PERSONS AND PARTIALITY: LIMITATIONS ON CONSEQUENTIALIST JUSTIFICATIONS

by

Timothy William Christie

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

DOCTOR OF PHILOSOPHY

in

The Faculty of Graduate Studies
(Philosophy)

THE UNIVERSITY OF BRITISH COLUMBIA

December 2007

© Tim William Christie, 2007
ABSTRACT

Should the authorities observe the rules regarding the treatment of enemy combatants, or is it morally justified for the authorities to violate some human rights in order to make everyone safer? Some moral theorists are committed to the claim that using torture for the greater good is not only permissible but also obligatory. One of the key goals of my thesis is to undermine this sort of claim.

Contemporary consequentialists, such as Philip Pettit, hold that an agent is always permitted to bring about a certain state of affairs solely on the grounds that the state of affairs is the best state of affairs, impersonally judged. Derek Parfit agrees with Pettit’s claim, arguing that a reductionist account of persons offers support for moral theories that fail to acknowledge the fact that each person is a separate unit of moral concern. I reject Parfit’s assumption that the natural separateness of persons is morally insignificant: if we imagine a species of person that is not naturally separate from each other, it is reasonable to suppose that the moral norms of this different species of person would be drastically different from deeply entrenched human moral norms. I conjecture that the separateness of persons offers a rationale for restrictions against grossly assaulting and killing innocent persons.

Samuel Scheffler argues that restrictions are so strong they are paradoxical. I counter this charge by arguing that restrictions need not categorically bar types of actions like killing innocent people, but rather should limit consequentialist justifications for these types of actions. Such a distinction address the air of paradox that surrounds restrictions because it allows for the possibility that agent-relative reasons justify why agents may assault or kill when the agent is confronted with a tragic moral dilemma.

Agent-relative reasons are relevant to moral justification because human persons value the world around them from the first person point of view. In order for morality to appropriately acknowledge this feature of human persons, it must be permissible for humans to adopt a partial attitude toward their own actions, lives and loved ones.
# Table of Contents

Abstract ........................................................................................................... ii

Table of Contents ........................................................................................... iii

Acknowledgments ............................................................................................ v

Dedication .......................................................................................................... vi

1 Introduction .................................................................................................. 1
  1.1 An Introduction to the Project ................................................................. 1
  1.2 An Argument for Consequentialism ......................................................... 3
  1.3 Further Claims About Consequentialism .............................................. 5
    1.3.1 Consequentialism is a Moral Theory of the Right ......................... 7
    1.3.2 Consequentialism has an Objective Goal ....................................... 8
      1.3.2.1 Objective Consequentialism is Primarily a Theory of Moral 
          Evaluation ......................................................................................... 8
      1.3.2.2 Objective Consequentialism is a Moral Theory that Applies Both
          Directly and Indirectly to Actions ................................................... 10
      1.3.2.3 Objective Consequentialism is a Theory of Value Promotion ........ 11
      1.3.2.4 Moral Justifications are Objective ........................................... 12
  1.4 Challenges to Consequentialism in the Literature ................................ 12
  1.5 A Plausible Methodology ......................................................................... 22
  1.6 An Overview of My Argument .................................................................. 24

2 The Separateness of Persons is Morally Significant ................................... 28
  2.1 Introduction ............................................................................................... 28
  2.2 Rawls and Nozick on the Separateness of Persons .................................. 30
    2.2.1 Rawls .............................................................................................. 30
    2.2.2 Nozick ............................................................................................ 35
  2.3 Parfit’s Metaphysics .................................................................................. 38
  2.4 Parfit: A Reductionist View of Persons Supports Consequentialism .... 40
  2.5 Korsgaard and Johnston on Parfitian Impact ......................................... 43
  2.6 Natural Separateness and our Person Practices ...................................... 45
  2.7 Morality and Our Species’ Nature: The Mergers and the Collectives .... 51
  2.8 An Objection and a Reply ....................................................................... 58
  2.9 Conclusion ................................................................................................. 62

3 Consequentialism, Intuition and Restrictions ............................................. 63
  3.1 Introduction ............................................................................................... 63
  3.2 Parfit on Restrictions ............................................................................... 63
  3.3 The Argument From Intuition ................................................................... 65
  3.4 Pettit’s Consequentialist-Based Restrictions ......................................... 71
  3.5 A Commonsense Practice in Search of a Foundation ............................... 77
  3.6 Two Restrictions ....................................................................................... 81
    3.6.1 Killing Innocent People .................................................................. 81
3.6.2 Gross Bodily Assault ................................................................. 82
3.7 Are There Other Restrictions? .................................................... 83
3.8 Conclusion ................................................................................ 84

4 Answering the Charge of Paradox .................................................... 85
4.1 Introduction .............................................................................. 85
4.2 Restrictions and the Charge of Paradox ....................................... 86
4.3 Tragic Dilemmas and Obligation .................................................. 90
4.4 Nagel on Agent-relative Reasons and Restrictions ....................... 91
4.5 Two Accounts of Restrictions: Restrictions_A and Restrictions_B 94
4.6 Value Promotion and Theoretical Parsimony .............................. 99
4.7 Agent-relativity and the Intrinsic Moral Significance of Causal Pathways 104
4.8 Conclusion .............................................................................. 110

5 Agent-relativity, Human Partiality and Options ............................... 112
5.1 Introduction .............................................................................. 112
5.2 Options and Kagan's Pro Tonto Reason to Promote the Good .......... 113
5.3 Defending Options: The Appeal to Costs or the First Person Point of View? 122
5.4 Response 1: Eliminating Projects ................................................. 130
5.5 Response 2: Partiality Justified Impartially .................................... 132
5.6 Why Objective Consequentialism Fails to be a Human Morality ......... 136
5.7 Why Top-secret Objective Consequentialism Fails to be a Human Morality 139
5.8 Incorporating Partiality with Agent-neutral limits: Scheffler's Hybrid Solution 143
5.9 Incorporating Partiality With Agent-Relative Requirement ............... 147
5.10 Conclusion .............................................................................. 150

Bibliography .................................................................................. 152

Appendices ..................................................................................... 163

Appendix A: Prospects for Application: The War on Terror ............... 163
A.1 The Treatment of Suspected Enemy Combatants ........................... 163
A.2 The Beslan Massacre and the Moscow Theatre Hostage Situation .... 165
A.3 Options and the Order to Storm the Building ................................. 167

Appendix B: An Extended Enforce .................................................... 169
B.1 Introduction .............................................................................. 169
B.2 Darwinian Revolution .................................................................. 177
B.3 The New Sciences: Managing the Human Animal ......................... 180
B.4 A Popular Moral Theory ............................................................. 185
B.5 Popenoe's and Johnson's Consequentialist Arguments for Negative Eugenics 189
B.6 Restrictions on Consequentialist Justifications ............................ 192

iv
ACKNOWLEDGEMENTS

I want to thank all of the faculty, staff and graduate students in the Department of Philosophy who have all been of great assistance throughout my time at the University of British Columbia. I would like to take this opportunity to specifically thank Rana Ahmad, Vincent Bergeron, Ian Brooks, Dan Christie, Patrick Edwards, Jillian Isenberg, Josh Johnston, James Kelleher, Brian Laetz, Nola Semcyzsyn, Jie Tian, Max Weiss, and Jim Westin for all of the thought-provoking and helpful discussion. I benefited immensely from discussion and written comments from Dominic Lopes, Andrew Irvine, Don Brown, Paul Russell and Cory Fairley. From the start of this project to its completion, both Zohar Geva and Selman Halabi have been a great source of encouragement, discussion and criticism. I would also like to thank all of my editors, including Kari Coleman, Lynda Christie and Heather Christie. I am very grateful for the two years of funding I received from the Social Science and Humanity Research Council.

Most importantly, I would like to acknowledge the members of my Dissertation Committee—John Beatty, Catherine Wilson and Scott Anderson—for their advice, guidance and insight.
To Heather Christie
1 INTRODUCTION

1.1 An Introduction to the Project

Moral theory is supposed to help organize our thinking about morality in general, but especially so in complex cases. Consider the recent increase in violent terrorist activity in many countries. Should the authorities observe the rules regarding the treatment of imprisoned enemy combatants outlined in the Geneva Convention, or is it morally justified for the authorities to secretly violate some human rights in order to make everyone safer? Some moral theorists are committed to the claim that using torture for the greater good is not only permissible but also obligatory. One of the key goals of my thesis is to undermine this sort of claim.

As with most projects in philosophy, the impetus behind this current research project started with a puzzle. A moral theory is an attempt to offer a systematic analysis of normative terms such as ‘right’, ‘wrong’ and ‘morally justified’. I simultaneously grant that the following two moral theses have plausibility. The first thesis is that moral permission and obligation are about making the world as good as possible. Actions can be morally justified by virtue of the overall good the actions bring about (the Consequentialist Thesis). The second thesis is that there are some kinds of actions, for example hurting or killing innocent people, which cannot be morally justified on the grounds of their contribution to the overall impersonal good (the Restriction Thesis). Conflicts arise because both theses make claims about the nature of moral justification.
Several recent books have shown the force of the problems associated with incorporating both theses into a coherent picture of morality.¹

I will argue that the Consequentialist Thesis needs to be supplemented by the Restriction Thesis. First, the Consequentialist Thesis is too broad: it includes implausible moral justifications for grossly assaulting or killing innocent persons, on the grounds that the assaults or killings would, from an impersonal or impartial perspective, bring about the best state of affairs. Second, the Consequentialist Thesis is too narrow: it rules out plausible moral justifications that stem from our natural partiality. These two errors arise because the Consequentialist Thesis fails to appropriately acknowledge morally significant features of human nature. More specifically, the Consequentialist Thesis fails to appropriately acknowledge facts about human persons that limit what counts as moral justifications (i.e., the separateness of human persons) and facts about human persons that provide moral justifications (i.e., human partiality).

Supplementing the Consequentialist Thesis with the Restriction Thesis is a way to emphasize the intrinsic moral importance of actions. A supporter of the Consequentialist Thesis holds that actions are of mere instrumental value: an action is justified if and only if that action brings about the best state of affairs. What is wrong with this instrumental picture of moral justification? On such an account of moral justification actions are simply one means among many that produce the best state of affairs. In light of the separate and natural partiality of humans, our actions have an intrinsic moral importance. The Consequentialist Thesis fails to adequately acknowledge this intrinsic importance.

1.2 An Argument for Consequentialism

The goal of this first chapter is to clearly identify what account of consequentialist moral justification my thesis will reject. The “Consequentialist Thesis” is a stand-in for a very complex and much discussed moral theory called “consequentialism”. Consequentialism is a theory for judging moral permissibility and impermissibility. Consequentialists hold that actions should be judged morally based on what the action brings about in the world. How do the consequentialists (and their critics) explain this moral theory? Scheffler, in his Introduction to Consequentialism and its Critics, clearly explains this theory:

Consequentialism in its purest and simplest form is a moral doctrine which says that the right act in any given situation is the one that will produce the best overall outcome, as judged from an impersonal standpoint which gives equal weight to the interests of everyone.²

Philip Pettit holds that some surprising implications follow from the claim that right action ought to be defined in terms of producing the best states of affairs, impersonally judged:

Consequentialism is the view that whatever values an individual or institutional agent adopts, the proper response to those values is to promote them. The agent should honour those values only in so far as honouring them is a part of promoting them, or is necessary in order to promote them…. Consequentialists see the relation between values and agents as an instrumental one: agents are required to produce whatever actions have the property of promoting a designated value, even actions that fail intuitively to honour it.³

Consequentialism holds that an agent ought to do whatever it takes to promote value, even when this includes the agent failing to personally act in a manner consistent with what the agent values. For instance, consider the case of honesty. While usually the way


an agent may promote the value of honesty is to not lie, sometimes telling a lie will bring about a state of affairs with fewer lies overall. Assuming that honest is an intrinsically good thing, the consequentialist says that the agent *ought to lie* in such a case.

What I will highlight throughout this thesis is that both Scheffler and Pettit agree that an agent is *always permitted* to bring about certain states of affairs solely on the grounds that they are the best states of affairs, impersonally judged (AP). Before I turn to a focussed examination of AP, I will continue with my overview of consequentialism more generally.

In contemporary debates, it is not common for proponents of consequentialism to offer arguments that advance their theory. It is far more common to find criticisms of consequentialism, and then consequentialists’ responses. My conjecture for why it is that positive arguments that advance consequentialism are not common in the current literature is that the argument is deemed to be so obvious or so basic that it is not really worth stating explicitly.

Consider what I will label The Basic Argument, an argument that purports to rationally support consequentialism:

1. States of affairs are the bearers of impersonal moral value. (Value Premise)

2. Bringing about the best state of affairs, impersonally judged, is always definitively morally preferable compared to bringing about sub-optimal states of affairs. (Promoting Value Premise)

3. Agents are capable of bringing about many different states of affairs. (Freedom Premise)

4. Therefore, it is definitively morally preferable that agents always bring about the state of affairs available to them that is impersonally judged to be best.

---

4 From here on I will refer to this claim as the "always permitted" claim, or AP.
To extend this argument's conclusion to explicitly offer agents a moral prescription, I can add:

5. Agents are morally justified when they produce states of affairs that are definitively morally preferable; agents are not morally justified when they produce states of affairs that are not morally preferable.

6. Agents ought to do acts that are morally justified; agents ought not do acts that are not morally justified.

7. Therefore, agents ought to bring about a state of affairs if and only if doing so produces the impersonally judged best state of affairs available.

On this argument for consequentialism, my goal is to challenge the rational acceptability of premise 2. While it is obviously true that we value the production of impersonally judged good states of affairs, I aim to reject the claim that this one goal is always definitively morally preferable. However, because this premise has some prima facie plausibility, I acknowledge the burden of proof to render premise 2 rationally unacceptable. It is the goal of my thesis to accomplish this task.

1.3 Further Claims About Consequentialism

There are many versions of consequentialism. The one I will address is prominent in the literature. Both Pettit (a defender of consequentialism) and Scheffler (a critic) hold that (1) Consequentialism is best understood as a theory of the right, and (2) Consequentialism is best understood to have an objective goal, namely, to produce the best state of affairs, impersonally judged. Pettit holds that (2) ought to be understood to entail the following:

(2.1) Consequentialism is primarily a theory of moral evaluation;

(2.2) Consequentialism is a moral theory that applies both directly and indirectly to actions and institutions;
(2.3) Consequentialism is a theory of value promotion;
(2.4) Moral justifications are objective.

Pettit’s account of consequentialism shares the salient features of the version of consequentialism I aim to reject. In the Introduction to Pettit’s edited text, *Consequentialism*, he writes:

I said earlier that I would assume henceforth as the standard line has it, that consequentialism is a theory of the right, not a theory of the good. I will assume, equally, that it is a theory of evaluation, not of moral deliberation; and that it is a theory for assessing all options that can face an agent or agency, not just a theory for assessing abstract rules of choice.5

Pettit clearly agrees that the “standard line” version of consequentialism includes (1) and (2), where (2) is reasonably interpreted to entail (2.1), (2.2), (2.3) and (2.4).6 Critics of consequentialism, such as Samuel Scheffler, take Pettit’s account of consequentialism seriously. Scheffler aims to reject a version of consequentialism that is consistent with (1) and (2).7

I think it is plausible that a proponent of consequentialism in whatever form will be interested in my arguments, even if he or she disagrees with some of the features of the version of consequentialism I aim to address. Perhaps it is in virtue of one or two of these features of consequentialism that the proponent of some of version will claim my

---


7 Where (2) entails (2.1), (2.3) and (2.4). While Scheffler thinks his arguments could be made to address a version of consequentialism that also includes (2.2), he does not explicitly develop this argument. See: Scheffler, *The Rejection of Consequentialism*, 2.
arguments fail, but showing this entailment would be interesting in itself. I will now turn to a brief overview of each of (1) and (2).

1.3.1 **Consequentialism is a Moral Theory of the Right**

Consequentialism, on its own, specifies no account of the good. On its own, consequentialism explicates the right, namely, the morally permissible and the morally impermissible. Drawing on Scheffler’s language, consequentialism needs to be coupled with a “ranking principle” in order for it to be a complete moral theory. A ranking principle is a theory of the good that offers a framework for judging which state of affairs is best. For example, hedonism is one such ranking principle: states of affairs are ranked from best to worse based on how much pleasure they contain. Another more contemporary example is an objective welfare ranking principle, where states of affairs are ranked from best to worse based on how much human flourishing they contain.

One of the corollaries of incorporating a ranking principle into one’s moral theory is that a state of affairs must occupy one position on the overall scale of impersonal value. What is essential for any ranking principle is that it is goal-based. A goal of a ranking principle may be unitary or pluralistic; I will not debate this complex issue here. As Sumner argues, if the goal is a pluralistic one then the ranking principle must include some system of trade-offs among the plurality of different goods in order for the ranking principle to be able to perform its function and rank states of affairs. Aside from these structural features of ranking principles, it is not necessary to couple consequentialism

---


with any *specific* ranking principle. For my purposes, it will be enough to define the best state of affairs, impersonally judged, as being the state of affairs that contains the most goodness (as defined by one’s favoured ranking principle, whatever that may be).

1.3.2 **Consequentialism has an Objective Goal**

I will assume that the goal of consequentialism is to actually produce the best state of affairs, impersonally judged. A moral theory that employs an impersonal ranking principle purports to be an objective moral theory. States of affairs attempt to capture the way the world is, not the way the world seems to a particular agent, or to a particular group of agents. To put this claim in Thomas Nagel’s language, impersonal judgements about the best state of affairs are supposed to present agents with agent-neutral reasons for action.¹⁰ Consequentialism aims to give people access to, for example, the good life, not people who merely *imagine* they have access to the good life.¹¹ Based on this objective account of the goal of consequentialism, several other debates in the consequentialist literature are decided. I will now outline these aspects of objective consequentialism (2.1)-(2.4).

1.3.2.1 **Objective Consequentialism is Primarily a Theory of Moral Evaluation**

Historically, consequentialism has been seen as a theory of both moral evaluation and moral decision-making. In fact, a common attack against some versions of consequentialism is that it would be difficult in everyday life to make decisions on its

---


basis, or that if people tried to make decisions on its basis they would have calculating and morally unappealing characters. Yet, according to Peter Railton, there is no reason to think consequentialism should be thought of as a theory of moral decision-making. Railton holds that consequentialism should be seen as a criterion of the right. An objective version of consequentialism “is the [theory] that the criterion of the rightness of an act or course of action is whether it in fact would most promote the good of those acts available to the agent.” Following Railton’s lead, consequentialism should be seen as an “objective” moral theory.

Pettit agrees with Railton on the objectivity of consequentialism. Pettit takes consequentialism to be a theory of moral evaluation and not a theory of moral decision-making. In order to support his point, Pettit offers the example of “calculatively elusive” goods, such as spontaneous action. As soon as one calculates that an act would be appropriately spontaneous, the act ceases to be spontaneous. Thus, in order to actually attain the best state of affairs (which, we are assuming, includes some spontaneous action) we must not always in the moment aim at this goal because doing so would actually make our goal unattainable. The solution to the problem of calculatively elusive goods such as spontaneity is to deny that consequentialism is a rubric for moral decision-making in favour of it being a theory of moral evaluation.


14 Ibid., 113.

1.3.2.2 Objective Consequentialism is a Moral Theory that Applies Both Directly and Indirectly to Actions and Social Institutions

In some recent debates, people have thought consequentialists must choose: either consequentialism should apply directly to specific actions ("act-consequentialism" or "extreme-consequentialism") or it should apply only directly to systems of moral rules, motives or character traits that people ought to have. According to a proponent of an objective account of consequentialism, the choice is perceived as ill-formed because having to choose between these two options would obfuscate more than it would clarify. Objective consequentialists advocate that we ought to apply consequentialism in the way that actually works the best. In a context when it is rational to hold that applying consequentialism to rules will work the best we ought to do so; when it is optimal in some specific cases to break the rules we ought to do that, too. According to the objective consequentialist, whatever will in fact produce the best state of affairs, impersonally judged, is what we ought to do.

If consequentialism directly applies only to our choice of, for example, which system of motives people ought to have, then an explanation is needed for why the theory is limited in such a way. What is the principled rationale for limiting the production of the best state of affairs? Pettit, for instance, goes so far as to wonder whether or not philosophers who hold such views are really committed to consequentialism. The lesson I take from Pettit is the following: if one promotes consequentialism in one sphere (say, in judging the moral justification for a set of rules, or in judging the moral justification

---

for a set of motives, etc.), it is a *prima facie* mystery why this consequentialism is not more comprehensive.\(^{17}\) So long as the good would thereby actually be brought about, why cannot agents produce the most good available to them in specific cases? If the goal of morality is to produce the best states of affairs, then limiting the attainment of the best state of affairs is at minimum mysterious, and at worse contradictory.

### 1.3.2.3 Objective Consequentialism is a Theory of Value Promotion

Pettit holds that a formulation of consequentialism that emphasizes the promotion of value is the most plausible.\(^{18}\) While there are accounts of consequentialism that deny that it is best interpreted as a theory of value promotion,\(^ {19}\) I think such versions are far less intuitive and more difficult to defend. The Basic Argument would lose its simplicity and elegance. If producing good is the basis for right action, and more good is better than less, how could it *not* be the case that the most good is best of all? A consequentialist who offers a non-value promoting form of consequentialism has some tough questions to answer: in what way can The Basic Argument be repaired if the value promotion premise is dropped? Simplicity, elegance and intuitive plausibility are left by the wayside if the value promotion premise of the basic argument is abandoned. I will follow Pettit’s lead and take consequentialism to be a moral theory that emphasizes value promotion.

---


18 Philip Pettit, "Consequentialism."

1.3.2.4 **Moral Justifications are Objective**

A subjective account of moral justification is one that holds that an action is justified (if it is justified) only by what the agent claims justifies it. In other words, a moral justification for an action depends on *what the agent thinks*. An objective account of moral justification holds that actions are morally justified or not *independently* of what the agent thinks it is that offers his or her action moral justification. Scheffler and Pettit agree with an objective account of moral justification. Thus, for example, even if a particular agent could not produce a moral justification for her doing act P in circumstance C, P is either morally justified or not, independently of this agent’s failure to produce an appropriate justification. Furthermore, even if an agent held that act P in circumstance C is morally justified by consideration A—and the agent is wrong because act P is really justified by B—an objective account of moral justification would hold that the agent is morally justified in doing act P in circumstance C.

In summation, when I refer to consequentialism, I refer to a moral theory that adopts (1) and (2), where (2) is understood to include (2.1), (2.2), (2.3), and (2.4).

1.4 **Challenges to Consequentialism in the Literature**

Three major challenges to consequentialism predominate critical examinations of this moral theory over the past 40 years or so. First, Philippa Foot challenges consequentialists’ vague talk about “the best states of affairs”. Second, Thomas Nagel critiques consequentialism on the grounds that it assumes the only reasons a person could have to do anything are agent-neutral reasons. Lastly, Bernard Williams argues against

---

20 An objective account of moral justification might take what the agent thinks justifies the action to be relevant for determining what action the agent performed.
consequentialism on the grounds that it represents an attack on integrity because it relies on an impoverished view of action and moral responsibility. I will offer a brief overview of each criticism and explain how each relates to my project.

Philippa Foot, in her essay “Utilitarianism and the Virtues”, develops a telling critique of some of the key assumptions that utilitarians—and consequentialists more broadly—accept, and often accept without critical reflection. While Foot and I both disagree with consequentialist justifications for actions such as torturing innocent people, the nature of her criticism challenges Premise 1 of The Basic Argument (whereas I aim to reject Premise 2). She writes,

Suppose, for instance, that some evil person threatens to kill or torture a number of victims unless we kill or torture one, and suppose we have every reason to believe that he will do as he says. Then in terms of their total outcomes (again consisting of the states of affairs made up of an action and its consequences) we have the choice between more killings or torturings and less, and a consequentialist will have to say that we are justified in killing or torturing the one person, and indeed that we are morally obliged to do it, always supposing that no indirect consequences have tipped the balance of good and evil. There will in fact be nothing that it will not be right to do to a perfectly innocent individual if that is the only way of preventing another agent from doing more things of the same kind.

Now I find this a totally unacceptable conclusion and note that it is not a conclusion of utilitarianism in particular but rather of consequentialism in any form.\(^{21}\)

Foot claims that consequentialism must have gone wrong somewhere, and she thinks the error is hidden in the consequentialist’s assumption that “there are better and worse states of affairs in the sense that consequentialism requires.”\(^{22}\) Foot challenges whether it


\(^{22}\) Ibid., 410.
makes sense to claim that states of affairs are bearers of impersonal moral value. That is
to say, The Basic Argument is flawed because Premise 1 is unintelligible.

Foot denies that we can make sense of claim that states of affairs are the bearers of
impersonal value. Take a simple example: the invention of a new type of car alarm both
assists and hinders the attainment of “the best state of affairs”—depending on whether we
ask the cops or the robbers. How can we understand states of affairs from an
impersonal, or agent-neutral, point of view? That is to ask: how can we make sense of a
point of view from no particular point of view? These questions are perplexing; yet, the
consequentialist insists that the judgement about states of affairs be made from no
particular point of view (not the cops’ or the robbers’). Perhaps the consequentialist
would clarify their position by asking the following: all things considered, is the car
alarm an invention that helps or hinders the attainment of the best state of affairs,
impersonally judged? Yet, even at this meta-level, the question still is not clear: do we
let the capitalists or the Marxists make the all things considered judgement?

As a response to the last rhetorical question, perhaps we should consider the
consequences of the invention of this new car alarm from “the moral point of view”.
What is the moral point of view? It is the view that considers the interests and welfare of
each equally, the consequentialist replies. Yet, it appears we are in a circle: we are
employing the idea of a moral point of view to make sense of how states of affairs can be
the bearers of impersonal value and thus get consequentialism “off the ground”. This
strategy is circular because the consequentialists are making use of a moral point of view
in order to offer an account of morality.

