President's Council on Bioethics (ed.), *Human*

*Dignity and Bioethics* (Washington, D.C.: U.S. Government Printing Office, 2008), pp. 129-153.

Rose, Margaret & Adams, David, "Evidence for Pain and Suffering in Other Animals," in: Langley, Gill (ed.), *Animal Experimentation: The Consensus Changes* (New York: Chapman and Hall, 1989), pp. 42-71.

Rosenblatt, Roger, "Defenders of the Faith," *Time* (November 12, 1984), p. 112.

Ross, W. D., *The Right and the Good* (Oxford: Oxford University Press, 1930).

Sapontzis, S. F., *Morals, Reason, and Animals* (Philadelphia: Temple University Press, 1987).

Sarjeant, Richard, *The Spectrum of Pain* (London: Hart Davis, 1969).

Scanlon, Thomas Michael, *What We Owe to Each Other* (Cambridge: Cambridge University Press, 1995).

Scheffler, S., *The Rejection of Consequentialism* (Oxford: Clarendon Press, 1982).

Searle, John R., *Consciousness and Language* (Cambridge: Cambridge University Press, 1998).

Sher, George, *Equality for Inegalitarians* (Cambridge University Press, 2014).

Shewmon, D. Alan, "Recovery from 'Brain Death': A Neurologist's Apologia," *Linacre Quarterly* 64 (1997), pp. 30-96.

Shoemaker, Sydney, *Self-Knowledge and Self-Identity* (Ithaca: Cornell University Press, 1963).

\_\_\_\_\_, "Persons and Their Pasts," *American Philosophical Quarterly* 7 (1970), pp. 269-285.

- \_\_\_\_\_, "On David Chalmers's The Conscious Mind," *Philosophy and Phenomenological Research* 59 (1999), pp. 439-444.
- Shultziner, Doron & Carmi, Guy E., "Human Dignity in National Constitutions: Functions, Promises and Dangers," *The American Journal of Comparative Law* 62 (2014), pp. 461-490.
- Sidgwick, Henry, *The Methods of Ethics* (1874; reprint, Indianapolis/Cambridge: Hackett Publishing, 1981).
- Siewert, Charles, "Consciousness and Intentionality", Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* (Fall 2011 Edition), URL = <http://plato.stanford.edu/archives/fall2011/entries/consciousness-intentionality/> [accessed on February 7, 2016].
- Singer, Peter & Kuhse, Helga, "Individuals, Humans, and Persons: The Issue of Moral Status," in: Singer, Peter, Kuhse, Helga, Buckle, Stephen, Dawson, Karen & Kasimba, Pascal (eds.), *Embryo Experimentation: Ethical, Legal and Social Issues* (Cambridge: Cambridge University Press, 1990), pp. 65-75.
- Singer, Peter, "Famine, Affluence and Morality," *Philosophy and Public Affairs* 1 (1972), pp. 229-243.
- \_\_\_\_\_, *Animal Liberation* (New York: New York Review/Random House, 1975).
- \_\_\_\_\_, *Practical Ethics* (Cambridge: Cambridge University Press, 1979a).
- \_\_\_\_\_, "Killing humans and killing animals," *Inquiry* 22 (1979b), pp. 145-156.
- \_\_\_\_\_, *Practical Ethics*, Third Edition (Cambridge: Cambridge University Press, 2011).
- \_\_\_\_\_, "Animal Liberation at 30," in: Holland, Stephen (ed.), *Arguing about Bioethics* (Abingdon/New York: Routledge, 2012), pp. 185-194.

- Slote, Michael, "Satisficing Consequentialism," *Proceedings of the Aristotelian Society* 58 (1984), pp.139-163.
- Smart, J. J. C., "An outline of a system of utilitarian ethics," in: Smart, J. C. C. & Williams, Bernard, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973), pp. 1-74.
- \_\_\_\_\_, "Extreme and Restricted Utilitarianism," in: Foot, Philippa (ed.), *Theories of Ethics* (Oxford: Oxford University Press, 1967), pp. 171-183.
- Smith, A. D., *Husserl and the Cartesian Meditations* (London: Routledge, 2003).
- Smith, E. S. & Lewin, G. R., "Nociceptors: A phylogenetic view," *Journal of Comparative Physiology A* 195 (2009), pp. 1089-1106.
- Smith, J. A., "A Question of Pain in Invertebrates," *ILAR Journal* 33 (1991), pp. 25-31.
- Sneddon, L. U., Braithwaite, V. A. & Gentle, M. J., "Do fish have nociceptors: evidence for the evolution of a vertebrate sensory system," *Proceedings of the Royal Society B* 270 (2003), pp. 1115-1121.
- Soto, Carlos, "Killing, wrongness, and equality," *Philosophical Studies* 164 (2013), pp. 543-559.
- Sprigge, Timothy L. S., "A utilitarian reply to Dr. McCloskey," *Inquiry* 8 (1965), pp. 264-291.
- Stone, Jim, "Why Potentiality Matters," *Canadian Journal of Philosophy* 17 (1987), pp. 815-830.
- Strawson, Peter Frederick, *Individuals* (London: Methuen, 1959).
- Stretton, Dean, "Essential properties and the right to life: A response to Lee," *Bioethics* 18 (2004), pp. 264-282.

- Strong, Carson, "Preembryo Personhood: An Assessment of the President's Council Arguments," *Theoretical Medicine and Bioethics* 27 (2006), pp. 433-453.
- Sulmasy, Daniel P., "Dignity and Bioethics: History, Theory, and Selected Applications," in: The President's Council on Bioethics (ed.), *Human Dignity and Bioethics* (Washington, D.C.: U.S. Government Printing Office, 2008), pp. 469-501.
- Sumner, W. L., "A Matter of Life and Death," *Noûs* 10 (1976), pp. 145-171.
- \_\_\_\_\_, *Welfare, Happiness and Ethics* (Oxford: Clarendon Press, 1996).
- Sutherland, Stuart, *The Macmillan Dictionary of Psychology* (London: Macmillan, 1995).
- Sverdlik, Steven, *Motives and Rightness* (Oxford: Oxford University Press, 2011).
- The Cambridge Declaration on Consciousness* (Cambridge, July 7, 2012), URL = <http://fcmconference.org/img/CambridgeDeclarationOnConsciousness.pdf> [accessed on February 7, 2016].
- Thomas, Larry L., "Human Potentiality: Its Moral Relevance," *The Personalist* 59 (1978), pp. 266-272.
- Thomson, Judith Jarvis, "The Trolley Problem," *Yale Law Journal* 94 (1985), pp. 1395-1415.
- \_\_\_\_\_, *The Realm of Rights* (Cambridge: Harvard University Press, 1990).
- Tooley, Michael, "Abortion and Infanticide," *Philosophy and Public Affairs* 2 (1972), pp. 37-65.
- \_\_\_\_\_, *Abortion and Infanticide* (Oxford: Clarendon Press, 1983).
- Unger, Peter, *Identity, Consciousness, and Value* (Oxford/New York: Oxford University Press, 1990).

- United States National Library of Medicine, "Anencephaly," *Genetic Home Reference* (November 18, 2015), URL = <<http://ghr.nlm.nih.gov/condition/anencephaly>> [accessed on November 19, 2015].
- Vallentyne, Peter, "Against Maximizing Act Consequentialism," in: Dreier, James (ed.), *Contemporary Debates in Moral Theory* (Oxford: Blackwell Publishing, 2006), pp. 21-36.
- Van Gulick, Robert, "Consciousness," in: Zalta, Edward N. (ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2014 Edition), URL = <<http://plato.stanford.edu/archives/spr2014/entries/consciousness/>> [accessed on February 7, 2016].
- Van Inwagen, Peter, *Material Beings* (Ithaca: Cornell University Press, 1995).
- Velleman, J. David, "Love as a Moral Emotion," *Ethics* 109 (1999), pp. 338-374.
- Waldron, Jeremy, *God, Locke, and Equality* (Cambridge: Cambridge University Press, 2002).
- Warren, Mary Anne, *Moral Status: Obligations to Persons and Other Living Things* (Oxford: Clarendon Press, 1997).
- Wells, M. J., *Octopus* (London: Chapman and Hall, 1978).
- White, Alan R., *Rights* (Oxford: Clarendon Press, 1984).
- Williams, Bernard, "A critique of utilitarianism," in: Smart, J. C. C. & Williams, Bernard, *Utilitarianism: For and Against* (Cambridge: Cambridge University Press, 1973), pp. 75-150.
- Wong, Kate, "Meet the Last Common Ancestor of Bats, Whales, Sloths and Humans," *Scientific American* (February 7, 2013), URL = <

<http://blogs.scientificamerican.com/observations/meet-the-last-common-ancestor-of-bats-whales-sloths-and-humans/> [accessed on December 5, 2015].