---

Ibid., 411.
In order that consequentialism “gets off the ground”, it needs to be assumed by proponents and opponents alike that at least some sense can be given to the claim that state of affairs can have impersonal value. Furthermore, if states of affairs can be the bearers of impersonal value, then some state of affairs among the possible ones will be best. While I do not have a principled reply to Foot, I think one observation is worth emphasizing. We do talk about what would be better and worse, morally speaking, overall. Some sense is given to these expressions. We use them in communication and people seem to understand what we mean when we say things such as: “the world would have been a better place, over all, had the tidal wave on Boxing Day 2004 never occurred”, or “the world will be made a better place, overall, on the day someone discovers a cure for childhood leukemia”. If we can make sense of “better” and “worse” states of affairs overall, then it might also be the case that we can make sense of the best state of affairs, overall. I admit that this entailment relation is a weak one, but in order for consequentialism to make any sense as a moral theory, we must agree we can make sense of “the best state of affairs, impersonally judged”. I grant the consequentialists the plausibility of holding that we can make sense of “the best state of affairs, impersonally judged”.

My project does share something in common with Foot’s. Foot’s motivation behind her criticism—her revulsion with the consequentialist claim that any action at all is potentially morally justifiable—is foundational for my current project. I agree with Foot

24 Ibid., 410; Philippa Foot appears to agree with this point.

25 This leap, from talk about “better” and “worse” states of affairs to “the best state of affairs” is what a “ranking principle” purports to permit (where the best state of affairs is the state of affairs that contains the most good as identified by a ranking principle).
that treating persons as mere means, even in pursuit of the overall good, is inconsistent with treating persons as each deserves. One of the goals of my thesis is to argue that it is inconsistent with the morally significant nature of human persons that our bodies and lives may be treated in such an instrumental manner.

Nagel’s criticism of consequentialism, while similar to Foot’s, focuses not on the vagueness inherent in “best states of affairs, impersonally judged”, but instead on the consequentialist’s claims regarding how the agent ought to respond to “the best states of affairs, impersonally judged”. Nagel challenges the consequentialist’s assumption in favour of transcendental impartiality. Consequentialists assume that all values are based solely on impersonal values. Nagel disagrees: “we can no more assume that all values are impersonal than that all reality is physical.” Nagel argues that a tension exists between objective reasons and subjective inclinations. He writes,

[The correct view of morality] will be arrived at by the exploration of this conflict rather than by the automatic victory of the most transcendent standpoint. In the conduct of life, of all the places, the rivalry between the view from within and the view from without must be taken seriously.

Nagel argues that the impersonal is merely a component of morality—albeit an essential component. Nagel agrees with the consequentialist that the pain and pleasures of others present agents with objective reasons for bringing about various states of affairs. Nagel writes: “Life is filled with basic pleasures and pains, and they matter”. Yet, morality—and ethics more broadly—is not only concerned with what states of affairs come to pass,


27 Ibid., 163.

28 Ibid.,164-165.
but also with what *actions* people should or may do. There are agent-relative reasons an agent may have for doing or not doing a certain thing.

Nagel considers three sorts of reasons: reasons of autonomy, deontological reasons, and special obligations. The consequentialist aims to reduce agent-relative reasons to agent-neutral reasons. Of special significance for my project is the inherent tension in attempting to reduce deontological values to agent-neutral values. Nagel thinks that *something* is missing in this attempted reduction of agent-relative reasons like deontological values to agent-neutral reasons; this reduction is not "the whole truth." Deontological constraints are agent-relative reasons and, "if they exist, they restrict what we may do in the service of either relative or neutral goals." Nagel claims that it is common to include many types of deontological constraints, including special obligations that stem from promises, various prohibitions against killing, robbing or torturing others, and perhaps deontological requirements of fairness. These reasons resist being understood in terms of agent-neutral values because "deontological reasons have their full force against your doing something—not just against its happening." Yet, Nagel does admit that these reasons are "formally puzzling": the constraint against murder purports to "prohibit murder even if it is necessary to prevent other murders." While the consequentialist might argue that the paradoxical nature of deontological restrictions

29 Ibid., 165.
30 Ibid., 166.
31 Ibid., 175.
32 Ibid., 176.
33 Ibid., 177.
34 Ibid., 179.
is exactly what their reductionist account would avoid, Nagel thinks that this reduction
omits the moral phenomenology of the case. If the consequentialists are correct, then
deontological constraints are illusions. Yet, Nagel continues:

> we seem to apprehend in each individual case an extremely powerful agent-relative reason not to harm an innocent person. This presents itself as the apprehension of the normative truth, not just the psychological inhibition. It needs to be analyzed and accounted for, and accepted or rejected according to whether the account gives it an adequate justification.\(^{35}\)

I think that Nagel is correct that agent-relative values do not neatly reduce into agent-neutral values without remainder; I also think he is correct that, commonly understood, such restrictions appear paradoxical. One of my goals in this thesis will be to address this appearance of paradox. I do so by denying that restrictions bar actions of a certain type and instead bar consequentialist justifications for actions of a certain type.

For the most part this distinction does not come into play: given the nature of murder, it is a pretty rare thing for an act of murder to have any reason that may count in its favour. However, in some rare and exceptional circumstances—such as when a murder would bring about fewer overall murders—the distinction allows for the possibility that committing the murder is morally justified on some grounds other than agent-neutral value. Such a distinction avoids the paradoxical implication of categorically barring restriction violations regardless of the horrific nature of the circumstances while still managing to rule out restriction violations when the only reasons for violating the restriction are agent-neutral reasons.

A critique of consequentialism that I will consider is that of Bernard Williams'.

Although he is primarily concerned with utilitarianism, Williams does offer criticisms of

\(^{35}\) Ibid., 179.
consequentialism in general, but "only in order to suggest that some undesirable features of utilitarianism follow from its general consequentialist structure." 36 Williams characterizes consequentialism as follows: "Very roughly speaking, consequentialism is the doctrine that the moral value of any action always lies in its consequences and that it is by reference to their consequences that actions, and indeed such things as institutions, laws, and practices, are to be justified if they can be justified at all." 37 In short, one of the major reasons why utilitarians view morality the way they do is because they are consequentialists.

Williams' main concern about consequentialist moral theories is that they do not focus on the individual qua actor. Specifically, consequentialism does not view actions as holding a special moral significance. Williams claims that the consequentialist holds that some things have instrumental value, and other things have intrinsic value: "I take it to be the central idea of consequentialism that the only kind of thing that has intrinsic value is states of affairs, and that anything else that has value has it because it conduces to some intrinsically valuable states of affairs." 38 While this central tenet of consequentialism might appear too vague, Williams argues that it is more useful if we take it to be implying what is not of intrinsic value: actions. According to consequentialism, the value of actions is not intrinsic. An action's value lies in its causal properties. 39 In other words, a consequentialist needs to deny that what ever can make a

36 Smart and Williams, Utilitarianism, 81.
37 Ibid., 73.
38 Ibid., 83.
39 Ibid., 84.
state of affairs the best state of affairs is that someone did some action, because the consequentialist argues the explanation is the other way around: what makes an action the right one is that it produced some state of affairs.\textsuperscript{40}

Williams points out that his debate with the consequentialist need not be all or nothing. It is simply not true that someone who denies consequentialism needs to argue that doing some action is right “whatever the consequences”.\textsuperscript{41} Such a moral theorist need only deny that “with respect to some type of action, there are some situations in which that [the action] would be the right thing to do, even though the state of affairs produced by one’s doing that would be worse than some other state of affairs accessible to one.”\textsuperscript{42} The denial of consequentialism does not imply that consequences are of no importance to moral justification, but rather the denial that they are the totality of moral justification.

Williams’ criticism of consequentialism centrally focuses on the consequentialist’s impoverished notion of action. Agents must bring about the best state of affairs available. States of affairs are said to be available to an agent if and only if the agent is situated in the causal nexus such that the agent may bring about the state of affairs:

Consequentialism is basically indifferent to whether a state of affairs consists in what I do, or is produced by what I do, where that notion is itself wide enough to include, for instance, situations in which other people do things which I have

\textsuperscript{40}Ibid., 87.
\textsuperscript{41} Ibid., 90 (italics in the original).
\textsuperscript{42} Ibid., 90.
made them do, or allowed them to do, or encouraged them to do, or given them a chance to do.\textsuperscript{43}

If what I do, or fail to do, would play an essential causal role in bringing about some states of affairs, then—depending on whether the state of affairs is best or not—I will be obligated or prohibited, respectively. According to Williams, the consequentialist's account of action is an impoverished one because it denies that the presence or absence of other agents in the causal nexus is of special relevance for assigning moral responsibility. Consequentialism essentially involves the notion of “negative responsibility”: “that if I am ever responsible for anything, then I must be just as much responsible for the things I allow or fail to prevent, as I am for things that I myself, in the more everyday restricted sense, bring about.”\textsuperscript{44} The consequentialist's notion of action—including as it does, negative responsibility—denies something that we take to be foundational to our moral thinking: “that each of us is specially responsible for what he does, rather than for what other people do.”\textsuperscript{45} At the root of Williams' challenge is the consequentialist's assumption in favour of an impersonal, or impartial, view of moral value. More specifically, Williams rejects the consequentialist's insistence that an impersonal account of value implies that actions cannot be of intrinsic moral concern.

Consequentialism needs to acknowledge the fact that human persons are partial: a human morality should permit persons to be partial toward their own actions, characters, and loved ones. Acknowledging our human partiality is essential for understanding why people are especially responsible for what they intentionally do, and hold limited

\textsuperscript{43} Ibid., 93.

\textsuperscript{44} Ibid., 95.

\textsuperscript{45} Ibid., 99.
responsibility for allowing or permitting the actions of others. I will argue that permitting persons to take an attitude of partiality toward their own actions, characters and loved ones is a rational response to the fact that persons value the world from the first person point of view. While I think partiality has limits, I do not think these limitations should be set from the impersonal perspective, as the consequentialist would insist.

1.5 A Plausible Methodology

I have claimed that a consequentialist account of moral justification fails to be a plausible account of moral justification for human persons. I will briefly explicate such a claim.

Robert Noggle argues that our morality needs to reflect our morally relevant features. First, Noggle argues for what he calls his “negative thesis”: we demand that moral theory may not out-and-out contradict the nature of persons. Any moral theory that demands that persons do the physically or psychological impossible would be a moral theory we reject tout court. Human moral theory needs to reflect the morally relevant nature of human persons. For instance, one popular argument against some versions of utilitarianism relies on the claim that it is impossible—in a practical sense—for persons to accurately calculate expected utility for each and every action performed. Such forms of utilitarianism make impossible demands on epistemically and temporally limited creatures like us; a moral theory for humans needs to acknowledge these facts.


47 The form of consequentialism I aim to reject in this thesis—a form of consequentialism that is not a theory of decision-making but rather a form of moral evaluation—does acknowledge the epistemic and temporal limitations of humans (Chapter 1, 3.2.1).
Consider a specific example: the human capacity to suffer is a morally significant feature of human persons that plays an essential role in determining the nature and content of foundational moral norms. The moral norm “agents should not cause other agents to needlessly suffer” can be seen as a response to the moral significance of the fact that human persons have the capacity to suffer. Any adequate theory of morality for human persons must acknowledge this feature of our nature in order for the theory to be an intelligible theory of human morality. I take this methodology to imply that any plausible moral theory cannot be a moral theory for human persons if it fails to appropriately acknowledge the significance of human persons’ morally significant traits.

Furthermore, and perhaps more controversially, some morally relevant traits have a “polarity”. Some morally relevant traits are lamented—we would wish ourselves rid of them if possible (e.g., our epistemic limitations). Such traits or features of persons have a negative polarity. Persons also have features that have a positive polarity (e.g., fellow feeling, or sympathy). Features of persons with a positive polarity are features of persons that a moral theory must not merely acknowledge, but must actually embrace in order for the moral theory to be a moral theory for human persons. These features are features that we would not wish to change about ourselves even if we could: these features are at the foundation of our being moral agents.

Noggle’s methodology allows further clarity regarding my claim that the consequentialist account of moral justification fails to be an adequate theory of moral justification for human persons. The consequentialist’s account of moral justification is too broad because it fails to adequately acknowledge the moral significance of the

---

separateness of human persons and it is too narrow because it fails to adequately acknowledge the fact that human persons value the world from the first person point of view. I will now turn to an overview of my argument for this critique of a consequentialist account of moral justification.

1.6 An Overview of My Argument

The consequentialist account of moral justification is both too broad and too narrow. It is too broad in that it includes moral justifications for assaults and killings on the grounds that the assaults or killings produce the best state of affairs; it is too narrow in that it excludes moral justifications that stem from our natural partiality. What makes my project novel is that I see the consequentialist's incorrect account of moral justification stemming from a general failure to adequately reflect the morally significant features of human persons. *Our* moral theory is for human persons: its norms have to adequately reflect morally relevant aspects of the nature of human persons.

My plan is the following. In Chapter 2, I will explore the metaphysics of the separateness of persons. Parfit's assumption that the natural separateness of human persons has no moral significance is rejected. I argue that this morally significant feature of persons grounds our "person practices". Chapter 3 argues that the separateness of persons is the rationale behind restrictions that protect innocent persons from being grossly assaulted or killed in the name of the greater good. Chapter 4 explores and eventually rejects the consequentialist's "charge of paradox" against restrictions by supporting a distinction between restrictions on action types and restrictions on consequentialist justifications. In Chapter 5 I argue that the consequentialist's account of
moral justification is too narrow: from a consequentialist perspective, human partiality is lamented instead of embraced. In the Appendices, I apply the limitations on consequentialist justifications I have defended, first by examining cases relevant to the "war on terror", and second I offer a more detailed examination of the historical case of forced sterilization for eugenic purposes. I will now offer a more detailed overview of each chapter in turn.

One morally significant feature of human persons that consequentialists overlook is our natural separateness. Consequentialists fail to treat each and every different person as a distinct unit of moral concern because their primary focus is on bringing about—the best state of affairs, impersonally judged. Thus, consequentialism may permit instances of grave injustice for the purpose of promoting overall value. Persons ought to be afforded protection against such injustices because our bodies and lives are necessary for human persons to be separate and distinct units of moral concern. If a moral theory is to reflect the nature of persons, then consequentialism needs to incorporate restrictions that will bar people from being treated as mere instruments for the purpose of bringing about the best state of affairs. Persons should be permitted to bring about a better world, but they cannot use this permission to justify actions that violate restrictions against grossly assaulting or killing innocent people.

My support for restrictions is not as tidy as I might wish: tragic dilemmas do occur. On the one hand, when the only way to prevent a moral catastrophe is to actively kill or assault an innocent person, the consequentialist holds that it is irrational not to minimize the number of killings overall; consequentialists claim that we should steel ourselves to
do what is, from an agent-neutral perspective, optimal because the action is optimal.\(^{49}\)

On the other hand, some supporters of restrictions will defend the claim that a person must not kill or assault an innocent person no matter the circumstance or impersonal costs.

While I grant that tragic dilemmas do occur, I reject the consequentialist's insistence that such cases thereby demonstrate the paradoxical nature of restrictions. A weaker account of restrictions—one that can coherently be defended by the separateness of persons and yet is not so blunt that it is inviolable—can find a middle path between the extremes of demanding that we minimize total killings in the name of the greater good and categorically forbidding killings no matter the tragic nature of the circumstances. Restrictions that limit the *consequentialist justifications* for assaulting or killing innocent people—rather than restrictions on all actions of this type—represent an account of restrictions that manages to avoid paradox. In order to defend this account of restrictions, I need to defend the claim that agent-relative reasons are relevant to solving such dilemmas against the consequentialist's insistence that the only thing relevant to morality is promoting agent-neutral value.

I will challenge the consequentialist's defence of agent-neutral value as the sole focus of morality by examining the moral significance of human partiality. Many have argued that the demands of moral theory must be constrained by the reality of how humans actually engage the world.\(^{50}\) Failing to acknowledge the significance of agent-relative reasons, the consequentialist proposes a moral theory that limits moral


justification that stem from our partial concern for our own actions, lives and loved one. Consequentialism fails to focus on the intrinsic moral value of actions. Consequentialists offer an account of morality that laments our natural partiality because they defend agent-neutral limitations on moral justifications that stem from our natural partiality.\footnote{Samuel Scheffler, \textit{The Rejection of Consequentialism}, 41-79.}

In the Appendices, I apply the limitations on consequentialist justification in two cases. First, I offer a few consequentialist arguments in favour of the “war on terror” and respond to each in turn. Second—and in much greater detail—I examine the consequentialist arguments that were offered in favour of forced eugenic sterilization at the turn of the 20\textsuperscript{th} century. I have chosen this historical case because educated and well-intended people defended the systematic practice of grossly assaulting innocent persons on the grounds that doing so would promote the overall good. These arguments managed to convince thousands of judges, social workers and health care workers: tens of thousands of such procedures were carried out in North America over the first half of the 20\textsuperscript{th} century. It is my hope that my examination of this historical case acts as a cautionary tale for the problems inherent in adopting an exclusively consequentialist approach to moral justification.
2 THE SEPARATENESS OF PERSONS IS MORALLY SIGNIFICANT

2.1 Introduction

My first main goal of this thesis is to deny the consequentialist's claim that people are permitted to bring about certain states of affairs solely on the grounds that they are the best states of affairs, impersonally judged. In other words, I will defend the claim that consequentialism is too broad: it offers moral justifications for actions and institutions that a moral theory for humans should not. The bulk of this chapter is a response to Parfit's claim that a reductionist account of persons offers some support for consequentialism. I will argue that, even if Parfit's metaphysical arguments about personal identity are correct, the natural separateness of persons is morally significant and cannot be overlooked or ignored.

That there are many different persons no one denies. Human persons are separate from each other in the sense that, barring special circumstances, each person has his or her own body that is physically distinct from the bodies of others: rarely is the health and life of one person directly related to the health and life of any other person. Although the fact that persons are separate from each other is not in dispute, the moral significance of this fact is a serious point of contention between consequentialists and non-consequentialists.

---

52 A version of this chapter has been accepted for publication. Tim W. Christie, "Natural Separateness: Why Parfit's Reductionist Account of Person Fails to Support Consequentialism", Journal of Moral Philosophy [forthcoming].

53 For human persons this claim is the norm, with the exception of some cases of conjoined twins. Unless otherwise specified, I will use "person" to refer to "human person" throughout this chapter.
On the one hand, most non-consequentialists appear to agree that the separateness of persons has a moral significance that counts against the plausibility of consequentialism. Some non-consequentialists, such as Rawls and Nozick, take the separateness of persons to be a deep fact about the nature of persons and use it to explicate such concepts as justice and property.\(^{54}\) On the other hand, consequentialists such as Derek Parfit have criticized non-consequentialists, arguing that it is morally appropriate to overlook the separateness of persons. According to Parfit, the moral point of view is one that overlooks the specifics of who receives benefits and burdens because doing so is the most efficient way to produce the best state of affairs. Consequentialists advocate person-neutrality because it is the size of harms and benefits that is of moral importance, not who is harmed or benefited. Parfit holds that a reductionist view of personal identity offers support for such a moral theory: it is because the separateness of persons is a shallow fact that Parfit believes that consequentialists are correct to overlook its purported moral significance.\(^{55}\)

Many people who discuss the separateness of persons—not only Rawls and Nozick, but also David Brink, Dennis McKerlie and Anthony Laden\(^{56}\)—focus on exploring

---


\(^{55}\) Derek Parfit, *Reasons and Persons* (Oxford: Clarendon Press, 1984), 342. Derek Parfit specifically claims that his reductionist theory of personal identity offers support for utilitarianism. It is because Parfit’s reductionist account of persons offers support for utilitarianism in virtue of utilitarianism’s focus on outcomes that his account would also offer support for consequentialism (or C) more generally. After all, Parfit defines utilitarianism as the combination of C (“There is one ultimate moral aim: that outcomes be as good as possible”) with a utilitarian account of ‘best outcomes’ (e.g. Hedonistic version, etc.). See Parfit, *Reasons and Persons*, 24-26.

metaphysically-laden accounts of its moral significance. None of these theorists adequately address the more general reductionist charge that the separateness of persons fails to be of any moral significance at all. However, both Johnston and Korsgaard have defended a non-metaphysical understanding of the nature of persons.\(^5^7\) I will draw on the practical and natural accounts of persons defended by Korsgaard and Johnston, respectively, in order to defend the moral significance of the separateness of persons. I will set out and criticize Parfit's reductionist challenge against the significance of the separateness of persons. Let us assume Parfit is correct that persons are not separate in a "deep" sense; it simply does not follow that the natural separateness of persons is morally irrelevant. Parfit's reductionist view of persons fails to supply support for consequentialism.

2.2 Rawls and Nozick on the Separateness of Persons

Before we examine Parfit's account of personal identity, and why he takes this account to offer some support for consequentialism, it will be fruitful to consider an overview of why some philosophers take the separateness of persons to count against the plausibility of consequentialism.

2.2.1 Rawls

In Rawls' *A Theory of Justice* he explicitly discusses the separateness of persons: he argues against utilitarianism on the grounds that it does not adequately acknowledge this

feature of persons. The problem lies in how utilitarians view persons in relation to the production of the good. Rawls suggests the utilitarian reasons as follows:

Just as the well-being of a person is constructed from the series of satisfactions that are experienced at different moments in the course of his life, so in very much the same way the well-being of society is to be constructed from the fulfillment of the system of desires of the many individuals who belong to it.\(^{58}\)

The utilitarian wishes to focus on the satisfactions a person experiences at different moments and then construct an account of well-being; furthermore, the utilitarian employs the analysis of an individual's well-being as an analogy for understanding the well-being of a society as a whole. The analogy not only fails because persons and societies are too different for the analogy to be credible, but also because it is questionable whether the utilitarian can make sense of the idea of "well-being" by looking at individual and temporary "satisfactions" and not the entire life a person lives.

The problem with utilitarianism is not that it totally ignores the separateness of persons, but rather that it does not \textit{appropriately} acknowledge this feature of persons. Historically, utilitarians have focused on pleasure and pain; there is an obvious sense in which such a focus acknowledges that each person has distinct experiences (i.e., pleasures and pains). The fact that each person has distinct experiences implies that each is separate from the other in some sense. The problem Rawls is highlighting does not simply reject the utilitarian conception of the good; more particularly, Rawls is concerned with \textit{how} the utilitarian conception of the good generates an account of right action. For the utilitarian, right action is defined exclusively in terms of producing the good (e.g., pleasure, well-being, etc.). This account of right action is the source of the problem: so

long as good states of affairs are produced, the separateness of persons has no intrinsic moral significance. Rawls sums up his critique as follows: “utilitarianism does not take seriously the distinctness between persons.”59

Implicit in this critique against utilitarianism is the view that the separateness of persons is morally significant. What is the particular nature of its significance? In order to answer this question, we need to first turn to a brief overview of Rawls’ views on distributive justice. Behind the veil of ignorance, rational agents looking out for their own best interest would choose an egalitarian distribution scheme such that the least advantaged members of society always have lexical priority in the distributions of benefits and burdens. That is to say, whenever a society is deciding between schemes for distributing benefits and burdens among its members, the scheme chosen always ought to make the worst off somewhat better than what they otherwise would be, before improving the conditions of other groups.60 Assuming that persons in the original position would choose a principle of equal liberty, the demand that social institutions maximize the overall good is a demand that simply would not be made.61

Several moral theorists, including Brink and McKerlie, have commented on the difficulty of interpreting Rawls’ critique of utilitarianism—and consequentialism more generally—in such a way that his criticism can be rendered consistent with the rest of his theory of justice. The best way to understand Rawls’ claims about the moral significance


60 Ibid., 132-33.

61 Ibid., 24.
of the separateness of persons is to focus on what has been labelled the "Objection to Balancing". McKerlie explains that the Objection to Balancing is the following claim:

a harm to one person cannot be morally outweighed by benefits to other people even though the benefits are greater than the harms. If benefits and harms belong to different lives they cannot morally cancel one another in the way that views like utilitarianism supposes.62

This interpretation sounds plausible. Consider the following passage from Rawls:

"Justice denies that the loss of freedom for some is made right by a greater good shared by others. The reasoning which balances the gains and losses of different persons as if they were one person is excluded."63

Yet, Rawls’ views resist such a straightforward reading. It is difficult to understand coherently what type of balancing Rawls wishes to rule out, in light of his principles of distributive justice. Is Rawls arguing that it is never the case that hardship on some people could be balanced away by a great amount of benefit enjoyed by others? Prima facie, there appear to be many counter examples. Consider our medical institutions: if there is a highly contagious outbreak of the plague, it is in the society’s best interest that social institutions temporarily curtail the liberty of the people who may have been exposed to the plague by quarantining them. Risk-averse agents behind the veil of ignorance would agree with such a policy. The case of permissible quarantine represents a situation where a few people suffer a temporary loss of liberty in the name of the greater good for everyone else; also, such a loss of liberty, prima facie at least, appears consistent with Rawlsian justice as fairness.


Can Rawls’ Objection to Balancing be seen as an account of the moral significance of the separateness of persons, and yet also be rendered consistent with his maximin principles of redistribution? Brink argues that the Objection to Balancing can be understood as a response to the fact that people have needs. Some of these needs are urgent, like our need to be protected from violence. When these urgent needs are not met, the person has grounds for “serious complaint.”[^64] If the boundaries among persons may be overlooked, one problem would be that the seriousness or size of a burden any one person could be asked to bear for the common good would be of no special moral concern.[^65] The gravity or seriousness of one person’s burden would be purely instrumental; it could be drowned-out by the minor goods that many thousands of people acquire. While it may be maximally efficient in terms of producing the impersonally judged best state of affairs, ignoring the boundaries among persons would appear to make urgent needs of no special moral significance.

Brink’s point about the moral significance of the size of certain burdens sheds light on McKerlie’s understanding of Rawls’ Objection to Balancing. When balancing benefits and burdens, individuals must be treated as “morally important units.”[^66] In order to treat people as units of moral concern, the balancing of benefits and burdens must be carried out in such a way that the boundaries between lives are recognized. Particularly onerous burdens may be permissibly balanced only when they are outweighed by a correspondingly large benefit occurring to one person (the benefit is not merely spread

[^64]: Brink, “The separateness of persons,” 270.

[^65]: Ibid., 267.

out across different lives).\(^{67}\) Rawls is objecting to the way utilitarianism—and consequentialism more generally—aggregates the claims of many different people, such that the relatively minor concerns of many people could outweigh the individual with the strongest single claim.

Brink and McKerlie have offered a plausible account of Rawls’ Objection to Balancing. We now understand what is central to the objection: persons are units of moral concern and consequentialist moral theories fail to adequately acknowledge this fact. Balancing extreme burdens on the few with greater benefits on many others denies that the person being grossly burdened is a unit of moral concern who has grounds for serious complaint. Granting people security from being forced to bear such extreme burdens ought to take priority over concerns of producing good states of affairs.

2.2.2 Nozick

Nozick draws on the significance of the separateness of persons in an effort to attack the consequentialist—and specifically utilitarian—position. Nozick’s view on the significance of the separateness of persons is best understood in the context of his metaphysically deep account of persons. Nozick begins *Anarchy, State and Utopia* with the following claim: “Individuals have rights, and there are things no person or group may do to them (without violating their rights).”\(^{68}\) While Nozick is upfront about the fact

---

\(^{67}\) McKerlie considers such a reading of Rawls’ Objection to Balancing; he calls it the Objection to Aggregation see: McKerlie, “Egalitarianism,” 222.

\(^{68}\) Nozick, *Anarchy, State and Utopia*, ix.
that his book does not offer moral arguments that support this assertion,\textsuperscript{69} he does offer a starting point. Nozick follows Locke’s lead and imagines what individuals would be like in the state of nature: “Individuals in Locke’s state of nature are in ‘a state of perfect freedom to order their actions and dispose of their possessions and persons as they see fit’ (sect. 4) … The bounds of the law of nature requires that ‘no one ought to harm another in his life, health, liberty or possessions’ (sect. 6).”\textsuperscript{70} Even in the state of nature, individuals have rights, and other individuals must respect these rights. The rights to life, health, liberty and possessions are part of our nature; Nozick holds that these rights are deep moral facts.

How should these rights be honoured? Nozick proposes that the way to honour these natural rights is that individuals must constrain their actions and goals when either involve other persons. In effect, rights represent “side constraints” on morally (and politically) acceptable courses of action. Nozick writes,

\begin{quote}
Side constraints upon action reflect the underling Kantian principle that individuals are ends and not merely means: they may not be sacrificed or used for the achieving of other ends without their consent. Individuals are inviolable.\textsuperscript{71}
\end{quote}

Nozick’s view is that, so long as a person does not violate another person’s rights, the person may do as he or she wishes.

Freedom—defined as making and acting upon morally permissible choices—is at the moral and political foundation of Nozick’s metaphysical picture of individuals. Freedom

\textsuperscript{69} Nozick writes, “This book does not present a precise theory of the moral basis for individual rights,” Nozick, \textit{Anarchy, State and Utopia}, xiv.

\textsuperscript{70} Nozick, \textit{Anarchy, State and Utopia}, 10.

\textsuperscript{71} Ibid., 30.
is at the basis of Nozick’s moral system because he thinks this freedom is a necessary condition on our generating a meaning of life. Nozick writes,

A person’s shaping his life in accordance with some overall plan is his way of giving meaning to his life; only a being with the capacity to so shape his life can have or strive for meaningful life.  

Only a person can give his or her life meaning. It is because persons are such creatures that their permissible choices are—and ought to be treated as—morally sacred.

Given this metaphysical account of persons, Nozick argues that consequentialism fails to recognize that each person is an inviolable unit of moral concern. Under consequentialism, the violation of side constraints would occur so long as it is optimal, overall, to do so. Nozick thinks consequentialism in general, but especially utilitarianism, would tend to ignore the fact that different persons live different lives because such kinds of moral theories mistakenly take the unit of moral concern to be the whole society. In *Anarchy, State, and Utopia* Nozick writes,

there is no social entity with a good that undergoes some sacrifice for its own good. There are only individual people, different individual people, with their own individual lives. Using one of these people for the benefit of others, uses him and benefits the others. Talk of the overall social good covers this up.

The individual becomes insignificant when the individual’s needs or ambitions are in conflict with this super-entity. Nozick’s charge is that consequentialism is flawed because its unit of moral concern is something that does not exist. Nozick proceeds to argue that, under consequentialism, the borders among persons that separate one person from another should be ignored. Yet, Nozick holds that it is improper to ignore these boundaries because doing so violates our nature as bearers of inviolable rights.

---

72 Ibid., 50.

73 Ibid., 32-33
Demanding that people incur uncompensated sacrifices is to violate the side-constraint against harming innocents and needlessly curtailing liberty.

As diverse as Rawls’ and Nozick’s moral views, they both hold that the separateness of persons has a moral significance that counts against utilitarianism specifically, and consequentialism in general, because such theories would fail to treat persons as units of moral concern. In order to treat people as units of moral concern, there need to be moral limits on what agents may do to other agents, even for morally worthwhile purposes. I will now turn to Parfit’s response to Rawls’ and Nozick’s views on the separateness of persons.

2.3 Parfit’s Metaphysics

Parfit points out that Rawls’ and Nozick’s moral theories depend upon metaphysically laden claims about the nature of persons. Parfit challenges both moral theorists’ implicit reliance on a non-reductionist account of persons. Parfit argues that a reductionist account of personal identity is a metaphysically superior account of the nature of persons compared to a non-reductionist one, and then argues that his reductionist view offers some explanation why it is appropriate for consequentialism to fail to acknowledge the separateness of persons. First, I will examine Parfit’s metaphysics, focusing specifically on his reductionism. Second, I will explain Parfit’s claim that his reductionist view of persons offers support for consequentialism.

Parfit relies on a dichotomy between deep and shallow facts. Deep facts are facts about things that exist independently of our practices. Deep facts are universal features
of reality that are the way they are independently of us. Facts that are not deep (i.e., shallow facts) are facts that are true because of us (facts about, e.g., chess, nations or clubs). While Parfit does not extend this terminology, allow me to take “deep entities” to be entities that exist independently of us, and “shallow entities” to be entities that depend on us for existence. Deep entities are metaphysically fundamental entities. If humans did in fact have Cartesian egos then this fact about us would be a deep fact (and Cartesian egos would be deep entities).  

Shallow entities are not, metaphysically speaking, fundamental. Most philosophers are reductionists about the identity of clubs or nations over time: a club’s or nation’s continued existence involves nothing other than (enough of) its members’ continued existence and its members’ behaving together in certain ways. Parfit writes, “Suppose that a certain club exists for several years, holding regular meetings. The meetings then cease. Some years later, some of the members of this club form a club with the same name, and same rules.” Is this second club the very same club? Or, is the second club another club that is exactly similar? According to Parfit, the statement, “This is the same club,” would be neither true nor false.

Parfit thinks that our views about personal identity over time can be informed by our views on the identity of a club over time. Persons are like clubs, Parfit argues, in the sense that both clubs and persons have intrinsically indeterminate identity conditions over

---

74 Parfit, *Reasons and Persons*, 336 and 339. This reading of Parfit is deeply indebted to Johnston, “Human Concerns.”


76 Ibid., 213.

77 Ibid., 213.
time. If we choose to draw a sharp line where no line exists—say, in a personal identity puzzle-case—then "we cannot believe that it [the line we drew] has, intrinsically, either rational or moral significance."78 The particular placement of the line is metaphysically arbitrary. Parfit holds that the line "cannot justify any claims about what matters" because the placement of the line merely depends on our practices.79

2.4 Parfit: A Reductionist View of Persons Supports Consequentialism

Parfit claims that the most common view about personal identity is a non-reductionist view. A non-reductionist view takes persons to be separately existing entities: persons are not just their bodies and psychology, but are something over and above these impersonally describable phenomena. The best-known non-reductionist views are those that hold persons to be purely mental entities such as Cartesian egos. In order to determine personal identity when a doubt or question arises, the non-reductionist just needs to know whether the person still has the very same ego.

Parfit's own view on personal identity over time is a reductionist view. The two most prevalent reductionist criteria of personal identity over time are the physical criterion and the psychological criterion. The physical criterion, as the name suggests, holds that personal identity over time depends on the continuity of a person's body and brain. The psychological criterion of personal identity over time is more functional in nature: so long as two persons, X and some past person Z, have enough "psychological continuity", then X and Z are identical persons (assuming that no other person also has

78 Ibid., 241.

79 Ibid., 241.
this same level of psychological continuity). On the widest version of the psychological criterion, any cause at all that produces enough psychological continuity will result in person X and Z being the same person.\textsuperscript{80} One key difference between reductionist and non-reductionist views of personal identity over time is that reductionist views accept that, in some cases, a person’s identity may be indeterminate; in contrast, the non-reductionist denies this claim.

Parfit employs many well-known thought experiments—such as the combined-spectrum thought experiment and the split-brain thought experiment—which purport to demonstrate the indeterminate nature of personal identity over time.\textsuperscript{81} According to the reductionist, if you found out that you were going to be the subject of a mad scientist’s neurological experiment, the issue of whether you survive your medical ordeal or not would be an ‘empty question’. An empty question is a question that has no answer unless we stipulate one. Parfit’s “combined spectrum” thought experiment is supposed to offer evidence for the indeterminacy of identity in at least some logically possible cases:

Combined Spectrum: Imagine a range of cases. At one end of the spectrum, I awake tomorrow after a mad scientist has conducted his experiments on me and find that I am both physically and psychologically identical with the person I was before the experiment. At the other end of the spectrum, the resulting person has no continuity with me as I am now, either physically or psychologically. In this later case, imagine that the scientist’s experiment destroyed my brain and body, and then created, out of new organic matter, a perfect Replica of someone else. Let us suppose this new person is Greta Garbo.\textsuperscript{82}

Parfit is interested the range of cases in the middle of the spectrum for the purpose of proving the indeterminacy of personal identity. These cases involve the scientist partially

\textsuperscript{80} Ibid, 207.

\textsuperscript{81} Ibid., 239.

\textsuperscript{82} Ibid., 236-237. This case is a paraphrase of Parfit’s thought experiment.
destroying his patient's brain and body, and proceeding to use an equal mix of the patient's organic matter and new organic matter to construct a person that is halfway between the patient and Greta Garbo. For the cases in the middle of the spectrum, Parfit argues that there would be no answer to the question of whether or not the patient survived the ordeal. The question of survival would be empty. Although Parfit does think that personal identity over time is indeterminate, he also argues that we should not let this fact bother us: what matters on issues of survival is not identity but rather psychological connectedness and continuity with any cause (Parfit's 'Relation R').

According to Parfit's view of personal identity, persons exist but they are merely shallow entities. Thus, persons can, in some cases, have indeterminate identity conditions over time because their identities are fully reducible without remainder to the holding of more particular facts that vary (or may vary) on a continuum. The four-dimensional boundaries of persons are intrinsically fuzzy. In such cases of indeterminacy, we do create conventions for everyday life—our 'person practices'—but, being shallow, the boundaries generated by our conventions are of little moral significance. Parfit explains:

[Reductionists] regard the unity of each life as, in its nature, less deep, and as a matter of degree. We may therefore think the boundaries between lives to be less like those between, say, squares on a chessboard, dividing what is all pure white from what is all jet black. We may think these boundaries to be more like those between different countries. They may then seem less morally important.

When it comes to persons and nations, we are not carving the world at the joints because there is no way-the-world-is such that it can be so carved. Parfit seems to think that,

---

83 It is worth emphasizing that Parfit is not saying that persons ought not have any moral significance because, like e.g. witches, persons fail to exist.

84 Parfit, Reasons and Persons, 339.
because we carve out these entities by convention, their boundaries lack moral significance or, at least, the significance of these boundaries may permissibly be overlooked. Parfit assumes that metaphysical primacy is a more appropriate carrier of moral significance than something that is not metaphysically primary, and he assumes that our level of caring should match metaphysical depth.

How does this view of persons offer support for consequentialism? Parfit claims that in virtue of utilitarianism’s consequentialist account of right action, his view on personal identity over time offers support for utilitarianism. The goal of a consequentialist moral theory is to produce good states of affairs. Parfit’s reductionist account of personal identity highlights one systemic source of inefficiency at producing good outcomes. When questions about personal identity are empty, the boundaries between ‘self’ and ‘other’ are intrinsically fuzzy. Such a finding is relevant to the doling out of benefits and burdens: maximizing the good should not be constrained by historical concerns of justice or by worries about which future person deserves particular benefits or burdens. Benefits and burdens ought to be distributed in whatever way most efficiently produces the best state of affairs. Treating persons as special, integral units with fixed identities and moral claims across time would thwart efficiency and, thus, should be elements of a moral theory that are sometimes overlooked or, more radically, simply jettisoned altogether.

---

2.5 Korsgaard and Johnston on Parfitian Impact

I argue that Parfit's reductionist account of personal identity over time does not offer support for consequentialism. In order to accomplish this goal, I will need to defend the moral significance of the separateness of persons without drawing on controversial metaphysical claims about the nature of personal identity over time. My view will follow the lead of Korsgaard and Johnston.

In "Personal Identity and the Unity of Agency," Christine Korsgaard argues that the underlying reasons for an agent to view him- or herself as the same agent who "will occupy my body in the future" are practical reasons and not metaphysical ones. There are two sources of practical necessity for agents viewing themselves as one entity at a time. First, having but one body with which to act, we must see ourselves as a unified agent or else fail to meet our basic biological needs (e.g., fail to consume nutrients). Second, there is the practical necessity of adopting a deliberative standpoint. Over time, Korsgaard claims, there is a practical necessity involved with viewing ourselves as a continuing agent: "In choosing our careers and in pursuing our friendships and family lives, we both presuppose and construct a continuity of identity and of agency." Based on these necessary practical concerns, Korsgaard argues that Parfit's claims about the practical significance of his reductionist account of persons are false because Parfit has assumed that the connection between unity and agency must be metaphysical. The lesson I take from Korsgaard's criticism of Parfit in relation to the separateness of person

87 Ibid., 111.
88 Ibid., 113.
89 Ibid., 115.
is that, like the connection between unity and agency, the separateness of persons need not be metaphysically deep in order for it to be of moral significance.

If persons are not separate in a metaphysical sense, then in what way are persons separate? In his paper “Human Concerns without Superlative Selves,” Johnston points in the direction of an answer to this question.90 Johnston is in general agreement with Korsgaard: it is not true that either our practices of taking persons to be separate over time must be based in deep entities like souls or else be morally insignificant. Johnston offers another alternative, namely, that our collective practices that demarcate the boundaries among persons over time—our ‘person practices’—take their basis in our natural features. Humans are naturally the kinds of creatures that have self-referential patterns of concern. This said, our person practices, while they have a natural basis, are not determined uniquely and specifically by nature; these practices are not immune to criticism. Thus, Johnston agrees that our person practices are not self-justifying, but he also thinks he can appeal “to a broadly coherentist view of justification” for the entire practice.91 He writes: “Concerns that are natural and fundamental have a certain kind of defeasible presumption in favour of their reasonableness; they cannot all be thrown into doubt at once, for then criticism would have no place from which to start.”92 While Johnston applies this insight to challenge Parfit’s view about what matters in cases of survival, I will focus on developing Johnston’s claim that our person practices have a natural basis.

90 Johnston, “Human Concerns,” 149.
91 Ibid., 158.
92 Ibid., 158.
2.6 Natural Separateness and our Person Practices

The separateness of human persons has a natural basis in the sense that humans have several natural features that constrain our person practices. First, all humans have a particular history. Access to the particulars of this history is a valuable resource for a human: telling the difference between friend and foe, food and poison, and mate-worthy or not mate-worthy are abilities that are bound up with recalling the details of one’s particular history. A human’s access to his or her particular history, especially his or her social history, is a key to individual and group survival. Second, each human is an individual living organism: humans are physically separate from other things in ways similar to how all organisms are separate from other things. It is empirically false that humans are a “super organism” like, e.g., a hive of bees or a colony of ants. Lastly, most humans are capable of rational prudence. The human ability to plan for the future allows humans not only to weather the tough times nature may throw our way, but also (possibly) to flourish. Planning for the future is one of the key evolutionary resources of our species. These natural features of humans place constraints on our person practices.

Our person practices reflect these natural features of humans. First, our person practices understand persons to be separated in the past. We access the past via memory. Our person practices take these memories to be a key basis for our self-knowledge and self-unity. Second, our person practices understand persons to be separate in the present, both physically and mentally. Persons are physically separate from others and may, for the most part, do as they wish with their own bodies. Persons are separate actors: when a

---

person tries to do something or think something, there is no other person that is also, thereby, doing or thinking something. Our person practices understand a person to bear a special responsibility relation for what he or she intentionally does and only what he or she intentionally does. Lastly, our person practices take persons to be separate in the future. Persons have plans, commitments, and projects. The only way that it would be possible for plans, commitments, and projects to be significant in the way that they are significant is if persons keep separate from one and other in the future.

It is worth emphasizing that, as we encounter them and as we are them, we take persons to be separate from each other; exceptions or confusing cases are few and far between. Our person practices, for the most part, operate undetected: they structure our thinking about our self and others, often doing so without our being consciously aware. Yet, strange or unusual cases may challenge our otherwise all but unnoticed person practices. Perhaps the most difficult of these types of cases are diseases or injuries that have long-term physiological and psychological impact. I argue that, even if Parfit is correct that in a case of this sort the metaphysical identity of the person is indeterminate, it nevertheless is false that our person practices lack a moral significance. In fact, it is in these confusing cases that our person practices become most salient.

Consider the following case: Ted is seriously injured in a work-place accident. While his brain surgery is considered successful, some tissue is removed and as a result Ted is left in a coma. Six weeks pass and then Ted starts to awaken. Ted’s long-term memories are absent: he does not remember graduating from high school, marrying his wife, or being a father of two young daughters. Ted suffers from amnesia. Luckily, Ted’s doctor says his condition is at least partially reversible. Undaunted by the
difficulties that Ted's recovery imposes on his two daughters and wife, Ted's family works tirelessly to assist him in whatever recovery he can manage.

The case of Ted is unusual but not unheard of. Furthermore, the attitudes and reactions of Ted, his family and his doctor do not strike us as odd or inappropriate. On Parfit's reductionist account of personal identity, there is no metaphysically justified answer to the following question: when the man in the hospital wakes up after being in a coma for 6 weeks (hereafter, post-coma man), is this person the same person he was before the accident (i.e., Ted)? The case of Ted represents a potential counter-example to the claim that persons are separate: on the Parfitian view of persons, it is plausible to claim that Ted and post-coma man are two persons that fail to be metaphysically distinct over time. The four-dimensional boundaries between Ted and post-coma man are intrinsically fuzzy. Parfit would claim that it is merely our person practices that decide the issue of whether post-coma man is identical with Ted. As mentioned above, Parfit holds that boundaries of identity drawn by us cannot "justify any claims about what matters." Consider the following questions: Is Ted's wife still married, or is she now a widow? What property, if any, does post-coma man own? Who is responsible for post-coma man's hospital bill? If Parfit's analysis is correct, then whatever practice-based answers we provide to these questions, we should view them as lacking justification.

Yet answers to these questions do have justifications: our person practices provide us with a reasonably clear basis from which to answer them. In particular, our response to almost all cases of patients who recover from a coma is to claim that the person who survives the ordeal is the same person as the person who suffered the trauma in the first

---

place.\textsuperscript{95} There might be no deep reason for demarcating the identity of the man in the hospital, but this lack of depth does not entail that there are no justified answers to questions that pertain to the man’s identity.

Parfit’s claim that his reductionist view of persons supplies support for consequentialism suggests one of two claims: when confronted with a claim of identity over time that merely depends on our practices, we should either consider abandoning some of our person practices in favour of direct appeals to consequentialism, or, more radically, consider abandoning all of our person practices. While some practices may be abandoned when it is especially inefficient to operate within the dictates of the practice in a particular instance, sometimes such a piecemeal approach does not work. On the one hand, consider the North American practice of driving on the right hand side of the road: in a particular instance when compliance is too costly, the practice ought to be abandoned. On the other hand, as Rawls argues, our practice of legal punishment ought not be abandoned in cases when the practice produces an unappealing result.\textsuperscript{96}

Rawls famously defends an important distinction between justifying an action within a practice and justifying the practice itself.\textsuperscript{97} Thus, while it might be the case that a society’s practice of punishing the guilty is a practice that maximizes utility, Rawls argues that concerns of utility should not figure in a particular instance of punishing the

\textsuperscript{95} While the case of Ted could be tweaked in order to cast doubt on the claim of identity, such convoluted and atypical cases bolster my argument because even in such extreme cases our default position is that the person who suffered the trauma and the post-coma person ought to be treated as though they are the same person.


\textsuperscript{97} Rawls, “Two Concepts of Rules,” 147.
guilty. Within the practice of punishing the guilty, the only concern is to determine whether the suspect did, indeed, commit the crime in question; the issue of how much utility an innocent or guilty verdict will generate in a particular case is beside the point. Rawls' insight about what may—and what may not—count as a justification for making a judgement within a practice is relevant to the issue of justifying claims of identity overtime when our person practices are involved.

In the case of Ted, the Rawlsian lesson applies: as Tamar Schapiro might explain, Parfit is assuming our person practices are mere methods for producing good states of affairs.98 A method may be abandoned when it is optimal, overall, to do so because a method's value lies in its efficiency at producing a particular goal. A practice, unlike a mere method for producing outcomes we value, represents a joint venture designed to attain a common end. We come to identify with our practices in the sense that we come to see them as what makes us who we are: as Shapiro notes, “Hence we might say that, whereas a method defines a way of solving a problem, only a practice defines a way of making a problem, along with its solution, count as ours.”99 What is true of practices in general is doubly true for our person practices: our person practices are significant and ought not be abandoned when it might be efficient to do so because they, in part, constitute our identity. While our person practices may be tested in a case like Ted’s, it is in cases of this exact sort that our person practices are morally significant for the people involved. Parfit is wrong to think that the metaphysical shallowness of personal identity


should lend support to the claim that our person practices may permissibly be overlooked when it would be, from a consequentialist perspective, optimal to do so.

2.7 Morality and Our Species' Nature: The Mergers and the Collectives

If I am correct so far, I have shown that our ‘person practices’ are morally significant and should not be abandoned in particular cases via a direct appeal to consequentialism. It is an open possibility, however, for the consequentialist to argue for radical revision, namely that our entire practice of thinking of persons as temporally-extended, moral entities is a practice that fails to be morally justified. Given that I have argued that our person practices reflect the natural separateness of persons, the next stage of my argument needs to address the following question: is the natural separateness of persons morally significant? I argue that it is: if our human nature were different, then it would be plausible to hold that the moral norms we take to be fundamental would have to be jettisoned or reconfigured in ways that would seem to leave, in their place, something less than moral norms. I take this relationship of dependence to constitute evidence that our natural separateness is morally significant.

Persons have features that are morally significant. A feature of persons plays a role in human-morality when that feature is a necessary aspect of morality for human persons. One salient example of a morally significant feature is the human capacity to suffer. One way to demonstrate this role is by a counter-factual thought experiment. Imagine a species of creatures exactly like us in every way except that this species lacks

---

the feature of persons being examined (e.g., the capacity to suffer). Is it plausible to hold that such a creature’s morality is different from human-morality? Yes: it is constitutive of human-morality that it involves, or says something about, the fact that we suffer. No complete human-morality would lack a principle that addresses the moral significance of human suffering, but the same is not true of a species that does not suffer. A reasonable hypothesis that explains the differences between a moral theory for human persons and a moral theory for persons of this other species is that our capacity to suffer is of moral significance.101

By way of contrast with suffering, some features of human persons would not be of moral significance. For instance, human persons are naturally bi-pedal. This feature is not of moral significance: if we conducted a similar thought experiment, it is not true that moral principles would change in any significant way. It is not plausible that human moral systems depend upon, or respond to, the fact that humans are naturally bi-pedal. In what follows, I argue that the natural separateness of persons is more like the fact that human persons are capable of suffering, and less like the fact that humans are naturally bi-pedal.

Suppose that there exists a race of beings that are like humans in all ways, with the exception that they lack the separateness of persons that human persons in fact have. I can imagine two ways that such beings would not be separate from each other. First, we could imagine these beings not being separate in body. Second, we could imagine these beings not being separate in consciousness. I discuss each of these possibilities in turn.

Imagine a race of beings exactly like humans except that they spontaneously merge bodies (the ‘mergers’). Once every few years, a merger would find him or herself to be a ‘single’ (an unattached merger). The moment two singles come in close physical proximity, the two individuals would reflexively and uncontrollably shake hands, and then the two hands would spontaneously merge. Once fused, the two beings would remain fused for several years; the health and life of each single would be hopelessly entangled with the other. Both members of the pair would be aware of what occurred to the shared appendage: if the appendage is tickled, they both laugh; if the appendage is cut, they both feel pain. Nonetheless, we may suppose that each member of the pair of mergers would still have the same sets of desires, intentions, and preferences that he or she had previous to the current merging. Furthermore, consent from both mergers is not necessary for one merger to try to initiate an action, move about, etc., and each merger could thwart or co-operate as he or she saw fit. In most cases, mutual co-operation would be essential for a merger to live his or her day-to-day life. After 3 or 4 years as a pair, the beings would spontaneously split and the process would start all over again. Compared to human-morality, would these beings’ morality be systematically different?

Obviously, it is difficult to imagine exhaustively the many differences between merger-morality and human-morality. Even so, the most salient moral difference between the moral lives of such beings and our own would be that, once merged, it would not be true that each person has his or her own life to live. For the most part, living would be a continuous compromise between oneself and one’s current partner. In a

---

102 I am indebted to an anonymous referee for pointing out that David Braddon-Mitchell and Caroline West have considered examples that bear some similarity to the one under discussion; see: David Braddon-Mitchell and Caroline West, “Temporal Phase Pluralism,” *Philosophy and Penomenological Research* 62, (2001): 59-83.
literal sense, one's body would not be entirely one's own. The very basic freedom of movement that is common for most human persons would not exist for mergers.

One particular upshot of this difference for the mergers would be that the morality of promise-keeping would need to be quite different because it would be far more common for it to be physically impossible for mergers to fulfill self-imposed obligations without the cooperation of their current partners, their future partners, or both. For instance, long-term commitments would be difficult to fulfill for mergers because the commitment might conflict with the equally important commitments of a new partner (e.g. a merger makes a promise to remain in a particular city, but then later becomes merged with a partner who has career commitments that necessitates frequent and lengthy travel). In such cases of conflicting commitments, only one of the mergers is physically capable of staying true to past obligations. Merger-morality would be quite different from human-morality, at least in terms of the details regarding self-imposed commitments like promises.