Würbel, Hanno, "Ethology applied to animal ethics," *Applied Animal Behaviour Science* 118 (2009), pp. 118-127.

Yeats, William Butler, *Poems of William Butler Yeats* (Raleigh: Hayes Barton Press, 1962).

## Appendix: Tom Regan's philosophy of animal rights

I anticipate that some readers, after having read this dissertation, will wonder how the view I endorse differs from Tom Regan's theory of animal rights.<sup>380</sup> The book in which Regan developed that theory was very briefly mentioned at the beginning of the chapter on McMahan's account of the wrongness of killing. That is no coincidence, as Regan's theory is very similar to McMahan's, in both structure and content, and problematic for much the same reasons. For Regan, it is not persons who have a special moral status, but subjects-of-a-life. Animals are subjects-of-a-life

“if they have beliefs and desires; perception, memory, and a sense of the future, including their own future; an emotional life together with feelings of pleasure and pain; preference- and welfare-interests; the ability to initiate action in pursuit of their desires and goals; a psychophysical identity over time; and an individual welfare in the sense that their experiential life fares well or ill for them, logically independently of their utility for others and logically independently of their being the object of anyone else's interests.”<sup>381</sup>

Even though this definition is written as if the psychological properties mentioned were a matter of either-or, it is clear that most or all of them are a matter of degree, and occur gradually in the normal development of human beings. Newborns have an emotional life and an individual welfare, but, if at all, they only have a very vague sense of the future, and even less a sense of their own future. Human beings are born without at least some of

---

<sup>380</sup> Of course, Regan's theory of animal rights, still one of the most comprehensive, robust, and popular such theories, is also of independent interest.

<sup>381</sup> Regan 2004, p. 243.



the psychological properties constitutive of being a subject-of-a-life, and there is, in Mary Anne Warren's words, no "magic moment at which the missing abilities spring into being."<sup>382</sup> Regan, like McMahan, endorses a threshold account of full moral status, which requires line-drawing. In the preface to the second edition of *The Case for Animal Rights*, Regan explicitly acknowledges that his position implies a sharp line between subjects-of-a-life and animals who are not subjects-of-a-life, yet he refrains from trying to draw that line.<sup>383</sup> He only says that, wherever we draw the line, "mentally normal mammals of a year or more"<sup>384</sup> are above it. As human beings in the earliest stages of their existence are plainly below the line, there is a moment in the life of every normal human being, at which he or she, barring premature death, will cross it. In Regan's theory, this moment has great moral significance. Unlike animals who are not subjects-of-a-life, subjects-of-a-life have intrinsic value (Regan uses the term "inherent value"), and they have it equally. Subjects-of-a-life are valuable as ends in themselves, and not just as means, and they have a right to be treated in a way that respects their intrinsic value, which includes a right not to be killed. The moral status of subjects-of-a-life in Regan's theory is similar to the status of persons in McMahan's, and the moments in the normal development of a human being at which he or she becomes a subject-of-a-life and a person, respectively, are of comparable moral importance. Regan too implausibly bases a radical difference in treatment on a mere difference in degree, which requires line-drawing that is necessarily arbitrary. Furthermore, just like McMahan, Regan grounds our intrinsic value and equality in the fact that you and I possess certain accidental properties which we once lacked, and might lose in the future. This seems superficial, and does not do justice to the

---

<sup>382</sup> Warren 1997, p. 118.

<sup>383</sup> Cf. Regan 2004, p. xvi.

<sup>384</sup> Regan 2004, p. 78.

intuitive understanding of equality, according to which the empirical ground of our moral equality must be an essential fact about us. In contrast, on my view, everybody who has intrinsic value has that value through the entirety of his or her existence. Further, my proposal bases a radical difference in moral status on a radical difference in reality, and does not require any arbitrary line-drawing.

## **Rainer Ebert**

- 05/2016      **Ph.D.**, Philosophy  
Rice University, United States of America
- 12/2014      **M.A.**, Philosophy  
Rice University, United States of America
- 06/2009      **Diplom** ( $\approx$  M.Sc.), Physics  
Ruprecht-Karls-Universität Heidelberg, Germany
- 03/2006      **Vordiplom** ( $\approx$  B.Sc.), Physics  
Ruprecht-Karls-Universität Heidelberg, Germany
- 06/2004      **Abitur**  
Hariolf-Gymnasium Ellwangen, Germany
- Born**            11 June 1985  
Ellwangen (Jagst), Germany
- URL**            <http://www.rainerebert.com>