Aside from freedom of movement and choice being seriously curtailed, another difference between merger-morality and human-morality would be these beings' norms that involve desert. If one member of a pair has been proven guilty in a capital case, almost all forms punishment would directly affect both members of the pair—i.e., two people and not one. Merger's norms that deal with capital punishment would systematically differ from humans'; merger-morality would also incorporate different moral principles on the appropriateness of lengthy incarceration. Consider the case of a merger who is a violent psychopath. In order to protect society from this merger's violent rampages, merger morality might permit the psychopath's incarceration even in
cases where his imprisonment entails the incarceration of his completely innocent partner. Until the pair of mergers separates, the innocent member of the pair would languish behind bars. While human-morality might condemn locking up known innocent persons, merger morality might view such a case as morally innocuous. Locking up a known innocent merger might be viewed as a morally acceptable cost of keeping merger society safe.

The mergers demonstrate that our separateness of persons is morally significant because, if we were not separate in body, it is plausible to hold that our morality would be systematically different. My hypothesis is that it is our natural separateness of persons that can account for this systematic difference. Yet, it might be argued, the example of the mergers shows only that some details of merger-morality are systematically different from human-morality. Perhaps a systematic difference in the details of the morality of species of persons is not enough evidence to claim that the separateness of persons is a morally significant feature. Merger-morality might be systematically different but it would not be different in kind from human-morality. Perhaps this criticism of my first thought experiment is plausible. However, even if the plausibility of such criticism is granted, there is little doubt about the significance of the second way in which a species of beings could fail to be separate.

Imagine a species of beings that have bodies just like ours but share a common mind: when one of the beings remembers something or imagines an image, the memory or image is common to all (the ‘collective’). When one member of the collective suggests

---

103 Perhaps the psychotic merger committed the violent rampages while he was a single, and it is only now that he has merged that the evidence of his foul deeds have come to light. Alternatively, the capital offences may have been committed after the psychotic merger drugged his current partner into a stupor.
to the common mind a course of action, all members of the collective know of the
suggestion and a common intention is reached. All members of the collective are
mentally transparent to each other. Benefits and burdens are experienced by all: one
member’s pain is another’s. If one member of the collective wins the lottery, all of the
members feel the elation of a large windfall. Similar science fiction examples
immediately come to mind (e.g., Star Trek: The Next Generation’s ‘Borg Collective’). 104
For the collective, how would morality be different from ordinary human-morality?

The moral differences between collective-morality and human-morality would be so
substantial that it is hard to imagine. The fact that pain and pleasure are common to all is
of obvious moral significance. Unlike human morality, epistemic issues regarding the
relative size of benefits and burdens would not be a source of concern for collective-
morality. Furthermore, issues involving individual responsibility would not arise for
collective-morality: each member of the collective does what the common mind, on
balance, intends. The collective as a whole may share responsibility for an act that one
member of the collective did, but it is not at all clear that individual members would bear
a responsibility for the actions at all.

Collectives would not be like a species of individuals at all; instead, they would be
more like a unitary moral agent. In the same way that the “cells of a human body”
compose an entity of special moral worth, each member of the collective would be more
like a unit that composes something of moral importance than like a thing that is itself of
special moral worth. 105 Without the type of separateness of consciousness that humans

104 The example is Robert Noggle’s; See also, Noggle, “Nature of Persons,”561.

105 Noggle, “From the Nature of Persons,” 560. See also: R. M. Chisholm, “Scattered Objects,” in On
do possess, the basic structure of collective-morality would be vastly different in kind when compared to human-morality. I hypothesize that the source of this difference is that these beings lack the separateness of persons. If this hypothesis is plausible, then the separateness of persons has a moral significance that our human morality ought not overlook. Moral theories cannot ignore or overlook the morally significant features of humans and expect to be a moral theory that applies to human persons.

It is worth noting at this point that the science fiction thought experiments I have considered above have obvious limitations: perhaps my intuitions are idiosyncratic, or at any rate, not shared widely. Thought experiments in normative morality, while common, are also recognized as holding limited significance. I will briefly discuss some real world analogues to both the ‘mergers’ and the ‘collectives’ that are at least somewhat familiar to moral theorists.

The obvious analogue for the ‘mergers’ is the case of conjoined twins. Conjoined twins are rare; conjoined twins that survive into adulthood are rarer still. The range of difficulties for conjoined twins—both physical and psychological—are well documented. Of particular interest for my present purposes are cases of a set of conjoined twins with different and mutually exclusive conceptions of, e.g., the good life. Consider the conjoined twins Eng and Chang Bunker (b. 1811): “If, as all available sources indicate, Eng was a complete abstainer then Chang must have been, indeed, a prodigious drinker.” Charged to the twin’s account on a hotel bill for June 1st of 1870

167-173. An object can be scattered around a room (e.g. a set of dishes, or a set of golf clubs). Sets of dishes and sets of golf clubs can be thought of as “scattered objects”. Perhaps the Collective would view itself as a “scattered agent”?

were two bottles of wine, eight bottles of beer, one flask of cognac and one bottle of whiskey.\textsuperscript{107} Chang and Eng were joined at the liver; in 1874, Chang died first, from a stroke, and a previously healthy Eng died two hours later.\textsuperscript{108} The case of Chang and Eng makes explicit my claim of dependence between natural features of humans and our morality: it is obvious that commonly accepted norms fail to apply to Chang and Eng precisely because of their natural conjoined status. The complex moral issues this actual case exhibits—issues dealing with personal autonomy for two different persons who literally share a physical body—are of exactly the type that I claimed would be common fare for merger-morality.

The analogue for the ‘collective’ in the actual world is much more common: corporations are considered to be \textit{legal} persons in at least some respects under United States law. Many moral theorists have argued in favour of corporate persons coming to be viewed as \textit{moral} persons, in at least some respects.\textsuperscript{109} For my present purposes, it is enough to note that there is much current debate about the appropriate range and extent of corporate morality vis-à-vis other agents and the existence of this debate lends credence to my claims about how different an exclusively collective-morality would be compared with human-morality. While the analogy between the ‘collective’ and corporations has intriguing prospects, I merely suggest the possibility of this analogue but will not defend its appropriateness.

\textsuperscript{107} Smith, \textit{Psychological Profiles}, 105.

\textsuperscript{108} Segal, \textit{Entwined Lives}, 306.

2.8 An Objection and a Reply

The consequentialist could argue that the non-consequentialist’s account of persons is—to borrow a phrase from J. L. Mackie—such a “queer” one that it is rationally unacceptable.\(^\text{110}\) It is far more parsimonious to hold that non-consequentialists are allowing desires to cloud their judgement. Just because many people want to believe that “personhood” as it has traditionally been conceived represents an account that survives rational scrutiny, wanting does not make it so. Let us call this objection the “queerness objection”.

My response to the queerness objection is, in part, to accept it. The nature of human persons does influence our thinking; to deny this claim would be absurd. What I deny from the above challenge is that ‘persons’—qua shallow entities with moral significance—are so queer that their queerness ought to concern us. At least a few other equally queer entities exist.

Consider an example squarely situated in the moral context, namely, a particular promise. Suppose I agree to paint my neighbour’s garage in exchange for dinner. This particular promise is a shallow entity, in the sense that the promise does not have an independent existence apart from our promise conventions. Yet, this promise I made has an obvious moral significance: the specific wording of the promise—in conjunction with our conventions that surround the giving and taking of promises—has morally bound me to uphold my end of the deal. Even a reductionist like Parfit agrees with such a view. Parfit argues that his complex view of personal identity challenges whether a promisor can “unbind” a promisee after long periods of time; nowhere does Parfit suggest that

promises are not of moral significance. In fact, Parfit’s views presuppose that promises have a moral significance that ought not be overlooked.111 The fact that promises are shallow entities does not in any way offer support for reducing their moral significance. Based on the case of promises, it appears as though having a ‘deep’ metaphysical nature is not a necessary condition on some entity having moral significance.

Examples of shallow entities with significance are not peculiar to morality; other examples abound. One such example is Johnston’s discussion of an ancient Roman statue.112 A reductionist view of a marble statue would hold that the object’s statue-hood only consists in the material from which it is made, its shape, and its other relational properties; there is no further fact of the matter regarding the existence of the statue. A statue’s four-dimensional boundaries depend on our aesthetic conventions: qua aesthetic object, the statue is a shallow entity. Consider a partially broken Roman statue (i.e., the arms are missing and its weather-worn surface has many chips and cracks). The statue is still an object of aesthetic interest: it is judged that there is enough of the original art-object remaining that it ought to still be regarded as a statue rather than as a destroyed statue, i.e., a weather-beaten, non-aesthetic lump of marble. The ancient Roman statue is aesthetically significant because of its unique properties and relations, including its particular historically determined properties. Even though statues, qua art-objects, are shallow, they are paradigm examples of objects with aesthetic significance.

111 Parfit, Reasons and Persons, 326-339.

112 Johnston, “Human Concerns,” 152.
Lastly, consider one of Hillary Putnam’s themes in *Ethics Without Ontology*, namely, the case of our mathematical practices.\(^\text{113}\) Putnam argues that a deep metaphysical nature is not a necessary condition on the rational significance of, e.g., a particular proof. Suppose a person questions the metaphysical status of a mathematical proof (e.g. “Is that mathematical proof really an ‘entity’?”). Generating a satisfactory response to such a question is not a necessary condition on our *caring* about the production of a particular proof. Mathematical proofs are examples of entities with rational significance; debates about whether they are *really* entities are beside the point because debates about what conventions ought to govern our use of ‘entity’ in mathematics are debates that leave the rational value of particular mathematical proofs unchallenged.\(^\text{114}\)

I have presented three separate types of shallow entities that have moral, aesthetic and rational significance. One of the costs of maintaining the queerness objection against persons would be either to hold that promises, statues and proofs are all equally queer (and thus have a type of significance that is appropriate to overlook) or else to deny that promises, statues and proofs are shallow entities. Presumably, both options would be seen as unsavoury. Johnston’s thesis of minimalism—“the view that the metaphysical picture of the justificatory undergirdings of our practices do not represent the real conditions of justifications of those practices”—offers a third alternative.\(^\text{115}\) The third alternative is that the metaphysical depth of a practice bears no relation to justifying the


\(^{115}\) Johnston, “Human Concerns,” 149-150.
rational, aesthetic, or moral significance of a part of the practice. Arguing that the natural separateness of persons is morally significant is, in part, an attempt to develop Johnston’s thesis.

2.9 Conclusion

Our person practices are a reflection of the empirical separateness of persons. The separateness of persons takes its basis in the morally relevant aspects of our species’ empirical nature; it has a moral significance that should not be over-looked. Persons must be treated as units of moral concern, where this is understood to imply that there are some burdens that people cannot be asked to bear, even when it would be morally expedient to impose such a burden. Parfit is wrong that his reductionist view of personal identity offers support for consequentialism in general, or utilitarianism specifically. Assuming that the consequentialists are attempting to offer a moral theory for human persons, they overlook the moral significance of the separateness of persons at great peril.\(^\text{116}\)

What is in peril? What specific problems arise for a theory that fails to adequately account for the moral significance of the separateness of persons? It is to these questions that I turn next.

\(^{116}\) I would like to thank the many people who have read drafts of this chapter and then took the time to offer succinct criticisms and encouragement, including Zohar Geva, Dominic Lopes, Andrew Irvine, Catherine Wilson, Cory Fairley, two anonymous referees and especially Scott Anderson.
3 CONSEQUENTIALISM, INTUITION AND RESTRICTIONS

3.1 Introduction
Both Scheffler and Pettit agree that an agent is *always permitted* to bring about certain states of affairs solely on the grounds that they are the best states of affairs, impersonally judged (AP).\(^{117}\) Any moral theory that affirms AP fails to incorporate "restrictions"—that is, moral limitations on certain action types. In this chapter I will argue that a moral theory must incorporate restrictions that protect the bodies and lives of persons in order for the theory to acknowledge the moral significance of the separateness of persons.

Here is an outline of what is to come: first, I will offer a brief overview of Parfit's argument from Chapter Two as it relates to restrictions. Second, I will present an argument for restrictions that rests on intuition. Third, I will consider a few replies to this argument, including an examination of Pettit's attempt to defend *consequentialist*-based restrictions. The problem with Pettit's restrictions is that they will not be capable of performing their intuitive function, namely, protecting people from being sacrificed merely for the greater good. Fourth, I agree with the consequentialist that restrictions cannot be properly supported by the appeal to intuition on its own, but then argue that at least two restrictions take their rationale from the separateness of persons: restrictions on killing the innocent and restrictions on grossly assaulting the innocent.

3.2 Parfit on Restrictions
An overview of Parfit's view on the moral significance of the separateness of persons is in order. Consequentialists overlook the separateness of persons. Moral theorists have

---

\(^{117}\) From here on I will refer to this claim as the 'always permitted' claim, or AP.
argued that ignoring the moral significance of the separateness of persons is a key (perhaps the key) failing of consequentialist theories like utilitarianism. Parfit critiques such arguments by holding that a reductionist view of persons offers support for utilitarianism—and consequentialism more broadly—because this view of persons explains why it is appropriate for consequentialists to overlook its purported moral significance.

On the one hand, Parfit admits his reductionist account of persons offers some support for expanding the scope of various moral duties and restrictions. For instance, take the case of the moral requirements of redistribution: in addition to redistributions taking place across lives, redistribution should also take place within a single life. The case Parfit considers is the “Child’s Burden”: imposing a hardship on a young child for the greater good of the future person he will become can, on the reductionist view, be seen as redistribution of goods within a single life.\(^{118}\) On the other hand, this expansion of distributive principles is morally irrelevant because Parfit holds that the consequentialist should deny that distributive principles have any moral significance, or “weight”.\(^{119}\) How does Parfit defend this claim?

In the same way that consequentialists do not care when benefits and burdens are distributed over time (they care only about the size of the net benefit), the consequentialist does not care who receives benefits or burdens. The consequentialist defends this view because it is appropriately impersonal. A reductionist account of persons offers support for such an impersonal morality because, as the reductionist view

---


\(^{119}\) Ibid., 335.
of persons has shown, the unity of each life “is in its nature less deep.” A person
should be treated in whatever manner leads to the maximization of overall impersonal
benefit. The reductionist view of persons represents the “partial disintegration” of
persons, and thus shows that the consequentialist is correct to assign no weight to the
separateness of persons. It is on these grounds that the consequentialist defends the
appropriateness of “overlook[ing] the boundaries among different lives.”

In Chapter 2, I developed two related arguments against Parfit’s position. First, I
argued that our person practices take their basis in the natural separateness of human
persons. Second, I argued that our person practices should not be abandoned on a case-
by-case basis—or in their entirety—because our natural separateness of persons has a
moral significance: many deeply held moral principles would have to be abandoned if
humans were not separate in the way that we in fact are separate from each other. I took
this dependence to constitute evidence for the moral significance of the separateness of
human persons. Based on these two related arguments, I rejected Parfit’s claim that the
consequentialist may overlook the separateness of persons.

Not being permitted to overlook the separateness of person, must consequentialism
acknowledge restrictions on how persons may be treated? It is to this issue I now turn.

3.3 The Argument From Intuition

Commonsense suggests that some actions—such as killing or grossly assaulting innocent
persons—are so heinous they are never permissible. “Restrictions” try to make clear and

---

120 Ibid., 336.

121 Or, at least, such little weight that it may permissibly be overlooked; See: Parfit, Reasons and Persons, 336.

122 Parfit, Reasons and Persons, 331.
specific what this commonsense thought suggests. On the most general account, restrictions purport to be some form of moral limitation: it is not merely the impersonal cost of violent and murderous actions that renders them morally impermissible, but rather because such actions have an intrinsically immoral nature. At first glance, restrictions appear to be at odds with consequentialism: consequentialists affirm AP, and therefore they do not categorically bar acts like killing or assaulting the innocent.\textsuperscript{123}

The defenders of restrictions (hereafter referred to as the "anti-consequentialists") draw on well-known cases involving run-away trolleys and calculating surgeons as evidence in favour of the overly permissive nature of any moral theory that incorporates AP. Anti-consequentialists simply deny that all acts that produce the best states of affairs are, ipso facto, morally permitted.

Restrictions purport to bar agents in certain ways, even when great good is at stake. Consider the following case:

\textbf{Case 1:} Suppose a very old man suddenly has a change of heart. He contacts his attorney and asks her to change his current will: instead of giving his millions to his Mafioso nephew, he now wishes to leave his money to a local charity group that assists the homeless and mentally handicapped. However, the change of heart turns out to be short lived. The next day he calls his attorney and wishes to change his will back to the way it had been. Upon her arrival at the house, the attorney meets the old man at the top of a long flight of steep, concrete steps. Suppose, just as he was about to sign the necessary documents, the old man takes a misstep. With but the slightest of a shove in the right direction the attorney knows that she has the opportunity to bring about the old man’s death and that the best state of affairs will ensue.\textsuperscript{124}

\textsuperscript{123} Scheffler holds that “one would always be permitted but not always required to perform the act that would produce the best available outcome overall.” Samuel Scheffler, Introduction, \textit{Consequentialism and its Critics}, ed. Samuel Scheffler (Oxford: Oxford University Press, 1988), 11. Therefore, if an agent is situated such that the only way to produce the best state of affairs, impartially judged, is to commit a murder, then the agent is permitted to do so.

\textsuperscript{124} In order to make Case 1 more plausible, feel free to add what details you see fit (i.e., no one else will ever know what happened, the lawyer will not be caught, etc.). Case 1 bears a passing resemblance to Shelly Kagan’s “rich uncle” example; see: Samuel Scheffler, \textit{The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions}, rev. ed. (New
On the one hand, the anti-consequentialist holds that it is impermissible for the attorney to shove her client down the steps. Actions like murder are simply inconsistent with treating persons as they deserve. On the other hand, consequentialists argue that, in the normal course of events, actions of the murderous variety are morally barred on the grounds that killing an innocent person lowers the overall value of the ensuing state of affairs. However, in the context of Case 1, the killing of an innocent person appears to be the means of attaining the best state of affairs. Consequentialists like Pettit and Scheffler would argue as follows: ensuring the money is left to charity will promote impersonal value far more so than standing by and allowing that the money goes to the nephew. Accordingly, in Case 1, the means of promoting the best state of affairs involves doing what it takes to ensure the inheritance goes to a charity group rather than a gangster. In virtue of AP, the consequentialist will hold that the attorney is permitted to shove her client down the steps.

Based on the consequentialist’s assessment of Case 1, the anti-consequentialist argues as follows:

1) Any moral theory that permits killing an innocent person on the grounds of impersonal costs is a moral theory that ignores the separateness of persons;
2) As Case 1 shows, consequentialism permits the killing of an innocent person on the grounds of impersonal costs;
3) Therefore, consequentialism ignores the separateness of persons;
4) A moral theory that ignores the separateness of persons is so unintuitive that it ought to be rejected;
5) Therefore, consequentialism is so unintuitive that it ought to be rejected.125

---


125 This argument will be referred to as the “argument from intuition.”
How might the consequentialist respond to the “argument from intuition”? One possible response is for the consequentialist to reject premise two, namely, the claim that consequentialism permits the killing of an innocent due to impersonal costs. There are two versions of this response: one version emphasizes epistemic concerns and another version emphasizes the nature of the good. I will consider both possibilities in turn.

The consequentialist may balk at the claim that consequentialism permits the lawyer to shove the old man down the steps: there are just too many variables that the lawyer would need to factor into the equation (e.g., Will I get caught? If caught, what long-term negative consequences will I have caused? etc.). Such a case involves many different values: the value of personal security, the value of autonomy, and perhaps others. There is no way to predict the long-term consequences of promoting and deterring so many different values. In short, the consequentialist could agree that in Case 1 it may well be the case that the way to bring about the best state of affairs would involve killing the old man, but then argue that no agent would ever be epistemically situated such that they could know this fact. As such, the lawyer in Case 1 is not morally permitted to shove the old man down the steps.

I think such a line of defence is a red herring. For the consequentialist to reject the example on the grounds that it is too complex is to confuse the epistemic issues regarding what a decision maker could, or should, reasonably know about the consequences of her actions with the factual claim of what will produce the best state of affairs. The version of consequentialism under discussion (a version that affirms AP) is not primarily a theory of moral decision-making but is a theory of moral evaluation. So long as the act

---

126 I argued in Chapter 1 why I am focussing on a form of consequentialism that purports to be objective. See Chapter 1, Section 3.2.
actually brings about the best state of affairs, the agent is permitted to bring these states of affairs about. Case 1 stipulates these epistemic worries do not arise, and I think this stipulation is reasonable. A reply to the argument from intuition along these epistemic lines is not a successful reply.\footnote{Unless, of course, the features of consequentialism I outlined in Chapter 1 are rejected.}

An alternative to this epistemic objection to premise 2 is for the consequentialist to simply reject the assessment of Case 1. The consequentialist could argue that killing the old man in Case 1 does not produce the best state of affairs. Consequentialism does not have to be coupled with any specific account of the good. It could be argued that a sophisticated account of the good would derive a ranking principle that accommodates our intuitions about Case 1. Certain acts, such as murder, lower the value of future states of affairs to such a great extent that these acts would never bring about the best state of affairs. A sophisticated account of the good would somehow factor into the assessment of future states of affairs that an act of murder is evil. Therefore, Case 1 does not present a consequentialist with a problem because the consequentialist could—in accordance with our intuition—claim that the lawyer should not kill the old man.

I acknowledge that this response to the argument from intuition has some plausibility. Case 1—and my assessment of it—does make some implicit assumptions about the nature of the good (e.g., that the good involves human welfare, respecting persons, autonomy, and perhaps other features). However, I do not think that my implicit assumptions about the good play an essential role in the proposed consequentialist assessment of Case 1. As I noted in Chapter 1, consequentialism is a goal-based moral theory; the focus of consequentialism is on producing the good. Thus, when some action
promotes the good—whatever it may be—consequentialism is incapable of barring the action. It is worthwhile to remember Pettit’s claim regarding the relationship between actions and agents: “Consequentialists see the relation between values and agents as an instrumental one: agents are required to produce whatever actions have the property of promoting a designated value.”\footnote{Philip Pettit, “Consequentialism,” in A Companion to Ethics, ed. Peter Singer, (New York: Blackwell, 1991), 231.} When assessing what action an agent ought to perform, a goal-based moral theory cannot appropriately acknowledge the intrinsic significance of actions and remain a goal-based moral theory. Therefore, consequentialists cannot claim that the theory of the good employed by a goal-based moral theory makes room for acknowledging the intrinsic moral significance of some types of actions.

While it might be the case that the goal-based forms of consequentialism I am focussing on cannot offer the criticism of premise 2 suggested above, does this mean that all versions of consequentialism cannot? That is to ask: is a non-goal-based version of consequentialism possible? While it is surely outside the scope of my enquiry to adequately explore these questions, I think that the answer to both questions is “no” because consequentialism is essentially goal-based.\footnote{Take a particular example: suppose a consequentialist put forward a ranking principle that offered lexical priority to all states of affairs that included no murders over states of affairs with murders. When a consequentialist of this sort examines Case 1, the consequentialist must grant that a state of affairs with one murder is better than a state of affairs with many, irrespective of who committed the murders (the lawyer or the nephew). In Chapter Four I will examine cases of this sort far more thoroughly.} I cannot fathom how consequentialism could be coupled with a ranking principle that manages to bar the promotion of overall value while also managing to remain a consequentialist moral
However, simply because I cannot imagine such a version does not mean that such an account is impossible. For now, I must leave these questions incompletely answered. 131

Thus, given my overview of consequentialism from Chapter 1, and the proponents of consequentialism my thesis aims to address, my initial assessment of Case 1 remains plausible: the consequentialist must grant that killing the old man would bring about the best state of affairs. Therefore, due to AP, it is permissible for the lawyer to kill the old man. However, the consequentialist does not have to agree that the Argument from Intuition is sound. The consequentialist can “bite the bullet” and deny the fourth premise, namely, that ignoring the separateness of persons is unintuitive. It is to this response to the argument from intuition that I now turn.

3.4 Pettit’s Consequentialist-Based Restrictions

Philip Pettit holds that, properly understood, permitting the killing of an innocent may ignore the separateness of persons but doing so, in some extreme cases, is not unintuitive. Pettit’s argument has two steps. First, Pettit argues that consequentialism can protect people from being sacrificed by restricting the deliberative strategies of agents: over the usual course of events, agents should not be permitted to calculate the good produced by various actions, such as murder. These restrictions capture what intuition requires for acknowledging the separateness of persons. Second, Pettit affirms AP by arguing that,

130 A non-optimizing version of consequentialism does not require the optimal act, but it does permit it.

131 I will consider a related objection, called the ‘charge of paradox’, in Chapter 4. The charge of paradox claims that it is paradoxical for an agent to be barred from minimizing the occurrences of forbidden actions, even when the only way to minimize the occurrence of forbidden actions involves committing a forbidden action.
when agents are presented with an emergency situation, restrictions on deliberative strategies ought to be "turned off". Thus, when these restrictions are turned off, any act that maximizes the good is permitted. As such, Pettit denies that a moral theory that ignores the separateness of persons is so unintuitive that it has grounds to be rejected. Let's turn to a close examination of Pettit's overall argument.

Pettit holds that consequentialism should place restrictions on the deliberative strategies of agents because some values cannot be realized without doing so. For instance, a consequentialist can defend the claim that people ought to "foreswear calculation and calculative monitoring" when deliberating about how to treat other persons.\(^{132}\) For example, in order to have access to the value of loyalty one must restrict oneself in such a way that one does not constantly have an eye on profitable back-stabbing opportunities. In more technical language, the good of loyalty is calculatively elusive and calculatively vulnerable. A good is said to be calculatively elusive in cases such that if "the agent calculates over [the good] at the first order level, he cannot hope to achieve it." A good is said to be calculatively vulnerable "in the sense that even if [the agent attempts to calculate] at the second order level, monitoring some appropriate restrictive procedure, he must equally despair of [the value's] realization."\(^{133}\) Pettit argues that similar claims can be made to defend why persons should hold a "distinctive respect for persons."\(^{134}\)


\(^{134}\) Pettit and Brennan, "Restrictive Consequentialism," 138.
In an effort to defend the claim that consequentialism can recognize rights and thereby respect persons appropriately, Pettit argues as follows:

The dignity of the people with whom you deal is likely to be important to you and so if you are a consequentialist you should be prepared to forswear calculation and calculative monitoring in favour of the commitment – in effect, maxims – distinctive of respect for persons.\(^{135}\)

Pettit, in his article “The Consequentialist can Recognize Rights,” expands his position. He explains that because the good of human dignity is calculatively elusive and calculatively vulnerable, the consequentialist must restrict calculative deliberation in normal circumstances in order to achieve the good of human dignity. Under the plausible assumption that human dignity is considered to be an end worthy of pursuit, I think Pettit’s claims are reasonable. It would appear as though Pettit could argue that the lawyer in Case 1 would be restricted from seeing that pushing the old man down the stairs would promote the best state of affairs. However, Pettit does not argue in such a fashion, because he thinks such restrictions on an individual’s calculative deliberations would apply only on occasions “on which normal circumstances obtain.”\(^{136}\)

Pettit states that in “exceptional” circumstances—the particulars of which remain undefined—the consequentialist must abandon restrictive maxims on deliberation. Pettit explains as follows:

The consequentialist will abandon the sort of maxim that produces virtue, or unselfconsciousness, or loyalty, if that is known in a given instance to be genuinely for the best. In such an instance – however unlikely – he will even violate the maxim that ensures respect and rights.\(^{137}\)

\(^{135}\) Ibid., 138.

\(^{136}\) Pettit, “The Consequentialist can Recognize Rights”, 54.

\(^{137}\) Pettit and Brennan, “Restrictive Consequentialism,”140.
When is an agent in exceptional circumstances? Pettit is opaque on this issue, other than saying that the violation “is known to be in a given instance to be genuinely for the best”. He never offers any concrete guideline for what occasions count as normal and what occasions count as an emergency. Think back to Case 1: does the possibility of ensuring an inheritance goes to a charity instead of a gangster count as a “normal” circumstance or does it count as “exceptional” circumstance? How ought this judgement be made? The most plausible answer is that the consequentialist will explicate the difference between normal circumstances and exceptional circumstances in terms of the amount of good that is at stake.

Pettit is clear that agents have a duty to notice when they must “turn off” restrictions on their deliberation strategies. According to Pettit, persons need to be able to detect when occurrences represent an emergency. We need an “auto-pilot” on our calculative restrictions in the sense that they sometimes need to be “disengaged” in cases that are not normal, e.g., emergency cases:

The metaphor of the autopilot is appropriate... Just as an autopilot will be disengaged in emergencies, so [the restrictive maxim] is subject to escape clauses... Thus if the agent comes to learn in any instance that the best thing for him to do there is to break with the maxim, then he can have no grounds for not doing so.\(^{138}\)

If Pettit’s argument for consequentialist-based rights is sound, then the consequentialist can recognize rights over the normal course of events. During an emergency situation, however, these rights may be violated. Pettit holds that it is not unintuitive to respond in such a fashion to an emergency situation.

\(^{138}\) Ibid., 132; emphasis added.
However, the anti-consequentialist will not be impressed with Pettit’s consequentialist-based restrictions: the cases at the heart of the Argument from Intuition are cases where a lot of impartial good is at stake. Pettit’s restrictions are far too weak. While it is in emergencies that morality is pushed the hardest, it is in these cases that it is the most important. Pettit’s consequentialist-based restrictions allow for the possibility that restrictions should be violated when a lot of good is at stake. The anti-consequentialist denies that acknowledging the moral significance of the separateness of persons is consistent with restrictions that may be violated under such conditions. How can the anti-consequentialist defend this assertion?

Pettit’s consequentialist-based restrictions represent a method for producing the best states of affairs.139 A method for producing the best state of affairs may be abandoned when it is optimal, overall, to do so because a method’s value lies in efficiency at producing a particular goal. The lawyer in Case 1 is situated such that she has the chance to bring about a great amount of good. Case 1 looks as though it should count as an emergency situation. As such, according to Pettit, the lawyer must turn off her restriction on killing the innocent because this restriction is merely instrumentally useful at producing good states of affairs. The restriction is clearly not of instrumental value in this case and thus may permissibly be abandoned.

For the anti-consequentialist, restrictions are not a mere method for protecting persons but are an essential part of the practice of treating persons as units of moral concern. As I argued in Chapter 2, practices are not like methods because practices may

---

139 I discussed the distinction between a method and a practice in Chapter 2; see: Tamar Schapiro, “Compliance, Complicity and the Nature of Nonideal Conditions,” The Journal of Philosophy100, no. 7 (July 2003): 341.
not be abandoned on a case-by-case basis. As such, the fact that killing the old man would maximize overall impersonal value is not a morally decisive reason in favour of killing him because the function of the practice is not to maximize impersonal value—it is to treat persons as units of moral concern. While practices might not be as impersonally efficient as a method, they are more deeply entrenched: according to Tamar Schapiro, our identities become bound up with being the kind of person who follows this practice and being from a society that adopts this practice. Concerns of overall agent-neutral value are not decisive reasons for failing to act in accordance with the practice of treating persons as units of moral concern.

An objection could be raised: the consequentialist could claim that permitting the violation of the restriction against killing is consistent with the practice of treating everyone involved as units of moral concern. In Case 1, the option is between one life going well and many lives going well. The consequentialist thinks that opting for more lives going well overall simply is to treat everyone involved as a unit of moral concern. The concerns of many dozens of people—viewed as units of moral concern—have greater weight than the concerns of one unit of moral concern. While this response appears promising, on closer examination it is not. The consequentialist has assumed that the focus of our practice of treating persons as units of moral concern is to produce the best agent-neutral outcome; however, as my analysis of the moral significance of the separateness of persons in Chapter 2 showed, the practice may not be overlooked on a case-by-case basis and remain a practice. Practices are not just a communal solution designed to produce a certain end; our practices also come to define our selves and our

140 Ibid., 341.
community. In order to remain *who we are* we follow our practices, even in the cases where it is impartially inefficient to do so.

Pettit’s consequentialist-restrictions would not protect persons in emergency situations. When persons are in the greatest need for protection, Pettit’s restrictions are cast aside. Pettit’s restrictions may have a clear consequentialist foundation, but in light of this foundation such restrictions fail to adequately acknowledge the moral significance of the separateness of persons. In some specific and rare cases, our practice of treating persons as units of moral concern might have a high cost, impersonally judged, but this cost must be paid if we wish to acknowledge the moral significance of the separateness of persons. Pettit’s response to the argument from intuition—for him to deny the claim that any moral theory that ignores the separateness of persons is a moral theory that is so unintuitive that it ought to be rejected—has not proven successful.

I have canvassed two types of responses to the argument from intuition and they both have failed. However, the consequentialist has another option before they resort to abandoning AP. The consequentialist may challenge the anti-consequentialist’s reliance on intuition. Do restrictions have a plausible rationale? It is to this question I now turn.

### 3.5 A Commonsense Practice in Search of a Foundation

Many consequentialists grant that a moral theory needs to adequately acknowledge the moral significance of the separateness of persons; furthermore, some consequentialists even concede that we do have a practice that restricts some types of actions, such as killing the innocent. However, the fact that we do have a practice is not a *defence* for the practice. The consequentialist can simply argue as follows: the anti-consequentialist’s
argument for restrictions rests on mere intuition and practice. Intuition and practice, on their own, are not enough.

Consider the following (well-known) case:

**Case 2:** Sam is a surgeon at a large and busy hospital. He has several patients all in grave need of organ transplant. Barring a miraculous turn of events, these patients will all die unless they receive healthy organs. Sometimes miracles happen: one day while performing a mundane surgery on a different and otherwise healthy patient, Sam suddenly realizes the following: 1) the patient he is currently operating on is a perfect match for all five of his other patients; 2) his current patient is an organ donor; 3) with one well-placed stroke of the scalpel, Sam knows he could bring about his current patient’s death; 4) his patient’s death will appear to be a non-culpable accident; 5) killing his current patient will, in all likelihood, save the lives of his other five patients. Therefore, Sam has to choose: should he kill one innocent person in order to save five innocent people from certain death?

Is it right for Sam to purposefully kill his current patient—assuming, for the sake of argument, it is true that this action would produce the best state of affairs, impersonally judged? The consistent consequentialist has no *particular* qualms about permitting Sam to kill his current patient.\(^{141}\) The argument from intuition boils down to the claim that the consequentialist’s assessment of Case 1 and Case 2 is *vicious*: some acts that consequentialism may prescribes in some extreme cases are held to be morally repugnant actions. The consequentialist asks, “So what?”

The consequentialist could hold that the argument from intuition merely begs the question. For a variety of reasons—many of them morally legitimate—it is a good thing that decent and morally upright people blanch at the idea of ever killing an innocent person. Yet, when an agent is confronted with an extreme situation where lives hang in the balance, feelings have to be overcome. Morality is tough, and consequentialism sets

---

\(^{141}\) Obviously a careful cost-benefit analysis would need to be conducted: what are the chances that he is caught and exposed? Are there any long-term negative consequences that need to be taken into consideration? These questions, though challenging, need to be explored.
a high standard. In short, if it is (at bottom) merely feelings of squeamishness that stay
our hand in Case 1 and Case 2, then we ought to steel ourselves to do the moral thing. It
might be unintuitive to permit what consequentialism suggests we ought to permit; it may
also be against our current practices. These points to one side, we cannot hide behind our
morally illegitimate squeamishness in the face of a moral legitimate duty. Unless a
principled rationale can be presented for restrictions that bar us from bringing about the
best state of affairs in Case 1 and Case 2, then restrictions should be seen as “dangling
distinctions”—rationally unsupported features of commonsense morality that have
grounds to be rejected.¹⁴²

I think the consequentialists are correct to demand that restrictions need to be
supported by a principled rationale and not merely rest on intuition and current practice.
However, the consequentialist is wrong to hold that restrictions fail to have a principled
rationale. The separateness of persons offers a principled rationale for restrictions that
provide protection for the necessary conditions of human persons being a unit of moral
concern: the living bodies of human persons must be protected from permanent harm
because our bodies are empirically essential for each human person being a unit of moral
concern. It is inconsistent to hold both that persons are units of moral concern and that
persons may be killed or grossly assaulted when, due to strange or extreme
circumstances, it would be impersonally worthwhile to do so. In order for human persons
to be treated as units of moral concern, it is essential that we have a moral guarantee we

will not be killed or assaulted solely for the purpose of promoting overall impersonal value.\textsuperscript{143}

The argument from intuition needs to be improved by replacing its fourth premise, which, in turn, limits the scope of its conclusion. In light of my arguments from Chapter 2 regarding the moral significance of the separateness of persons, intuition is not needed in order to support at least some restrictions. The new version of the argument—now renamed the Argument From the Separateness of Persons—is the following:

1) Any moral theory that permits the killing an innocent person due to impersonal costs is a moral theory that ignores the separateness of persons;
2) As Case 1 and Case 2 show, consequentialism permits the killing of an innocent person due to impersonal costs;
3) Therefore, consequentialism ignores the separateness of persons;
4) A moral theory that ignores the separateness of persons is not a moral theory for human persons;
5) Therefore, consequentialism is not a moral theory for human persons.

Premise 4 rests on the account of the moral significance of the separateness of persons I defended in Chapter 2: human persons must be treated as units of moral concern, where being treated as a unit of moral concern is taken to imply that people cannot be forced to bear great burdens, such as being assaulted or killed, even when it would be morally expedient to impose such burdens. In light of this new 4\textsuperscript{th} premise, the new conclusion rejects consequentialism as an appropriate moral theory for human persons.

What burdens may not be imposed on human persons? That is to ask: what restrictions are generated by the Argument From the Separateness of Persons? I will defend two restrictions: restrictions on grossly assaulting and killing innocent human persons. These restrictions are a rational response to the power that agency provides:

\textsuperscript{143} It is worth emphasizing that I am identifying two necessary features of treating persons as units of moral concern; I do not even attempt to offer sufficient conditions on treating persons as units of moral concern.
restrictions are moral limitations on how this power may permissibly be used in the context of causing severe harm to human persons.

3.6 Two Restrictions

In order to support the claim that killings and gross assaults should be restricted, I will emphasize three aspects of these actions: the harm caused must be permanent, sufficiently costly, and non-fungible. A burden is permanent if it will not naturally heal. A burden is sufficiently costly when it ends or seriously curtails the quality of life for a person. Lastly, a burden is non-fungible when what is lost cannot be replaced by a different token of the same type without a resulting loss in (commensurate) value. For instance, if someone takes a sledgehammer to my car, my loss is fungible. The person can be sued for damages, and my car can be replaced by a different token of the same type (without loss in value). If someone takes a sledgehammer to my hand, my losses are non-fungible: no amount of money can (exactly) offset the damages I have incurred.  

3.6.1 Killing Innocent People

Killing-justified-by-the-greater-good is the most obvious denial that a person is a separate unit of moral concern. A necessary condition for an agent to continue to be a unit of moral concern is that the agent’s life is not terminated. Obviously, one of the features

\[144\] Or, more accurately, given current medical technology nothing is a fungible replacement for one’s hand. However, what is non-fungible today, may one day be a fungible loss in 100 years. How do we determine whether a claim to a non-fungible burden is appropriate? Surely some grey areas are inevitable: suppose a thief steals and melts down my 1911 gold coin. The loss might be considered fungible because the coin could be replaced by a type-identical coin. However, if what I valued with that coin is its history—that it was my Grandpa’s coin that he set aside for me the day I was born—then the coin is in fact non-fungible. While property crime would not represent a violation of any restriction I am proposing, the case of the gold coin might present my account of restrictions with some problems. One solution that may help in some cases would be that claims to non-fungibility must be inter-subjectively reasonable.
that render killing so horrible is that its effects are permanent. If a person is killed in error there is no way to “undo” this wrong; being killed is a non-fungible harm. Furthermore, being killed obviously passes the threshold for costliness.

Why is it not the case that restrictions apply to agent-caused deaths only, instead of all human deaths, regardless of cause? While I will not offer anywhere near a full answer to this challenging question at this time, I will explore it in more detail in Chapter 4. That said, I will offer a few points. Every human will die one day. Morality is not at odds with human death per se, but rather with persons playing an immoral role in another’s death. Some deaths are morally faultless deaths. For instance, most deaths caused by an earthquake, tsunami or other natural disaster are morally faultless deaths. While morally faultless deaths are bad or regrettable occurrences, no one is at fault because no human was involved.

3.6.2 Gross Bodily Assault

I will define gross-bodily assault as physical or psychological harm committed to a person’s body or mind that has permanent, costly and non-fungible repercussions. Also, the costliness must pass an inter-subjectively defined minimum threshold as grounds for legitimate complaint. Paradigm examples of such gross bodily assault are injuries caused by an agent that results in the loss of a limb or a person’s eyes. Given the current state of medical technology, such injuries will not naturally heal on their own (they are permanent), they have a sufficiently costly impact on the victim (a heavy enough burden that passes the threshold for legitimate complaint), and even the best and most expensive prosthesis in the case of a limb is not a fungible replacement for a person’s arm or leg.
It might be argued that a slippery slope threatens. If an injury that causes the loss of a limb counts as a violation of the restriction against gross bodily assault, what portion of the limb must be lost in order for the harm to merit the protection of a restriction? What of an injury that causes partial hand loss, or even the loss of a thumb? A pinkie finger? A fingernail? Where, and on what principled basis, are the lines drawn? With out clear boundaries for harms of this sort, a lot of actions that are plausibly justified on consequentialist grounds will turn out to be violating a restriction against gross bodily assault (perhaps in a health-care context: e.g., allowing economic factors to influence the justification for some medically expedient procedure that involves the amputation of a baby toe). It could be argued that due to the difficulty—if not outright impossibility—of drawing such lines, restrictions against gross bodily assault will be useless.

While there will be 'grey' or unclear cases, part of my definition will assist. "Costly impact" must pass an inter-subjectively valid minimum threshold as grounds for legitimate complaint. Furthermore, when we judge the negative impact of an injury, we may take into consideration specifics of the person’s life in order to make judgements regarding the costliness of the harm. Such concerns might help make reasonable judgments that stop the slippery slope from starting.

3.7 Are There Other Restrictions?

While a case can be made that there are other plausible restrictions, I will not defend such a claim. For example, a restriction against all actions that harm might be seen as plausible.¹⁴⁵ Some Kantians might wish to argue in favour of a restriction against lying.

Libertarians may lobby in favour of a restriction against "forced labour" (i.e., governments using coercive regimes to collect taxes for the purposes of wealth redistribution).\textsuperscript{146} Whatever the independent merits of broadening the list of restrictions, I think that it is not plausible that the \textit{separateness of persons} offers principled support for these other restrictions. Perhaps some other principled rationale could be found, but I will not take a position one way or the other on the likelihood of a successful search.

3.8 Conclusion

In order for a moral theory to be a moral theory for human person, it must adequately acknowledge the separateness of persons. Incorporating restrictions against grossly assaulting or killing the innocent, even in the name of the greater good, is necessary for a moral theory to adequately acknowledge the separateness of persons. The version of consequentialism I am examining—a version that affirms AP—fails to incorporate such restrictions and thus has grounds to be rejected.

\textsuperscript{146} e.g., Robert Nozick, \textit{Anarchy, State and Utopia} (New York: Basic Books, 1974).
4 ANSWERING THE CHARGE OF PARADOX

4.1 Introduction

The goal of this chapter is to respond to an objection against the two restrictions outlined in Chapter 3:

Restriction 1: Agents are restricted from grossly assaulting an innocent person, even when doing so would produce the best state of affairs, impersonally judged.
Restriction 2: Agents are restricted from killing an innocent person, even when doing so would produce the best state of affairs, impersonally judged.

More specifically, I aim to defend these two restrictions from the "charge of paradox". The charge of paradox is an argument put forward by consequentialists against the overall plausibility of restrictions: practical reason suggests that a greater number of overall restriction violations are objectively worse than fewer. Restrictions go against practical reason because they bar agents from minimizing overall total restriction violations. A restriction violation occurs when an agent or agency acts contrary to Restriction 1 or Restriction 2 (i.e., assaults or kills an innocent person). If anti-consequentialists hold that restriction violations are very bad actions, then it is against practical reason to bar agents from minimizing the total overall number of restriction violations. Thus, being against practical reason, restrictions are paradoxical. My response to this argument is to deny that the restrictions asserted by the anti-consequentialist must categorically bar restriction violations. I will introduce a distinction between restrictions on actions—restrictions of a sort that categorically bar action types (restrictions\textsubscript{A})—and restrictions on moral justification—restrictions of a sort that limit consequentialist moral justifications for particular action types (restrictions\textsubscript{J}).
While both forms of restrictions entail that AP is false, only restrictions manage to avoid the consequentialist's challenge.

In order to respond to the charge of paradox, I will closely examine the types of cases that are at the root of the charge of paradox. I will refer to these types of cases as "tragic dilemmas". Tragic dilemmas have the following basic structure: an agent, through no fault of her own, finds herself in a situation where she can violate a restriction and this violation will produce a state of affairs with considerably fewer identical restriction violations, or the agent can do nothing and allow a greater number of equally horrible restriction violations to occur. Confronted with a tragic dilemma, it is plausible for the agent to hold that there is nothing he or she can do that will obviously count as following the practice of treating persons as units of moral concern. In such a case, what ought an agent do? I will argue that in a tragic dilemma an agent is always permitted to kill an innocent just so long as there are compelling agent-relative reasons for doing so. I defend the relevance of agent-relative reasons for solving tragic dilemmas because agent-relative reasons, unlike agent-neutral reasons, capture the intrinsic moral importance of the causal pathways of actions and harms.

4.2 Restrictions and the Charge of Paradox

As we found in Chapter 3, the anti-consequentialist argues in favour of restrictions that entail a denial of AP: restrictions, at least sometimes, bar agents from doing what it takes to bring about the best state of affairs. The consequentialist makes two demands on any defence for such restrictions: (1) The mere fact that such restrictions are intuitively plausible is not enough: restrictions need a principled rationale; (2) Granting that
provision (1) is met, restrictions cannot be contrary to the dictates of practical reason: no acceptable part of a moral theory should lead to paradox. For my present purposes, I will grant the consequentialist the appropriateness of making these demands on any defence of restrictions.\(^{147}\)

In Chapter 3, I set out to identify a plausible rationale for two restrictions and thereby meet the consequentialist's first provision. I argued that any moral theory that adequately acknowledged the separateness of persons must incorporate moral restrictions against assaulting and killing the innocent. Restrictions that protect the essential basis for human persons being separate units of moral concern—our bodies and our lives—take their rationale in the natural separateness of persons. In this chapter, my focus will be on meeting condition (2).

Why are these restrictions paradoxical? Scheffler argues that even if someone offers a plausible rationale for restrictions, they are too strong because they rule out morally justified actions. Consider Bernard Williams' "Jim in the Jungle" scenario:

**Case 3:** Jim is a Canadian citizen on vacation in the jungles of Peru. Jim comes across a small native village that has armed soldiers lining up young men against a brick wall. After Jim's papers are checked, the Captain explains what is going on: in order to stop the native people from protesting against the government, the Captain has decided to execute twenty random native men. However, the Captain is in a charitable mood and makes the following proposal to Jim: if Jim shoots and kills one of the men, then in honour of Jim doing so, the soldiers will allow all the

---

\(^{147}\) Some critics of consequentialism have not been as charitable on this point. For instance, Paul Hurley challenges Scheffler’s unsupported value-neutral framework for assessing the acceptability of a rationale for restrictions: "The demand that a rationale must appeal to the value of states of affairs... is legitimate only within the context of a particular grounding conception of value that is characterized by three central components: (1) an account of the evaluation of states of affairs, according to which states of affairs are appropriately evaluated from the impersonal standpoint; (2) an account of the relation between the evaluation of actions and the evaluation of states of affairs, according to which the former is based entirely on the latter; (3) an account of the evaluations of actions... according to which actions as well are appropriately evaluated from the impersonal standpoint... This conception of value clearly precludes a rationale for traditional agent-centred restrictions." See: Paul Hurley, "Agent-Centred Restrictions: Clearing the Air of Paradox," *Ethics* 108, no. 1 (October 1997): 132.
other natives to go free. The natives appear to understand the proposal, and are begging Jim to accept it. Thus, the dilemma Jim faces is the following: if he kills one of the twenty men standing in front of him, this act would save the other nineteen from being murdered; if, on the other hand, he abstains from killing one random man, all twenty will be murdered.\footnote{J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against* (London: Cambridge University Press, 1973), 98-99.}

I will call the Jim and the Jungle case a "tragic dilemma". An agent finds him or herself in a tragic dilemma when, no matter what the agent does, some person or persons will be assaulted or killed, and the only way to save many innocent persons from being assaulted or killed is for the agent to assault or kill a fewer number of innocent persons. Williams employed this case in an effort to criticize consequentialism. Williams argued that one of the failings of consequentialism is that, according to the consequentialist, Jim does not have a moral dilemma. Williams thinks that it is a serious defect in any moral theory where theory-proponents fail to acknowledge the appropriateness of Jim's experiencing anguish.\footnote{Williams also employs this example in order to highlight the counterintuitive implications of the consequentialist's thesis of negative responsibility, namely, that people are equally responsible for states of affairs they bring about as states of affairs they fail to prevent; I will discuss this issue later in the chapter.}

Samuel Scheffler hijacks the type of example Williams employed in order to defend AP. Scheffler claims that the "Jim in the Jungle" scenario and others like it demonstrate the paradoxical aspect of restrictions: the anti-consequentialist would be forced to prohibit Jim from killing the one aboriginal, even though 19 more persons will be killed as a result of this prohibition. Such a prohibition is paradoxical: if murder is such a bad thing, then why is Jim not permitted to do what it takes to minimize the total number of murders? I might add that Scheffler's analysis has an initial plausibility. In my experience, most people who I have talked with about Case 3 have granted that it would
be permissible for Jim to kill one of the men about to be executed. Many have said that they are not sure what they would do, due to fear and the like, but most grant that the case is such a bizarre and exceptional one that killing one of the men is morally permissible.

Scheffler points out that if the anti-consequentialist thinks that it is so bad for one unjust execution to take place, then it would be rational for them to hold that Jim should try to minimize the total number of unjust executions.\(^{150}\) Regardless of the degree of disvalue that results from Jim killing one random native, in this case *not killing* the one native will bring about greater disvalue overall. Yet, supporters of restrictions hold that minimizing the number of murders is not the right thing to do. This position is odd: one of our primary intuitions is that if one thing is bad, nineteen additional actions of the exact same type are worse. If defenders of restrictions hold that murder is a bad thing and this is *why* we have a restriction against it, then they fall prey to this argument from paradox.

Given my argument for the rationale for restrictions in Chapter 3, it might appear as though I have a reply: If Jim murders one aboriginal man then he would be ignoring the fact that this one man is a *person*. The aboriginal man is a unit of moral concern that deserves special protections from being forced to bear heavy burdens merely for the common good. Jim is barred from killing an innocent person not because of the badness his action brings about but because the nature of the act is *evil*. However, Scheffler could still respond in the same way he did with the appeal to the extreme degree of harm that murder entails. Scheffler can argue that even though murdering the one aboriginal man

will undermine or violate the moral significance of persons, surely not killing the one man who is about to be murdered anyway will result in a greater degree of disrespecting of the moral significance of persons overall.\textsuperscript{151} Scheffler claims that it is irrational not to allow for minimizing the number of such violations.\textsuperscript{152} Even with the rationale for restrictions in mind, it seems that an air of paradox still surrounds restrictions. My goal for the remainder of the chapter is to address the charge that moral restrictions against killing the innocent are paradoxical.

4.3 \textbf{Tragic Dilemmas and Obligation}

Tragic dilemmas beg us to answer the following question: can people intent on violating restrictions morally \textit{obligate} other agents to bring about slightly less evil states of affairs that what would otherwise have taken place? I will not presently answer this question because my goal in this chapter is to focus on moral \textit{permission}, not moral obligation.

In order to side-step the debates involved in addressing the issue of moral obligation, my examination of the Jim in the Jungle case will focus on moral permission by assuming that, all things considered, Jim \textit{is willing} to kill an innocent person (so long as doing so is morally justified). Given this assumption about Jim, I can now turn to the issue of whether or not he is \textit{permitted} to violate the restriction against killing the innocent, and, if so, what morally justifies such an action. In Chapter 5, I will deal more fully with the general issue about the limits of consequentialism \textit{obligation}.

\footnotesize{\textsuperscript{151} Scheffler, \textit{The Rejection of Consequentialism}, 89.}

\footnotesize{\textsuperscript{152} Ibid., 107 and 111-112.}
4.4 Nagel on Agent-relative Reasons and Restrictions

As a first step toward addressing the consequentialist’s charge of paradox, I will examine some of Thomas Nagel’s work. Nagel thinks that agents have two different types of moral reasons: agent-neutral reasons and agent-relative reasons. Agent-neutral reasons apply to all: everyone has equal reason to, e.g., ameliorate human suffering. Agent-relative reasons are different from agent-neutral reasons because agent-relative reasons make essential reference to a particular agent. For instance, the fact that Mr. Smith’s son is suffering presents Mr. Smith with a reason for helping him, above and beyond the agent-neutral reason for limiting suffering in general. Nagel’s view is that moral dilemmas are often challenging because these two types of reason come into direct conflict.

Consider the following case: suppose you encounter a car-accident on a lonely country road.\(^{153}\) The three occupants of the car all have sustained life-threatening injuries and will surely die if they are not brought to a hospital. Luckily, there is a nearby farmhouse and you run to the door screaming for help. An elderly woman and her grandson are in the farmhouse; however, because of your screaming and yelling, the old woman fails to understand the nature of the situation and refuses to help. In the first version of this case, suppose you see a phone on the kitchen counter. While you may have agent-relative reasons against breaking and entering (you do not want to be the kind of person that breaks into homes), there are compelling agent-neutral reasons to smash the window in order to call for help. In the second version of this case, suppose that the

---

\(^{153}\) Nagel, *The View From Nowhere* (New York: Oxford University Press, 1986), 176. I have changed the example slightly.
farmhouse does not have a phone, but does have a vehicle in the driveway. In a panic, the woman locks herself in a bedroom. After pleading with the woman for her car keys, you realize that it is no use: the car-accident victims are going to die unless you do something drastic. It occurs to you that if you twist the grandson’s arm, perhaps his cries of agony will make the woman listen to reason. While the agent-neutral reasons in favour of saving the three accident victims remain the same, the agent-relative reasons have become much weightier: deontological restrictions bar you from torturing a child.

Both versions of Nagel’s case represent an extreme situation: the lives of three strangers hang in the balance. What is of interest for my present purposes is the radically different nature between the first version of the case and the second. It is commonly thought that there is a major difference between breaking into a house in order to use the phone, and torturing a child. While both actions—arguably, at any rate—would bring about the best state of affairs, impersonally judged, the agent-relative reasons against torturing children are commonly judged to be much weightier compared to the agent-relative reasons against breaking-and-entering. According to Nagel, one problem with an impersonal morality like consequentialism is that it overlooks the significance of our commonsense view about the weight of some agent-relative reasons.

Yet, as plausible as our commonsense views are, Nagel agrees that they have a paradoxical aspect to them. While it is intuitively plausible to support deontological restrictions that protect agents from certain types of harms, such restrictions resist any simple agent-neutral reduction. Furthermore, Nagel holds that deontological restrictions do not simply bar certain events from happening but also represent agent-relative reasons against individuals doing certain things. Yet, as Nagel continues,
it is hard to understand how there could be such a thing. One would expect that reasons stemming from the interests of others would be neutral and not relative reasons. How can a claim based on the interests of others apply to those who may infringe it directly or intentionally in a way that it does not apply to those whose actions may damage the same interests just as much indirectly? After all, it is no worse for the victim to be killed or injured deliberately than accidentally... a murder does not seem to be a significantly worse event, impersonally considered, than an accidental or incidental death. Some entirely different kind of value must be brought in to explain the idea that one should not kill one person even to prevent a number of accidental deaths: murder is not just an evil that everyone has reason to prevent, but an act that everyone has reason to avoid.\(^{154}\)

What produces the paradoxical nature of deontological restrictions is the “collision between the subjective and objective viewpoints.”\(^{155}\) Impersonally considered, cases arise where the violation of deontological restrictions will make the world a better place. On the one hand, Nagel does not think that the impartial standpoint dominates the partial one: “the dominance of this neutral conception of value is not complete.”\(^{156}\) On the other hand, seen from the impartial perspective, deontological restrictions appear “primitive, even superstitious.”\(^{157}\)

Nagel does not think there are easy and decisive answers about tough moral dilemmas. If we are to be guided by the objective self, we will be guided to do what is evil in order that fewer evil acts overall occur; if we are to be guided by the subjective self—the self that is embedded within a world and comes from a particular point of view—the decision is “not merely choices of states of the world, but of actions.”\(^{158}\)

Nagel is clear that the internal, or agent-relative, perspective needs to be examined in

\(^{154}\) Ibid., 178.
\(^{155}\) Ibid., 180.
\(^{156}\) Ibid., 181.
\(^{157}\) Ibid., 181.
\(^{158}\) Ibid., 183.
order to understand our intuitions about deontological restrictions. Yet, he is also clear that an examination is not a defence of the correctness of these intuitions.

I agree with Nagel that in the case he examines there exists a very strong agent-relative reason for not torturing the child, even though harming the child is assumed to be the most efficient means of maximizing overall agent-neutral value. In addition to being useful for thinking about moral dilemmas in general, I think Nagel's distinction between agent-relative and agent-neutral reasons is useful for thinking about tragic dilemmas like the Jim in the Jungle case. In a tragic dilemma—a situation where there is nothing an agent could do that would obviously count as acknowledging the practice of treating persons as units of moral concern—I will argue that agent-relative reasons may justify killing an innocent person, while agent-neutral reasons, on their own, fail to justify killing an innocent person.

4.5 Two Accounts of Restrictions: RestrictionsA and RestrictionsJ

Let's consider the example of a restriction on killing innocent persons. On this example, I will hold that an agent commits a "restriction violation" if he or she kills an innocent person. There are several conceivable ways that one may characterize why a restriction violation is prohibited. While the traditional view of restrictions is that they bar people from committing certain kinds of actions, I will also consider the possibility that restrictions limit consequentialist justifications for certain kinds of actions.

Restrictions on actions (restrictionA) directly bar agents from performing a restricted act. Restrictions on justification (restrictionsJ) indirectly bar agents from committing a restricted act because restrictionsJ limit acceptable moral justifications for the restricted
act. More specifically, restrictions\textsubscript{j} deny that agent-neutral concerns (on their own) offer an adequate moral justification for violating a restriction. Restrictions\textsubscript{j} leave room for the possibility that it is morally permissible for an agent to commit a restriction violation (i.e., kill an innocent person) because of some consideration in addition to agent-neutral concerns. Ultimately, I think restrictions\textsubscript{j} are the more plausible construal of restrictions.

Both restrictions\textsubscript{\textnormal{A}} and restrictions\textsubscript{j} are limitations on the production of the best state of affairs. Both accounts of restrictions represent a denial that every action can be justified solely in virtue of the consequences that the action will bring about. Cases 1 and 2 from Chapter 3 represent plausible instances where killing one innocent person will bring about the best state of affairs.\textsuperscript{159} If there are no restrictions on killing the innocent, then it would be morally permissible for the lawyer in Case 1, and the doctor in Case 2, to kill the innocent. Defending either restrictions\textsubscript{\textnormal{A}} or restrictions\textsubscript{j} is one way to deny this entailment. I will examine both cases in turn.

Restrictions\textsubscript{\textnormal{A}} bar the lawyer in Case 1, and Sam in Case 2, from killing innocent persons because restrictions\textsubscript{\textnormal{A}} categorically forbid killing innocent persons under any circumstances. This result is consistent with my claims in Chapter 3 regarding the role of restrictions. Sometimes restrictions are—from an impersonal perspective—inefficient. Case 1 and Case 2 represent situations where acknowledging the moral significance of the separateness of persons has an impersonal cost, but it is a cost we must bear if we wish to treat each person as a unit of moral concern.

Restrictions\textsubscript{j}, like restrictions\textsubscript{\textnormal{A}}, bar the lawyer and Sam from killing the one person. In Case 1 and Case 2, the only plausible moral justification for killing the innocent is the impersonal value the killing will produce. Restrictions\textsubscript{j} limit moral justifications for

\textsuperscript{159} Case 1 appears on page 66 (Chapter 3, section 3) and Case 2 appears on page 78 (Chapter 3, section 5).
killing the innocent based on agent-neutral concerns only; as such, both the lawyer and Sam may not kill. Restrictions, just like restrictions, are consistent with the practice of acknowledging the moral significance of the separateness of persons for which I have argued in Chapter 3.

It should be emphasized that Case 1 and Case 2 have important distinguishing differences from Case 3 because Case 1 and Case 2 can be solved by actions that are consistent with the practice of treating persons as units of moral concern (namely, by simply not killing an innocent person). While some people in Case 1 will not receive some needed financial assistance, and some people in Case 2 will die because of organ failure, it is not true that these people have been burdened by anyone. While it is an unfortunate event to not receive needed financial assistance, and it is an extremely unfortunate event to die because of organ failure, our practice of treating persons as units of moral concern is not a practice designed to maximize overall agent-neutral value. Rather, as I argued in Chapter 3, the function of this practice is to ensure that people are not grossly burdened (e.g., assaulted or killed), even when doing so would be for the greater good.160

My goal in what follows is to defend the rejection of AP, while simultaneously granting that some tragic dilemmas represent situations where it is morally permissible to kill an innocent and thereby violate a restriction. We are finally in a position to apply my

160 My argument, as defended in Chapter 3, supported the claim that a human morality ought to include the practice of treating persons as units of moral concern, where being treated in such a way entailed that innocent persons are not grossly burdened (assaulted or killed), even when such acts would produce the best state of affairs. Consistent with this claim is that we have other moral norms that serve the function of increasing overall agent-neutral value (e.g., moral requirements of aid, generosity, etc.). Usually these two sources of moral norms are not in conflict. However, when these norms do come in conflict, an appropriately human morality holds that our practice of treating persons as units of moral concern has precedence over these other agent-neutral concerns (hence the ‘even when such acts would produce the best state of affairs’ clause); see Chapter 3, 75-77.
distinction between restrictions_A and restrictions_J to tragic dilemmas like Case 3. Tragic dilemmas are especially acute moral dilemmas because action and inaction both result in restriction violations. The case is tragic not only because many people will be killed at the hands of others but also because there is no action the agent facing the dilemma could perform that is obviously consistent with following the practice of treating persons as units of moral concern. If the agent kills one person then the agent has failed to acknowledge the moral significance of the fact that the victim was a separate unit of moral concern. If the agent does not kill one person, then the agent permits 19 additional people to be treated in the exact same manner that the one is treated. A morally conscientious person would experience anguish about what he or she should do.

Suppose Jim kills the one aboriginal. The consequentialist will hold that Jim’s action is justified on the grounds of the great good the murder brought about. The defender of restrictions_A would hold that Jim’s action is wrong because he violated the restriction_A against killing an innocent person. The supporter of restrictions_J would hold that Jim has violated the restriction on killing innocent people, but done so permissibly so long as some agent-relative features permit Jim to kill. What agent-relative justifications for killing the innocent does Case 3 present? Jim is permitted to shoot one person in this case because the twenty potential victims are all “begging Jim to accept the offer”. Furthermore, Jim might find it easier to live with himself if he kills the one, rather than walking away. These agent-relative reasons for killing the one innocent person render Jim’s action a permissible one.161

161 It is worth noting that different agent-relative reasons may offer Jim the permission to abstain from killing the one (e.g., supposing Jim is a pacifist). I will defend the claim that agent-relative reasons offer agents moral justifications for acting sub-optimally in Chapter 5.
 Unlike restrictions, restrictionsA have the effect of categorically barring Jim from killing one of the people about to be executed. I agree with Scheffler that restrictionsA have an air of paradox about them. When an agent is presented with a tragic dilemma, I think it is paradoxical to hold that it is categorically impermissible to violate restrictions. Even though Case 3 has the appearance of being a tragic dilemma, supporters of restrictionsA see nothing of the sort: killing the innocent is categorically barred. Scheffler has made the case that restrictionsA are paradoxical, and I agree. Tragic dilemmas bring the paradoxical nature of restrictionsA to the foreground. Due to this paradox, restrictionsA should be rejected.

Consider a permutation on the Jim in the Jungle scenario. Suppose everything about the case is the same, with the exception that the captain has asked you to kill yourself in order to save the twenty aboriginal men. Any moral theory that incorporated restrictionsA on killing the innocent would be committed to morally barring you from deciding to take your own life in this situation. The defender of restrictionsA cannot now plead special circumstances and permit self-sacrifice; such permission is purely ad hoc. I can imagine some staunch defenders of restrictionsA might find this entailment of restrictionA fitting. I can also imagine, however, that most supporters of restrictionsA would agree that restrictions that always and categorically bar self-sacrifice are too strong.

---

162 Some defenders of deontological restrictions claim that the restrictions have thresholds: when enough lives are at stake, we ought to revert to a consequential analysis, e.g., Jeff McMahan, The Ethics of Killing: Problems at the Margins of Life (Oxford: Oxford University Press, 2002). Such “restrictionsA with thresholds” sidestep the charge of paradox, but they do so at the cost of failing to outright deny AP. As I argued earlier when I discussed Pettit’s account of restrictions, I do not think a defence of restrictionsA with thresholds would meet the original demands of the anti-consequentialist.

163 I think this permutation on the Jim in the Jungle case represents an especially acute problem for the defender of restrictionsA on killing the innocent because defenders of restrictionsA often wish to defend the
When confronted with a tragic dilemma, restrictions\textsubscript{I}, unlike restrictions\textsubscript{A}, do not necessarily bar agents from assaulting or killing the innocent because it is an open question whether a plausible agent-relative moral justification renders the assault or killing morally permissible. Restrictions\textsubscript{I} manage to avoid paradox and yet still entail the denial of AP.

4.6 Value Promotion and Theoretical Parsimony

In all likelihood, consequentialists will not be persuaded by the distinction between the two types of restrictions I have suggested. For instance, Pettit thinks that one of the virtues of consequentialism is its unified response to value: while the consequentialist says all values ought to be promoted,\textsuperscript{164} my position entails that only some values ought to be promoted. While restrictions\textsubscript{I} have managed to avoid paradox, they have done so at the cost of simplicity and theoretical parsimony. The introduction of agent-relative reasons adds complexity where it is not needed. The consequentialist reply to my analysis of Case 3—claiming that Jim is permitted to kill the one man on grounds that agent-relative reasons offer moral permission for killing the innocent—is the following: why have both agent-neutral and agent-relative reasons that justify choice instead of only agent-neutral reasons, like the consequentialist? Consequentialism unifies our response to value. By virtue of theoretical parsimony, consequentialism is the superior account. The consequentialist can simply insist on re-interpreting any agent-relative justification in permissibility of moral options for self-sacrifice (a licence for agents to act altruistically, even when the altruistic act is non-optimal). I will examine options more thoroughly in Chapter 5.

consequentialist terms. Permitting the killing in Case 3 brings about the best state of affairs; ultimately, the consequentialist will claim, this fact provides the true justification for Jim.

While I agree that the consequentialist will be able to re-interpret any agent-relative justification for killing in consequentialist terms, I do not think that the mere presence of this re-interpretation represents a demolishing critique. We have two conflicting hypotheses for the judgement of moral permissibility in Case 3: (1) The Consequentialist Hypothesis that says Jim is justified in killing because his action, among those available, promotes impersonal value; and (2) The Agent-Relative Hypothesis that states that Jim is justified in killing because of (a) the heinous nature of the circumstances he faces; (b) Jim feels as though he has to do something—he could not live with himself otherwise; and (c) the twenty men about to be executed all want him to do it; and maybe other agent-relative reasons. The consequentialist can insist as follows: The Consequentialist Hypothesis is parsimonious whereas the Agent-Relative Hypothesis is merely a hodge-podge of various agent-relative reasons. The Consequentialist Hypothesis is the more coherent analysis of Case 3 and thus should be preferred.

My reply to this counter argument is that the Consequentialist Hypothesis’ parsimonious nature implies a generality that should be seen as having a very high cost—unless one is in favour of radical and sweeping moral reform. For example, if we permit Jim to kill an innocent because the killing promotes impersonal value, then we should—by parity of reasoning—permit Sam, the doctor from Case 2, to kill his current patient in order to redistribute organs. While people may experience anguish about what to do in tragic dilemmas like Case 3, most people do not feel as much distress about what is the
right course of action in Case 2. Furthermore, the Consequentialist Hypothesis makes Case 3 look like an *easy* case. Due to the tragic nature of Case 3—the fact that many restriction violations will occur no matter what the agent in the dilemma does—perhaps the Agent-relative Hypothesis’ highly specific nature is a virtue and not a vice. While the focus on agent-relative reasons may be judged to be inefficient on agent-neutral grounds, such a criticism merely begs the question against the defender of the Agent-relative Hypothesis.

We might seem to be at an impasse: how can we decide between the Consequentialist Hypothesis and the Agent-relative hypothesis? In order to break this stalemate, I will (1) consider independent reasons for rejecting the Consequentialist Hypothesis, and (2) consider independent reasons for affirming the Agent-relative Hypothesis. I will consider (1) and (2) in turn.

In an effort to reject unbridled value promotion, consider the following case invented by Cora Diamond.¹⁶⁵ Imagine you could save one of two people from drowning. The only detail that you know about the drowning victims is that one of them is missing a right leg and the other is able-bodied. Diamond holds that “many contemporary versions of utilitarianism imply that it would normally be wrong for you to choose to save the one-legged person.”¹⁶⁶ Utilitarianism offers a ludicrous moral analysis of an agent’s choice of either saving a disabled person or an able-bodied person:

To choose to save the one-legged person is not just wrong, but it is wrong for almost the same reasons for which it is wrong to cause a person to lose his leg in the first place... If you choose to save the one-legged person rather than the two-

---


legged, the number of legs in the world (or, better, legs attached to people) is down by one in comparison to what it would have been had you chosen differently.\textsuperscript{167}

There are a host of interesting issues that Diamond's criticism of utilitarianism highlights: the moral difference between doing and allowing, actions and inactions and agent-caused suffering and non-agent-caused suffering. I will not focus on Diamond's analysis of what we ought to do in this case, but instead on her critique of how a \textit{consequentialist} would analyse this case.

A consequentialist such as Parfit openly acknowledges that his moral theory fails to view persons as units of moral concern: "When we are trying to relieve suffering, neither persons nor lives are the morally significant unit."\textsuperscript{168} One consequence of this failure is that different persons will be worth more or less, impersonally judged, based on the instrumental value the person's continued existence has from an impersonal perspective. To consider a particular form of consequentialism, most utilitarians would claim that a "disabled person loses less when he loses his life" compared to an able-bodied person.\textsuperscript{169} From the perspective of a utilitarian, the presence of a disabled person brings about states of affairs with "greater frustration, less pleasure, more pain and so on", compared to the presence of an able-bodied person.\textsuperscript{170} Therefore, when a person has the choice to cause

\textsuperscript{167} Ibid., 150.


\textsuperscript{169} Diamond, "How Many Legs?" 154.

\textsuperscript{170} Ibid., 150.
an able-bodied person to continue to exist or a disabled person to continue to exist, the person should—all else being equal—choose to cause the able-bodied person to exist.171

Diamond's argument offers good grounds to deny value promotion as the sole goal of morality. While it might be true that more impersonal value is lost when the able-bodied person is not saved (and of course such a claim is contentious and rests on some very controversial assumptions about how the impersonal value is measured), it is far more clear that both persons will lose the same thing if they are not saved—their life. Both able-bodied and disabled people are equally deserving of being saved because both of their lives will be equally ended if they are not saved. Focussing on the impersonal value of saving one person over the other is to fail to view each person as a unit of moral concern and thereby acknowledge the moral significance of the separateness of persons.172

Diamond's case, and her criticism of the utilitarian's analysis of it, offers some independent grounds for rejecting the Consequentialist Hypothesis. Simply put, in cases where lives hang in the balance, it is implausible to hold that the promotion of value ought to be the sole focus of moral concern. In addition to being concerned about impartial value, we also are concerned with treating each person as he or she deserves and the intrinsic value of some types of actions. A value promotion interpretation of tragic dilemmas is cast in doubt because it would either deny these other concerns are morally appropriate or else reduce their significance to agent-neutral value. Furthermore,

171 Diamond cites Parfit, Reasons and Persons, 358-361, as an instance of a contemporary version of utilitarianism that entails such a view.

172 I will return to the issue of how the disabled tend to be viewed by supporters of a consequentialist moral theory in the Appendix (Section 3).
unlike a value promotion interpretation, Diamond’s point of view sheds light on (and helps explain) the commonsense judgement that the lawyer (Case 1) and the doctor (Case 2) ought not kill the innocent. In light of Diamond’s analysis, the Consequentialist Hypothesis has grounds to be rejected.

4.7 Agent-relativity and the Intrinsic Moral Significance of Causal Pathways

In order to complete my defence of the Agent-relative Hypothesis, I will offer some independent evidence in its favour. Consequentialism is a “purely recipient-oriented” moral theory. A purely recipient-oriented moral theory is one that focuses on how recipients of the moral theory fare under the system. Thomas Pogge has recently been highly critical of any moral theory that adopts a purely recipient-orientated approach to morality.\textsuperscript{173} Acknowledging the moral significance of agent-relative reasons for solving tragic dilemmas allows a moral theory to be more than a purely recipient-orientated moral theory.

Many recent papers and books have been written on the moral differences and similarities between killing and allowing to die, killing and withdrawing aid, and killing as an intended consequences and killing as a foreseen but unintended side effect.\textsuperscript{174} All of these debates focus on examining the moral significance of different causal pathways of actions, especially those actions that produce harm. The Agent-relative Hypothesis


gains support because it can be seen as a rational response to the fact that we care intrinsically about causal pathways.

From the perspective of a recipient of a consequentialist moral theory, the causal pathways that bring about benefits or harms are of mere instrumental significance: *qua* a recipient of the good, an agent can be equally harmed by a sly assassin or a large meteor. From a different perspective—focussing on agents *qua* beings who initiate actions on the basis of agent-relative reasons—the causal pathways of various harms are taken to be of the gravest of moral importance. From the perspective of aiming to live in a just society, for instance, the moral significance of the causal pathways of actions, especially harmful ones, are essential to making judgements between guilt and innocence, responsibility and irresponsibility, and unfortunate event and moral catastrophe.175

On the one hand, the Consequentialist Hypothesis claims the only relevant type of reasons for solving tragic dilemmas are agent-neutral reasons. On the other hand, the Agent-relative Hypothesis acknowledges the special moral significance of different causal pathways that produce harm. From the first person point of view, we do not simply care about *what* states of affairs are brought about, but *who* does *what*. For instance, consider the following agent-relative reason for killing the one man: all 20 aboriginals are begging Jim to kill one of them. The fact that the victim’s wishes figure in the causal pathway of Jim’s action is surely of moral significance. In the event that Jim shoots one of them, the fact that Jim kills an innocent person who strongly expressed the preference for taking the risk of being killed—instead of the certainty of being

175 Catherine Wilson, *Moral Animals: Ideals and Constraints in Moral Theory*, (Oxford: Clarendon Press, 2004), 169-170. On the topic of moral culpability and moral disasters, Wilson writes, “An earthquake is not a *moral* disaster unless someone stood to benefit from shoddy building practices, or some persons were able to help the victims, but had self-interested reasons for not doing so”, 169; emphasis added.
killed—is morally relevant. The Agent-relative Hypothesis offers a framework for acknowledging the intrinsic moral relevance of causal pathways that lead to the action that solves a tragic dilemma.

Undoubtedly, the consequentialist will deny the intrinsic significance of causal pathways. Surely, it will be granted, we do take the causal pathways of actions to be of moral significance, but so what? The fact that we do something cannot be taken as evidence that we ought to be doing it that way. The consequentialist is often a moral revolutionary: if consequentialism ineluctably demands changes to our current moral practices, no matter how sweeping, the changes must be embraced.

I reject this reply because it is susceptible to a reductio ad absurdum. For instance, among the moral catastrophes of the 20th century are the Nazi Death Camps, the Rape of Manchuria, the Soviet gulags, and the firebombing of Japanese cities during WWII. If the consequentialist is correct to reject the commonsense moral difference between human deaths with a causal pathway that involved the deliberate and intentional actions of humans and deaths with a causal pathway that do not involve the deliberate and intentional actions of humans, then we should add a diverse list of events to this list of moral catastrophes: the influenza outbreak of 1917-18, the invention and production of the automobile and the mass production and distribution of alcohol. Although these later three occurrences all involve many millions of deaths (and many undeserving victims), the causal pathways that brought about these harms are not of a type that makes these events moral catastrophes. Although some car accidents and some alcohol-related deaths are morally inappropriate (and some cases of people contracting influenza are as well, e.g. when an agent fails to follow quarantine protocol), surely the vast majority of such
deaths are not. It is false that these events are moral catastrophes. There is no agent who imposes the burden of influenza, car accidents and alcoholism on anyone, even if it is true that in some of these cases various agents and agencies permitted or allowed other agents to bring about such bad states of affairs. Even though deaths by car accident, influenza and alcoholism are negative events, the majority of such deaths are morally faultless deaths. Sometimes bad things happen (dying as a result of influenza, car accident, or alcoholism) and there is no overarching agent at moral fault. While preventing morally faultless deaths is a good thing to do, this fact does not entail that we should be willing to kill people to promote this good. After all, barring a moral disaster in our lives, we are all going to die a morally faultless death.

Does my reductio beg the question? Perhaps it could be argued that I am making assumptions about when, in fact, a person plays an immoral role in causing harm to others. The consequentialist’s position is simply that an agent plays an immoral role in e.g., the death of someone else if and only if the agent could have prevented the death and it would have been optimal, impersonally judged, to do so. On my account, when does a person play an immoral role in e.g., the death of a person?

The clearest case is when one agent e.g., stabs another agent to death. An obvious case when no person played any role in someone’s death is when a large meteor strikes and kills someone. Are there some grey cases? Yes. When a hurricane makes landfall on a very poor island, thousands of the island’s inhabitants die from mudslides and wind-blown debris. However, if the island’s inhabitants had have been able to afford more protective shelter, a far smaller number of deaths would have occurred. It could be argued that at least some of these deaths are the result of the negligence of agents (e.g.,
corrupt agents of the local government, etc.), and thus persons played an immoral role in the deaths of some of these people.

My quick response to a case of this sort is that, yes, it is likely that some people played a morally culpable role in bringing about the deaths of innocents (e.g., a corrupt local official taking bribes from lax building contractors). Cases of a similar sort are many: famines, disease, and other problems associated with allocating scarce resources all show that the line I wish to draw between deaths caused by persons for which the person is morally at fault (e.g. Agent A stabs agent B to death) and deaths caused by natural phenomenon with some human involvement (hurricane, famine, etc.) is a distinction that is not easily accomplished.

Consider a non-obvious grey case: building a large bridge. Statistically speaking, we know that it is possible that several workers will be seriously injured or even killed over the course of the bridge’s construction. These deaths represent a burden that a few people pay so that many millions will have the mere convenience of a shorter travel time between Point A and Point B. Furthermore, these foreseeable deaths are not naturally caused, but—ultimately at any rate—are brought about by human agency (e.g. the constituents who voted for the bridge to be built). Do restrictions rule out dangerous mega-projects like building a bridge? No, for the workers—at least for the most part in the Western world—autonomously choose to incur the burden that building the bridge entails: the risk of death is not imposed on them by other agents. If it is the case that the workers are, literally or practically speaking, forced to work on a project that entailed serious risk of death, then a good case could be made that restrictions are being violated.
Obviously, it is beyond the scope of my project to offer a complete and exhaustive account of the difference between morally faultless deaths and morally blameworthy deaths. While I have admitted that there are grey cases where it is not entirely clear whether a particular death is a morally blameworthy death or not, all I need to address the consequentialist’s charge that I have begged the question is that there are at least some deaths that could have been prevented which are nonetheless morally faultless deaths. The examples I considered in my original reductio argument—many millions of deaths caused by influenza, alcoholism and auto accidents—are just such examples. I think it is absurd to claim that all of these deaths are necessarily morally blameworthy deaths merely because various agents and agencies did not do everything possible to limit the total number of these deaths.176

Consequentialism, being a goal-based moral theory, fails to acknowledge the intrinsic moral significance of some causal pathways over others. Consequentialists refuse to acknowledge agent-relative reasons for or against a particular agent doing or not doing some action. The consequentialist simply insists that permitting agent-relative reasons should be abandoned on the grounds that it does not promote agent-neutral value. However, insisting on the promotion of agent-neutral value is merely to beg the question. Accepting the appropriateness of agent-relative reasons can be seen as a rational response to the fact that, from the agent-relative point of view, we care intrinsically about some causal pathways over others. Confronted with a tragic dilemma, an agent should be permitted to acknowledge the relevance of agent-relative reasons.

176 Ibid., 170. Wilson emphasizes that it is a contingent feature (and not a necessary feature) of our world that natural disasters are moral disasters because there is frequently “a clever or conniving human being or two, operating at the expense of a weaker party” (170).
I have argued that the way to solve tragic dilemmas is to admit that killing the innocent may well be permitted in these horrific cases. However, I have denied that the states of affairs that the restriction violation brings about offers a complete moral justification for the violation; rather, I have argued, that plausible agent-relative justifications are necessary for killing the innocent. I have then defended the overall plausibility of agent-relative reasons from the consequentialist's charge that incorporating them lacked theoretical parsimony by suggesting they represent a rational response to the fact that we take certain causal pathways to be of intrinsic moral importance.

4.8 Conclusion

Perhaps surprisingly, both the supporter of restrictions\textsubscript{A} and the consequentialist have something in common: they view tragic dilemmas like the Jim in the Jungle case as moral situations with obviously correct answers—they just disagree about what this answer should be. Restrictions\textsubscript{J} offer a far less heavy-handed approach when applied to a tragic dilemma. A supporter of restrictions\textsubscript{J} thinks that anyone who approaches tragic dilemmas without an open mind is failing to appropriately respond to the morally horrific nature of these cases.

AP states an agent is always permitted to bring about a state of affairs simply because it is the best state of affairs, impersonally judged. Restrictions\textsubscript{J} deny AP because they represent a denial of the claim that all actions that produce the best state of affairs are always permitted. Furthermore, restrictions\textsubscript{J} avoid paradox because they allow agents to acknowledge agent-relative reasons in favour of solving tragic dilemmas by killing the innocent. Such a claim is consistent with the denial of AP because the action
is justified by agent-relative reasons and not justified solely on the grounds of the states of affairs the killing brings about.
5 AGENT-RELATIVITY, HUMAN PARTIALITY AND OPTIONS

5.1 Introduction

In Chapter 4 I argued that agent-relative reasons are relevant to solving tragic dilemmas. My overall goal in this chapter is to defend the relevance of agent-relative reasons to moral justification more broadly. An agent-relative reason for some action is a reason that stems from our natural partiality because it makes essential reference to a particular agent: "I saved him because he is my son", or "I did not kill her because I am a pacifist" are examples of agent-relative moral justification. An agent-relative moral justification is contrasted with an agent-neutral moral justification—a moral justification that does not make reference to any specific agent—such as: "I saved him because doing so will produce the best state of affairs."

More specifically, I will defend what Shelly Kagan calls an "option"—a licence to perform impersonally judged non-optimal act—on the grounds that human persons are partial. Our natural partiality should not be denied or lamented because human persons value the world around them from the first person point of view. We value the world as we do because humans are separate from each other. Each person has his or her own individual perspective: we value our own plans, commitments and projects from our partial, or agent-relative, point of view. Our nature to be partial is a feature of human persons that needs to be appropriately acknowledged by a moral theory and not ignored.

177 Samuel Scheffler, The Rejection of Consequentialism: A Philosophical Investigation of the Considerations Underlying Rival Moral Conceptions, rev. ed. (New York: Oxford University Press, 1995). Scheffler refers to ‘options’ as ‘agent-centred prerogatives’ and I will use these two terms interchangeably, unless otherwise noted.
or lamented. Incorporating the option for agents to act non-optimally represents a way to appropriately acknowledge this feature of persons.

Here is an overview of this chapter: I will critically examine several attempts to defend the acceptability of moral justifications that stem from our natural partiality. First, I consider—but ultimately reject—the consequentialist attempt to justify human partiality impartially. I then turn to an examination of Scheffler’s “agent-centred prerogatives” and Catherine Wilson’s preference based approach for defending the existence of, and limiting the extent of, agent-relative moral justifications. I end the chapter by arguing that Wilson’s strategy is superior because her approach justifies limitations on partiality from within a framework that embraces our natural partiality. However, by way of introduction to agent-relative moral justifications, I will start this chapter by addressing Kagan’s argument that purports to exclude all agent-relative moral justifications.

5.2 Options and Kagan’s Pro Tanto Reason to Promote the Good

In The Limits of Morality, Kagan argues that people are always under the moral obligation to produce the best state of affairs, impersonally judged (so long as what they do is not otherwise morally forbidden). Kagan’s position, if defensible, is directly relevant to the debate in Chapter 4 on moral dilemmas. I ended Chapter 4 by arguing that agents are permitted to solve tragic dilemmas by killing an innocent person, so long as plausible agent-relative reasons justify the killing. Such a solution to tragic dilemmas manages to deny AP and avoid the charge of paradox against restrictions. However,

178 AP is the “always permitted” claim: an agent is always permitted to bring about certain states of affairs solely on the grounds that they are the best states of affairs, impersonally judged.
this permission entails a licence for agents to act non-optimally; Kagan’s position that we are always under the moral obligation to produce the best state of affairs rules out my solution to tragic dilemmas. Kagan’s position needs to be addressed. Second, addressing Kagan’s claim about moral obligation is a necessary first step in defending my claim that consequentialism offers an account of moral justification that is too narrow for human persons: consequentialism fails to acknowledge moral justifications that stem from our natural partiality and thus it has grounds to be rejected as an appropriate account of human moral justification.

Kagan holds that all persons must always do the acts that are reasonably thought to be impersonally optimal. Let’s examine Kagan’s position:

Morality requires that [I] perform—of those acts not otherwise forbidden—that act which can be reasonably expected to lead to the best consequences overall.

...To live in accordance with such demands would drastically alter my life. In a sense, neither my time, nor my goods, nor my plans would be my own. On this view, the demands of morality pervade every aspect and moment of our lives—and we all fail to meet its standards. This is why I suggested that few of us believe the claim, and that none of us live in accordance with it. It strikes us as outrageously extreme in its demands—so much so that I shall call its defender the extremist. The claim is deeply counterintuitive. But it is true.179

Think back to my solution to tragic dilemmas: suppose Kagan agrees with my argument that killing an innocent person merely to produce the best state of affairs is morally forbidden because doing so would fail to appropriately acknowledge the moral significance of the separateness of persons. Even if Kagan granted the plausibility of my argument, it is at this point that our thinking diverges: although I have argued that agents would have the option to solve a tragic dilemma by either killing or not killing, Kagan’s position is that, so long as the act is not morally forbidden, agents must do the act that is

reasonably held to lead to the best consequences overall. Thus, Kagan would hold that if an agent is permitted to kill when confronted with a tragic dilemma, then he is *obligated* to kill.

Kagan argues that the moral theory I defend—the “moderate” position—is characterized by three moral theses: (1) a *pro tanto* reason to promote the good; (2) constraints (or restrictions) that limit the pursuit of the good; and (3) options that licence agents to act in an (otherwise morally permissible) non-optimal manner. An “extremist”, like Kagan, acknowledges (1), is neutral with respects to (2), and rejects (3). The consequentialist only acknowledges (1) and thus holds that “one ought to perform the optimal act, *simpliciter*”.180 (To complete our lexicon, I should note that the moral “minimalist” rejects all of (1), (2), and (3), e.g., the moral nihilist). Let’s examine (1), (2) and (3) in turn.

According to Kagan, all of the ordinary moral theorist, the extremist, and the consequentialist assent to (1). What is a “*pro tanto* reason to promote the good”? Let’s look at a *pro tanto* reason in general. A *pro tanto* reason “has genuine weight, but nonetheless may be outweighed by other considerations.”181 A *pro tanto* reason is *not* a mere apparent reason. For instance, the fact that Mr. Smith’s fingerprint is found on (what is believed to be) the murder weapon would appear to be a reason for holding that Mr. Smith is the guilty party. But the reason is defeasible. If it turns out that the knife with Mr. Smith’s fingerprint is *not*, in fact, the murder weapon, then the fact that his fingerprint was found where it was found may now, in light of this new information,

---

180 Kagan, *The Limits of Morality*, 8. For the purposes of this chapter, I will usually not need to distinguish between the extremist and the consequentialist: when I use “extremist” I refer to both the “extremist” and the “consequentialist”, unless I specifically state otherwise.

181 Ibid., 17.
become evidence supporting Mr. Smith’s innocence. By contrast, a *pro tanto* reason, unlike a merely apparent reason, maintains its weight even though it may be outweighed by other, contrary reasons.

A *pro tanto* reason in a moral context is of course similar to a *pro tanto* reason in general: it is a standing commitment that is unequivocal in its support (even though, like a *pro tanto* reason in general, it may or may not be decisive.) A *pro tanto* reason to *promote the good* will gain clarity with some examples:

(A) If a child is drowning in front of me, and I can save her by throwing a life preserver, ordinary morality certainly requires me to do so—even though it takes some slight physical effort and my clothes may become soaked.\(^{182}\)
(B) Suppose a building is on fire. Upon entering, I find a child and a bird trapped within. Needing one hand to free a clear path back outside, I can save only one of the two, and I hastily pick up—and escape with—the caged bird.\(^{183}\)

Case (A) is straightforward enough: the good of saving the child represents a reason for the agent to incur some minor hardship. Kagan believes that the ordinary moral theorist would hold that a person is morally obligated to save the child because doing so will promote the good. Kagan takes the ordinary moral theorist’s analysis of Case (A) as constituting evidence for the claim that ordinary morality incorporates (1). Case (B) is a little trickier: Kagan argues that although the defender of ordinary morality may not think it is obligatory for the agent to go into a burning building in the first place, once the agent enters the house, he or she is morally obligated to save the child and not the bird. Kagan interprets the ordinary moral theorist’s judgement about the person who decides to enter the house as being evidence in favour of the ordinary moral theorists’ acceptance of (1). Saving the child instead of the bird promotes the good and it is for *this reason*, according

---

\(^{182}\) Ibid., 3-4.

\(^{183}\) Ibid., 16.
to Kagan, that the ordinary moral theorist holds that the agent ought to save the child and not the bird.

Let's examine (2). Kagan holds that the ordinary moral theorist is committed to constraints on the production of the good. Constraints represent the denial of the claim that agents are always permitted to produce the best state of affairs, impersonally judged, because at least sometimes the only way to produce the best state of affairs will be morally forbidden. The restrictions I have defended on killing and assaulting the innocent in the name of the greater good represent two instances of constraints.

Lastly, let's look at (3). Kagan holds that the ordinary moral theorist is committed to options, namely, the moral licence to refrain from producing the best state of affairs. The ordinary moral theorist holds that a person need not do all that they can to make the world as good as possible; while it would be morally meritorious for an agent to, for example, work tirelessly assisting those in need, it is not morally mandatory to do so. Ordinary moral theorists accept that agents have the option to not produce the best states of affairs, impersonally judged, if the agent so desires.

From among (1), (2) and (3), the most contentious moral thesis is (1). Kagan holds that the acceptance of a pro tanto reason to promote the good is the best explanation for many of our commonsense moral judgements: “if the moderates were asked to account for his judgements [about cases (A) and (B)], the best explanation seems to involve the quite general thesis that one always has a (morally acceptable) reason to promote the good.”\textsuperscript{184} Kagan is careful to emphasize that the defender of ordinary morality is not committed to the claim that the pro tanto reason to promote the good is morally decisive,

\textsuperscript{184} Ibid., 16-17.
but only that it offers unequivocal support, even when the *pro tanto* reason is not morally
decisive (due to the *pro tanto* reason being outweighed). Consider a case relevant to the
debates from Chapter 4:

[Suppose that] killing one innocent person is the only way to save two others from
being killed. The moderate believes that it is forbidden to kill the one; but I take it
that he does not want to say that there is no reason at all to save the two—i.e. that
there is no reason at all that speaks in favour of killing the one. Rather, he
recognizes that there is a reason to kill the one (i.e., that it would save the two).\(^{185}\)

I agree with Kagan that the fact that killing the one person will save two others from
being killed is a fact that offers at least some reason for performing the act.

Kagan argues that the defender of ordinary morality occupies a fractious middle
ground between the extremist position and the minimalist position. On the one hand, the
defender of ordinary morality wishes to uphold our commonsense judgements about, for
example, case (A) against the moral nihilist: the man has a genuine moral obligation to
save the drowning child. On the other hand, the defender of ordinary morality aims to
support options in order to reject both extremism and consequentialism. Kagan’s overall
argument against ordinary morality is that the position cannot be defended against the
simultaneous pressures from the moral nihilist and the extremist. (1) and (3) are in direct
tension: why would a moral theory incorporate *both* a standing reason to promote the
good, and offer an option to refrain from promoting the good? In light of this tension,
the moderate needs to modify or abandon either (1) or (3). When judging the overall
plausibility of (1) and (3), Kagan argues that (1) is a “given” so we should consider
modifying or abandoning (3).

---

\(^{185}\) Ibid., 50.
On what grounds does Kagan hold that (1) is central to ordinary morality? Kagan holds that the best explanation of ordinary morality is that it incorporates (1). In light of the tension between (1) and (3), Kagan’s abductive argument needs to be supplemented with an additional argument showing that there are no alternatives to (1) that are not in tension with (3). However, there are other alternatives, besides (1), for the ordinary moral theorist’s judgements about (A) and (B).

Consider the following more specific explanation of the moderate’s judgement about Case (A): the fact that saving the child will produce good offers the agent a pro tanto reason in favour of saving the child.\(^{186}\) This interpretation of the moderate’s judgement in Case (A) offers a plausible explanation of the moral judgement the defender of ordinary morality offers in this case but does not explain what the agent did in terms of “promoting the good”. The moderate can hold that the agent in case (A) does not so happen to save the child in order to promote the good, but rather so happens to promote the good in order to save the child. For illustrative purposes, consider the following mundane example. Suppose I go for a walk. While I may acknowledge one reason for taking a walk is to wear in my new shoes, it does not necessarily follow that the best explanation of what I am doing is wearing in my new shoes. Perhaps I am relaxing, or getting some fresh air, and these actions have the added side benefit of wearing in my new shoes.\(^{187}\) The fact that saving the child promotes the good may count as a reason for

\(^{186}\) As opposed to explaining the moderate’s judgement by the very general moral thesis (1), namely, a pro tanto reason to promote the good.

\(^{187}\) I am indebted to Scott Anderson for pointing out examples of this type.
saving the child, but this reason is merely one among many. The moderate position need not incorporate (1).

Let’s turn to Case (B). Kagan is correct that the ordinary moral theorist will demand that the agent save the child and not the bird; however, Kagan is wrong that the best explanation of this judgement is that ordinary morality incorporates (1). On my more specific explanation of this judgement, the fact that saving the child promotes the good offers some reason for the agent to save the child and not the bird. Consistent with this more specific explanation of Case (B) is that the agent has other reasons to save the child, namely, the duty to help innocent persons in grave danger. As in case (A), the agent in case (B) so happens to promote the good in order to save the child. The ordinary moral theorist can coherently maintain this judgement about Case (B) without incorporating (1).

According to the ordinary moral theorist, why does the agent in case (B) have to save the child and not the bird, given that rushing into a burning building is supererogatory in the first place? The ordinary moral theorist can answer as follows. Sometimes a person’s voluntary actions generate moral duties that the person would not have otherwise been under. Once the agent in case (B) puts herself into the situation she chose to put herself into, she has a self-imposed duty to save the child that most others would not have.

Consider a parallel example: electronics experts do not have a general moral duty to try to defuse ticking bombs. However, an electronics expert who chooses to undertake special training with the bomb squad acquires this duty that he or she otherwise would not have had. Likewise, the agent in Case (B) has decided to enter a burning building,

---

188 E.g., saving the child is one’s duty; saving the child is benevolent and not cruel; and perhaps others.
and now has acquired the duty to save the child (a duty he or she otherwise would not have acquired).

It is false that best explanation of ordinary morality’s relation to the good is the incorporation of a standing commitment to promote the good when other plausible and more specific characterizations, such as the one I considered, are possibilities. Furthermore, because these other more specific explanations are not in tension with other aspects of ordinary morality (e.g., options), these other interpretations are vastly preferable. Understanding a pro tanto reason as a pro tanto reason to promote the good is a Trojan horse that renders the moderate’s position open to unnecessary internal tension. Kagan’s abductive argument has failed: taking ordinary morality to incorporate a pro tanto reason to promote the good is not the best explanation of our ordinary moral judgements. In light of the failure of Kagan’s abductive argument, it is appropriate to hold that the ordinary moral theorist may best resolve the tension between (1) and (3) by modifying or rejected (1).

I started this current section by pointing out that my solution to tragic dilemmas entailed that agents had the option to refrain from bringing about the best states of affairs overall. Kagan’s position is that agents should always perform those acts that reasonably lead to the best states of affairs, impersonally judged, and thus would rule out my solution to tragic dilemmas. I rejected Kagan’s argument on the grounds that it rests on a dubious abductive claim about what the moderate’s position incorporates: a pro tanto reason to promote the good. Unencumbered by the pro tanto reason to promote the

\[189\] As I will argue in the next section, this inherent weakness in the moderate’s position is one that Kagan proceeds to exploit.
good, the moderate’s reliance on agent-relative reasons to solve tragic dilemmas is not ruled out.

My overall goal for this chapter is not yet complete: I do not wish to merely defend agent-relative reasons from Kagan’s challenge. More broadly, I wish to offer a principled defence for options. In fact, I wish to argue that the consequentialist’s denial of options—defended on the grounds that they are a rational response to the fact that persons value the world from the first person point of view—represent proof that a consequentialist account of moral justification is too narrow of an account of moral justification for human persons. In order to meet this more ambitious goal, I need to carefully examine what is at the root of the problem with a moral theory that fails to incorporate options.

5.3 Defending Options: The Appeal to Costs or the First Person Point of View?

It is revealing to explore the consequentialist’s construal of the moderate’s defence for options. My diagnosis why this defence is not as strong as it could be suggests the type of defence I think will be successful. As I indicated in the introduction to this chapter, I think the best defence for options is to highlight the essential role the first person point of view plays in human values, and then argue that moral theories must include options in order to respond appropriately to this feature of our nature. My explication of the best defence for options is understood most clearly if we first examine some alternatives that fail.

Bjorn Eriksson offers the following account of the moderate’s defence for options. Any moral theory that fails to offer options will be a moral theory that is too costly: from the individual’s perspective, such a moral theory would require that the individual gives
up too much. Let’s label this defence for options the “appeal to costs”. Eriksson, in his book *Heavy Duty: On the Demands of Consequentialism*, argues that what is at the heart of the moderate’s appeal to costs is that any moral theory that denies options (e.g., consequentialism and extremism) demands agents endure hardships that are inappropriately large. Promoting the good is intolerably burdensome: persons are motivated to act in such ways that further their own interests, not the interests of all people. If morality demands that individuals act in an optimal manner then they will often have to forgo the pursuit of their own interests. Eriksson presents the following cases:

1. You have $45 and can (a) spend that sum on a recording of Le Nozze di Figaro, which you admire highly, or you could (b) give the sum to charity;
2. You can either (a) give your child some benefit or (b) give a much greater benefits to some unfortunate stranger.\(^{190}\)

In both cases, the extremist demands you do (b); yet, in both (1) and (2) most people would agree that (a) is morally permissible. In fact, it is exactly cases of the sort represented by (1) and (2) that many non-consequentialists argue that extremism is overly burdensome.

Eriksson argues that what is at the root of the appeal to costs is that consequentialism and extremism allow no room for agents to give special weight to their own concerns.

For a precise formulation, he postulates:

The Special Relevance of Costs Thesis (SRC):
The Costs, \(C\), an agent, \(A\), would suffer if he performed an action, \(a\), is relevant for the question whether \(A\) ought to perform \(a\) in the way that even if \(a\) is optimal, \(A\) may, because of \(C\), be morally permitted to perform an alternative to \(a\) that is

sub-optimal [i.e. produces a state of affairs that contains less than the largest possible amount of impartial value that the agent could have produced.]\(^{191}\)

Eriksson’s SRC principle, if defensible, would offer permission for agents to act sub-optimally and thus avoid the overly burdensome nature of consequentialism. Ultimately, Eriksson’s project is to undermine the acceptability of the SRC, and thereby defend consequentialism from the charge that it is overly burdensome.\(^{192}\) Yet, I do not think Eriksson is correct that his overall strategy has addressed “the best versions of the objection from too heavy demands.”\(^{193}\) Any moral theory that demands that agents make sacrifices for the sake of others will be a costly moral theory. The moderate does take morality to be demanding in some cases, potentially obligating that agents even give up their lives. Eriksson’s focus on the size and quantity of burdens fails to offer much insight into the moderate’s case for options because the moderate does not wish to claim that morality is not—at least sometimes—costly.

Defending the importance of options is not best justified on the grounds of costliness. Humans have many interests, some of which are small and relatively unimportant and others that are large and central to the individuals’ entire life. I will follow Bernard Williams’ lead and refer to these larger interests as “projects”. Projects include such things as long-term relationships, membership in a religion, and various

\(^{191}\) Ibid., 29. Italics added for readability.

\(^{192}\) In Eriksson’s “more constructive” final chapter, he set out to “try to develop some thoughts on what I find plausible in the objection from too heavy demands” (156). Eriksson proposes a scalar account of obligation, one that would “respond to the charge of insensitively demanding that agents perform burdensome actions” (156). Thus, Eriksson’s considered view is that consequentialism would still demand that agents produce the best state of affairs, but the demands would be stronger or weaker for each agent based on how able (in a motivational sense) the agent is to carry out the obligation. While I think it is a mistake to link moral obligation this closely with an agent’s particular motivational set, it is outside the scope of my present task to consider these arguments (i.e. I find it prima facie implausible that people should have prudential reasons to develop miserly characters in order to mitigate the strength of their obligations to help needy strangers).

\(^{193}\) Ibid., 156.
social roles. Obligating persons to act in an optimal manner would all but rule out the possibility of agents adopting and maintaining the vast majority of non-malevolent life-framing projects. For instance, it is likely that almost all deep and meaningful relationships would not be maintained in favour of reading to the blind, assisting the elderly, and working an extra part time job in order to send the money to famine relief. Our loved ones would have to be abandoned if our time could be used more efficiently helping strangers. In short, acting in an optimal manner, as judged from an agent-neutral perspective, would involve abandoning all projects that do not generate as much good as possible.

Williams argues that projects are not merely especially important sources of utility that only need to be properly counted in the impartial calculus. Projects are not a mere means of producing value. The fact that persons value the world around them from the first person point of view is a practical and natural condition for the possibility of value for human persons. As such, obligating agents to promote the good should not be thought of as asking agents to incur a cost in the name of the greater good (in the same vein as asking agents to send money to famine relief in lieu of a night at the movies would represent the agent incurring a cost for the greater good). Focussing on the costs an agent incurs by being asked to promote the good is not the best strategy for defending options because such a focus overlooks the special significance of our long-term projects and commitments.

194 J. J. C. Smart and Bernard Williams, *Utilitarianism: For and Against* (London: Cambridge University Press, 1973), 111. In a discussion of Williams’ views on the value of integrity, Alan Thomas considers the charge that consequentialism represents an intolerable attack on moral integrity. A sufficient reply to this charge is not that an agent’s loss of integrity can simply be factored into the impartial calculus. Such a view takes an agent’s integrity to be an impartial valuable like any other value, instead of being the basis from which all agent values. See: Alan Thomas, “Consequentialism, Integrity and Demandingness” (unpublished manuscript).
Kagan agrees with Williams on the importance of projects in human life. Kagan offers a more subtle account of the appeal to cost compared to that of Eriksson's. Kagan explains that, "Persons have a point of view from which certain objects (the objects of their desires, concerns and interests), take on an importance and are assigned weight disproportionate to the weight an impartial perspective would assign." For example, one of my current projects—my Ph.D. thesis—is one of the most important things in my life. In the short- to medium-term, my life is structured around this project. What makes this activity of mine a project is that I am partial toward it in Kagan's sense. Kagan does not see the moderate as supporting options merely on the general basis of the size and quantity of costs extremist's demand the agent incurs. Rather, Kagan understands the moderate's criticism as focussing on the types of costs that will have to be sacrificed:

If my resources are being devoted to the promotion of the good, then I will generally be unable to favour those things which I might be particularly concerned. These might include my own welfare, or the welfare of various other individuals whom I love or with whom I am friends. They might also include various projects, goals, or other endeavours to which I am committed.... Worse still, in a situation in which I must choose between saving some loved one, or two strangers, I apparently must choose to save the strangers. Kagan is right to focus on the nature of what agents may be obligated to abandon on account of morality that denies options. Such an account of the appeal to costs might appear consistent with Williams' claim that an absence of options entails an absence of projects—an entailment that is so intolerably costly that such a moral theory has grounds to be rejected. Thus, the appeal to costs—when the costs are understood in Williams'


196 Ibid., 233.

197 Smart and Williams, Utilitarianism, 116.
sense—represents a defence for options that permit agents to pursue their life-framing projects.

I think Kagan’s strategy to focus on projects is a good one, but ultimately fails to be the best way to defend options. The best defence of options is not based on an appeal to costs at all. Options are best defended by rejecting the extremist’s understanding of human partiality as an obstacle to morality, in favour of understanding our partiality—the source of projects and commitments—as being a foundation of human morality. Therefore, the best defence of options is to argue that they represent a way for a moral theory to appropriately acknowledge the central importance of projects for human persons and thereby reflect the moral significance of the fact that human persons value the world from the first person point of view. In short, a moral theory for human persons should not advocate that persons begrudge or lament the way we value the world, but should advocate that we embrace the way we value the world.

Consider the following passage where Kagan displays his hostile attitude toward human partiality:

When persons assess events from their point of view, they assign weights to those events in which they take interest out of proportion to the weights which an impartial perspective would assign: for example, the impersonal evaluation tells us that one death is objectively better than two; but the subjective evaluation made from my personal perspective may differ, if I am the one. The weight we assign to an event reflects a bias in favour of the objects of our desires. Matters which are near to us, both psychically and physically, loom large; we consider them more important. And since our reactions reflect our assessment of the various possible outcomes, our actions too typically reflect this bias to what is near to us.198

Kagan’s general characterization as the personal point of view as reflecting a bias implies that our human partiality is something that clouds our judgement; the personal point of

view is something that should be overcome. Why does Kagan adopt such a hostile attitude toward the first person point of view?

Kagan’s negative attitude is entailed by the thought that “the requirement to pursue the good pervades an agent’s entire life—all its aspects, every moment”. The fact that we engage the world via the personal point of view is continuously in at odds with Kagan’s pro tanto reason to promote the good. However, given my challenge to Kagan’s abductive argument, we are left with no good reason for thinking that valuing the world from the personal point of view is morally contentious. Furthermore, in light of Williams’ claims regarding the centrality of projects for human value, we may be inclined to view this feature of person in a much more favourable light.

In contrast with Kagan’s negative attitude toward the personal point of view, Samuel Scheffler, in his book *The Rejection of Consequentialism*, adopts a positive attitude toward this feature of persons. The personal point of view has an independence that needs to be recognized by moral theories. Scheffler argues that a moral theory that incorporate options (what he calls “agent-centred prerogatives”) will “recognize and mirror” the nature of persons far better than the extremist’s position or consequentialism.

Scheffler defends his positive attitude toward the personal point of view as follows:

> To have a personal point of view is to have a source for the generation and pursuit of personal commitments and concerns that is independent of the impersonal perspective. And, it might be said, consequentialism ignores this feature of

---

199 Ibid., 19.

200 As I discussed in the last section, Kagan holds that the moderate is wedded to the pro tanto reason to promote the good on the grounds that the pro tanto reason to promote the good is the best explanation of a wide range of moral judgements the moderate wishes to make.

persons. For it requires each person always to act as if he had no further concern for his projects and plans once the impersonal assessment was in. In other words, it requires that he devote energy to his projects in strict proportion to the weight from the impersonal standpoint of his doing so. We must surely reject any regulative principle for persons which ignores the independence of the personal point of view.

Scheffler thinks that the extremist position fails because it does not recognize the moral significance of the independence of the personal point of view. A form of consequentialism that is supplemented by options—such as Scheffler’s “hybrid consequentialism”—would be a rational strategy for recognizing the moral significance of the fact that persons value the world from the first person point of view. The personal point of view is what allows humans access to projects and commitments. Even though the personal point of view can be a source of bias in some cases, the appropriate general attitude toward this feature of persons is a positive one because it is in virtue of the fact that we engage the world in the way that we do that we care about projects at all. As Williams argued, our projects and commitments make our lives worth living.

Kagan does consider Scheffler’s strategy for defending options, and I will examine why Kagan’s counter-argument fails. Kagan agrees that a moral theory should minimally reflect the morally relevant features of persons: “If there are any features [of persons] which are forced upon a moral system by the recognition of the nature of persons, then clearly any system which lacks those features will also be unacceptable.” Kagan proceeds to argue that extremism does minimally reflect the fact that persons engage the world via the first person point of view because, for instance, this fact “might support structuring rewards and punishments so that an agent can best promote her own interests

---

202 Ibid., 58.

by promoting the overall good". However, my response is to again highlight Kagan’s antagonistic attitude toward human partiality: it is a feature of our nature that moral theory is “forced” to recognize. Given my rejection of the *pro tanto* reason to promote the good, this general negative attitude appears groundless.

Before we turn to the question of what role our partiality should play in moral justification—and the limits on justifications that stem from our natural partiality—I wish to first consider two objections against the defence of options I have so far advanced. The first reply is to challenge the moderate’s claim that projects—and the first person point of view more generally—should be a feature of human persons that we ought to be proud of. I will consider this objection in Section 4. In Section 5, I will consider what I take to be a more feasible criticism of my defence for the moderate’s position, namely, that consequentialism can incorporate the moral significance of the first person point of view because a sophisticated consequentialist could justify partiality impartially. That is to say, a sophisticated consequentialist would recognize that making room for projects is the *means* of attaining the best states of affairs.

### 5.4 Response 1: Eliminating Projects

The extremist might pose an immediate challenge against my defence for options via the moral significance of the personal point of view: the extremist could argue that our partiality to our own projects is based in an uncritical acceptance of selfish impulses. It is undoubtedly true that—in the Western world at any rate—we have been brought up to think that justifications that stem from our natural partiality are morally appropriate; yet,
thinking something does not make it so. While this strategy might sound hopelessly counterintuitive, the consequentialist might remind us of some historical cases that challenge these initial reactions (e.g., in the same way that treating all people as equals may have originally sounded hopelessly counter-intuitive). Impartiality is at the core of consequentialism. Everybody is counted once and only once. Why should an agent count him or herself twice: once impartially and then again from the first person point of view? Consequentialism is a demanding moral theory. The extremist and the consequentialist would hold that projects should be allowed only when they are reasonably expected to promote the good.\(^{205}\) When persons act with partiality they are being selfish and being selfish is almost always not the way to produce the best state of affairs. We ought, morally, to rid ourselves of projects that inhibit the attainment of the best state of affairs, says the consequentialist.

I will not spend too much time responding to this side of the debate, besides pointing out it appears to violate the “ought implies can” principle.\(^{206}\) Surely, people are morally obligated to do only what they are physically and psychologically capable of doing. For instance, no moral theory should fault a person for failing to save a child drowning in a raging, flood-swollen river. On the grounds that no human \textit{could} swim well enough to save the victim, no one is morally obligated to do so. Likewise, if humans cannot help engaging the world in a manner that demonstrates partiality (at least some of the time), demanding that we \textit{never} do so would be for a moral theory to demand the impossible.

\(^{205}\) Kagan writes: “What each agent is required to do is to act in such a way that she can make her greatest possible contribution to the overall good (given her own particular talents). Very likely this involves taking a hard look at her life plans, and reshaping them accordingly...” Kagan, \textit{The Limits of Morality}, 8.

\(^{206}\) It also violates Kagan’s claim that no moral theory should fail to minimally reflect morally relevant features of human persons.
The consequentialist might agree that no moral theory should imply that we ought to do something that is physically or psychologically impossible. However, this response is consistent with the consequentialist viewing our projects—due to their usual impersonal inefficiency—with hostility. Furthermore, humans are plastic creatures: perhaps we could alter future generations of humans (via radical experiments in education, genetic manipulation, etc.,) in order for us to become capable of doing what is presently impossible. On the other hand, such attitudes toward future generations might be seen as a *reductio ad absurdum* of this whole line of the consequentialist response. Even a self-avowed “extremist” such as Kagan rejects such a view of the nature of human persons as overly hostile to our own morally significant features.207

5.5 Response 2: Partiality Justified Impartially

The consequentialist may respond to the moderate’s defence for options by initially agreeing that a moral system must acknowledge the essential role that projects play in living the good life, but subsequently disagree with the moderate’s charge that a moral theory that fails to incorporate options rules them out. A sophisticated version of consequentialism may account for the value of projects by making room for them. While projects may appear to represent inefficiencies at attaining the best state of affairs, they may be offered a justification on *consequentialist* grounds. That is to say, allowing people to structure their lives around various projects may be the *means* of bringing about the best state of affairs, impartially judged.

---

Consider the example of a father treating his own child with partiality. It is plausible to hold that the means of achieving the best state of affairs is for parents to exhibit a general disposition of partiality toward their own children. A general disposition of partiality towards kin can be morally justified on the grounds that the predominance of this disposition helps ensure that each child will have the best chance of receiving a nurturing and loving environment. In times of scarcity, for instance, parents are the best judges of their own children's needs. Thus, even though the occasional situation may arise in which offspring partiality inhibits the best state of affairs, it is plausible to maintain that the entire strategy is the most likely means of producing the best state of affairs. A sophisticated form of consequentialism permits the partiality that, for example, a father shows to his own child.\textsuperscript{208}

While the consequentialist is correct to focus on the consequences of character dispositions and various social institutions (e.g., the institution of the family), Bernard Williams thinks that the consequentialist's reply is not problem-free. Williams points out that a consequentialist who thought like a consequentialist all the time would have a character that is, to say the least, morally unappealing. A moral theory needs to address what kind of moral characters proponents of the theory would possess: the parent who loves his child on the grounds that doing so is an efficient means of bringing about the best state of affairs would be a morally abhorrent parent. Williams criticizes such a consequentialist, claiming this parent has had "one thought too many".\textsuperscript{209} A parent with a


morally appealing character acts in a partial manner toward his or her child because this is what a *good* parent does; wondering whether or not a moral theory would endorse this partiality is to have “one thought too many.”

Williams is correct that a consequentialist who *always* thought like a consequentialist would have a morally repugnant character. Yet, the consequentialist has a reply to Williams’ challenge. As outlined in Chapter 1, a sophisticated consequentialist may recognise that obtaining the best state of affairs may come about by *not always* thinking in consequentialist terms.\(^ {210}\) Given Williams’ challenge, there are reasons against taking consequentialism to be a subjective theory of right decision.\(^ {211}\) Peter Railton develops a view he calls “objective consequentialism”: “the view that the criterion of the rightness of an act or course of action is whether it in fact would most promote the good of those acts available to the agent.”\(^ {212}\) Such a form of consequentialism is distinguished from a subjective consequentialism, a view that holds that “whenever one faces a choice of actions, one should attempt to determine which act of those available would most promote the good, and should then try to act accordingly.”\(^ {213}\)

Objective consequentialism is a moral theory that informs us which actions and institutions are the right ones; it says nothing specifically and directly on how a person should morally deliberate. Railton’s objective consequentialism is outcome-focussed. The *means* of producing outcomes are—intrinsically at any rate—irrelevant. Whatever actually does produce the best state of affairs is what objective consequentialism

---

\(^{210}\) As argued in Chapter 1, section 3.2.1.


\(^{212}\) Ibid., 113.

\(^{213}\) Ibid., 113.
demands. At the time when an individual is confronted with some moral decision, he or she may permissibly think anything at all, so long as the right action is performed.

Given that one of the functions of moral theory is to aid moral deliberation by offering agents guidance, Railton's objective consequentialism is somewhat counter-intuitive. Objective consequentialism fails to offer agents guidance about how to deliberate when confronted with a morally complex case. In an effort to explain away the counter-intuitive aspect of his objective consequentialism, Railton draws on common sporting examples. Consider a professional golfer. While her overall goal may be to win as many golf tournaments as possible, during a tournament it is counter-productive to dwell on this big-picture goal. What is more effective at producing the desired result is to focus on the specifics of the game being played. The professional golfer develops two separate modes of thought about golf: "in-play mode," and "out-of-play mode". While the professional golfer is in-play mode, she must force herself to not think about her overall goal because doing so is the means of achieving it. A consequentialist is like the professional golfer in the sense that sometimes she must not think like a consequentialist because doing so is the means to achieve her actual goal—producing the best state of affairs, impersonally judged. I will label this response to Williams' objection the "two modes of thought" response.

There is nothing novel about developing two modes of thought toward one thing.\textsuperscript{214} From tight-rope walking, to catching a ball, people realize that what a person ought to think in specific cases might not be one's overall goal but instead whatever it is that helps one actually capture one's overall goal. Railton's two modes of thought response offers

an adequate reply to Williams’ one thought too many objection because the negative implications about a person’s character do not arise: Railton’s two modes of thought response shows why consequentialism would not demand that a consequentialist always think like a consequentialist.

5.6 Why Objective Consequentialism Fails to be a Human Morality

Yet, even though Railton’s strategy is successful at addressing Williams’ one thought too many objection, objective consequentialism does not successfully incorporate partiality into consequentialism. Railton’s attempt to incorporate partiality on impartial grounds fails because such a theory entails that its advocates should lament their own partiality. In order to defend this response to Railton, I will employ a distinction between a “proponent” of objective consequentialism and an “adherent” of objective consequentialism. A proponent of objective consequentialism does what objective consequentialism demands because of his or her belief in the truth of objective consequentialism; an adherent of objective consequentialism merely acts in accordance with the dictates of objective consequentialism. Given this distinction, Railton’s objective consequentialism would entail that a person should not always be a proponent of consequentialism. During the act, it can be necessary to merely be an adherent of consequentialism.

Let’s look in close detail at Railton’s two-modes of thought response as it applies in the example of Juan:

Consider again Juan and Linda, whom we imagine to have a commuting marriage. They normally get together only every other week, but one week she seems a bit depressed and harried, and so he decides to take an extra trip in order to be with her. If he did not travel, he would have a fairly large sum that he could send to Oxfam to dig a well in a drought-stricken village. Even reckoning in
Linda’s uninterrupted malaise, Juan’s guilt, and any ill effects on their relationship, it may be that for Juan to contribute the fare to Oxfam would produce better consequences overall than the unscheduled trip. Let us suppose that Juan knows this, and... given Juan’s character... will travel to see Linda. The objective act consequentialist will say that Juan did the wrong act on this occasion.... Thus, it may be that Juan should have (should develop, encourage, and so on) a character such that he sometimes knowingly and deliberately acts contrary to his objective consequentialist duty. Any other character, of those actually available to him, would lead him to depart still further from an objectively consequentialist life.215

Railton is clear that Juan’s going to visit his wife is, according to objective consequentialism, the wrong act for Juan to do. Juan’s character leads him to act inefficiently, as judged from an impersonal perspective. However, in order for Juan to make room for important relationships in his life, he must have the type of character that responds to the needs of his loved ones. Thus, in light of Juan’s character, seeing Linda is the best possible act that Juan could perform: Juan’s act is morally permissible. What Railton has overlooked, however, is what it would be like to be Juan in this case. When should Juan be a proponent of consequentialism and when should he be an adherent? That is to ask: when is Juan in-the-act and when is he out-of-the-act?

Unlike tight-rope walking and golf tournaments, some morally assessable acts fail to have relatively clear “boundary conditions”.216 Assuming that Juan is an objective consequentialist, he must adopt two different modes of thought about his own action: on the one hand, he wants to see his wife and look after her because he loves her (in-the-act). On the other hand, we may assume, he knows it is his moral duty to produce the best state of affairs, impersonally judged (out-of-the-act). Railton thinks Juan will not be frozen


216 It is quite clear when one is walking a tight-rope or playing a game of golf and when one is not; such relatively clear boundaries are not true of Juan’s act of going to visit his sick wife (is Juan’s act the whole commute trip, the evening he spends comforting his wife, or some specific number of hours?).
with indecision because Juan knows that objective consequentialism “makes room” for projects in our lives. If you ask Juan right before he leaves the house the justification for his action, he would say it’s because he loves his wife. Sometime later, if you ask Juan the same question, he gives a much different answer: “Well that act was, strictly speaking, immoral. But, it was a necessary evil due to the way humans naturally function. We have to permit ourselves to do some inefficient acts like what I did in order that—in the big picture—people have a character that is disposed to do the right thing most of the time.” Yet, how much later does this second justification cross Juan’s mind? Three days after he has returned home from visiting his wife? After he has soothed her to sleep? Once the ferry has left the berth and Juan has passed the point of it being objectively optimal to return? These questions cannot be all that clearly answered because—unlike the sporting examples—there are no sharp boundaries that delineate when Juan is in-the-act and when he is out-of-the-act. Therefore, there is no clear separation between Juan the adherent of objective consequentialism and Juan the proponent of consequentialism. Juan might help his wife because he loves her, but he will also, on some level, simultaneously lament the fact that he has a character that leads to such impersonally inefficient—hence, immoral—actions.

Even if the problem of clearly delineating the boundary conditions of an action can be resolved, another problem emerges. Qua a proponent of objective consequentialism, Juan adopts an instrumental attitude toward his relationship with his wife. Juan maintains his relationship with his wife because his relationship is conducive to bringing about the best state of affairs. This instrumental attitude diminishes the possibility of having a deep

---

217 Railton, “Alienation,” 133.
and meaningful relationship because deep and meaningful relationships involve the agent taking the other person to be intrinsically important, and not merely of instrumental value for producing impersonal good. *Making room for our partiality* is not enough: we cannot turn off our partiality when (and only when) it is impartially efficient to do so. Our partiality toward a loved one does not come with an off switch. The fact that we value the world around us from the first person point of view is not something from which one may opt out. In short, being a proponent of objective consequentialism entails putting deep and meaningful relationships in jeopardy. In light of the significance of the fact that persons value the world from the first person point of view, this entailment is an unacceptable entailment for a *human* moral theory.

Proponents of objective consequentialism “make room for” their own partiality, but simultaneously wish it were otherwise. The case of Juan highlights this problem with an impartial justification for partiality: proponents of objective consequentialism ultimately take an attitude of disdain toward partiality in general, and his or her own projects in particular. The consequentialist has tried to justify limited partiality on impartial grounds. This response to the moderate’s challenge fails to be a successful response because our partiality ought to be *embraced* by moral theory, not grudgingly accommodated.

### 5.7 Why Top-secret Objective Consequentialism fails to be a Human Morality

In some cases, consequentialists hold that the way to attain the best state of affairs is to not think about morality at all. While this may be plausible in some extreme cases (especially when time is of the essence and acting—and not morally deliberating—is necessary for an agent to do the right thing), the objective consequentialist may think
such a strategy ought to be employed more widely. In fact, one response to my argument from the last section might develop this line of thought. In the case of Juan, the problem is that, out-of-the-act, he is a proponent of objective consequentialism. While a proponent of objective consequentialism may sometimes have his moral thinking thwart the production of the best state of affairs, the same is not true of an adherent of objective consequentialism. Thus, one response to my argument in the last section is to bar people from being a proponent of objective consequentialism. An objective consequentialism is "top-secret" if and only if it advocates that (almost all) agents are adherents of the theory, and bars (almost all) agents from being proponents of the theory. Some consequentialists, like Railton, do not see any problem with a top-secret morality. Yet, such a moral theory would fail to have "transparency." A moral theory that lacks transparency implies that whatever moral reasoning adherents actually engage in is false (or, at least, necessarily incomplete). The problem I wish to focus on is the one suggested by Williams: defending a moral theory that lacks transparency amounts to prohibiting people's access to moral integrity.

I take it that a person lives a life of integrity when, at least, their actions and character are consistent with their beliefs about what their actions and character ought to


219 Almost all agents, and not all agents, because at least some agents, at some point in time (e.g., the moral theorists under discussion) will write about top-secret objective consequentialism.

220 A moral theory lacks transparency if—by its own lights—it would be wrong for the theory to be made public; see: Railton, "Alienation," 116.

221 Sometimes it may be efficient for adherents to hold false beliefs (when manipulation is more efficient at attaining the best state of affairs than the truth). In other cases, adherents may be permitted to know mid-level justifications, but—by definition—adherents are not permitted to know a complete account of moral justification.

be. The problem with top-secret objective consequentialism is that most people would not even be permitted to try to find out the truth about morality.\textsuperscript{223} Adherents of objective consequentialism are merely being manipulated to act in accordance with the demands of objective consequentialism. Adherents of objective consequentialism may be barred from living a life of integrity because they are not permitted to engage in any complete moral thinking. Barring agents from thinking about their own actions and lives in certain ways constitutes an unacceptable attack on their integrity.

Historically, some philosophers have advocated ethical or moral theories that suggest the masses ought not be permitted to learn the true nature of these theories. While Plato’s Myth of the Metals comes to mind, some more recent examples exist. Margaret Walker, in her essay “Feminism, Ethics, and the Question of Theory”, offers a critical examination of Sidgwick’s claim that it is permissible for a moral theory to lack transparency. Sidgwick did not think his utilitarianism was for the common folk. A morally superior sub-section of society had a burden to “covertly deploy this rigorous systematic moral theory for the public weal under the ‘actual condition of civilized communities, in which there are so many different degrees of intellectual and moral development’.”\textsuperscript{224} Sidgwick’s utilitarianism was a moral theory for the elite, used as a tool for “managing” the rest of society. Walker quite rightly criticizes such views on the grounds of its implied elitism, not to mention its obvious anti-egalitarian implications. She argues for a general sense of caution when confronted with approaches to moral

\textsuperscript{223} Consider Williams’ related comments on “government house morality”; see: Williams, The Limits of Philosophy, 109-110.

theory that yield such results: "This unpleasant fusion of images of the superior moral caste, internalized remote control, and manipulative influence should give us pause."225

What is wrong with a moral theory that lacks transparency? A life cannot have moral integrity by chance or coincidence; someone else cannot manipulate a person in such a way that he or she ends up living a life with moral integrity. How can an agent lead a morally consistent life of integrity if she is not permitted to know the truth about morality? For naturally reflective and intellectually curious beings like human persons, actually engaging in genuine moral deliberation is a necessary condition for living a life with moral integrity. Top-secret objective consequentialism would all but rule out the permissibility of living a life with moral integrity. Top-secret objective consequentialism puts almost everyone's moral integrity in jeopardy and thus is not a moral theory for beings like us.226

I have considered two replies to the claim that consequentialism needs to acknowledge the fact that persons value the world from the first person point of view: the rejection of the claim that projects are central to the good life (Section 4) and the rejection of the claim that consequentialism fails to acknowledge the moral significance of projects. I have showed why both replies have good grounds to be rejected. I will now turn to other possibilities for incorporating partiality in general—and projects more specifically—into a moral theory.


226 With the exception of the few people privy to the secret.
5.8 Incorporating Partiality with Agent-neutral limits: Scheffler’s Hybrid Solution

In *The Rejection of Consequentialism*, Scheffler agrees with the claim that people need to be liberated, at least sometimes, from the duty to always act in an optimal manner. In this section I will outline the specifics of Scheffler’s account of the nature and extent of the role for human partiality in moral justification. Ultimately, I raise some doubts that he has offered the moderate the right type of solution for incorporating options into a moral theory.

A defence of options—Scheffler refers to them as “agent-centred prerogatives”—would offer permission for agents to make impersonally judged sub-optimal choices (within some unspecified limitations). For instance, an agent-centred prerogative would permit a father to treat his own son partially even if it is inefficient to do so, impersonally judged. Agent-centred prerogatives, if defensible, would entail the denial of extremism and consequentialism: people would not be obligated to always produce the (otherwise morally permissible) best state of affairs. Scheffler explains as follows:

Scheffler knows, however, that intuitions on their own are not enough. As I pointed out earlier, Scheffler argues that agent-centred prerogatives indeed have a principled rationale: the independence of the personal point of view.

One virtue of Scheffler's account of agent-centred prerogatives is that it does a good job of addressing the problems associated with a moral theory that laments the fact that human persons value the world around them from the first person point of view. Scheffler is not worried merely about the size of the burdens (such as the appeal to costs defence for options), but instead focuses on the implications associated with a moral theory that imposes such burdens. Scheffler is especially concerned with how strict consequentialism would not reflect, or mirror, our nature as beings with a personal point of view. Scheffler's hybrid consequentialism is a response to these worries.

How much out of proportion ought agent-centred prerogatives permit agents to value their own projects and commitments? Scheffler (wisely) leaves this proportion unspecified. In *Moral Animals*, Catherine Wilson argues that the appropriate size of the proportion is not easy to identify. If the proportion is too small, then we fail to have room for many plausible non-malevolent projects. For instance, options have to be large enough that they permit agents to at least sometimes fail to save lives merely because the agent does not want to (e.g., I would be permitted to—at least occasionally—spend $200 on a show and a nice dinner instead of famine relief). However, if the proportion is too large, then Scheffler's hybrid-consequentialism would be indistinguishable in practice from egoism. Options would have to be small enough that they fail to permit me to directly harm others for my own enjoyment (e.g., I cannot permissibly hurt innocent

---

people for my own amusement). Wilson worries that agents could evoke Scheffler’s agent-centred prerogatives in order to defend the permissibility of bringing about greatest personal satisfaction. Such an agent “no longer seems to be operating within a normative system at all.”

The great difficulty associated with identifying some plausible proportion for an agent-centred prerogative begs that the following question be answered: is Scheffler’s agent-centred prerogative the right type of solution to the problem of a moral theory that denies the moral significance of the fact that persons engage the world from the first person point of view? I will use two arguments by analogy to challenge Scheffler’s solution. The first argument by analogy is aimed at Railton’s objective consequentialism, and then the second will develop this analogy in order to undermine the spirit of Scheffler’s solution to the problem.

**Analogy 1:** Consider a couple, Ted and Sally; suppose they have been dating seriously for 1 year. Of late, however, Ted finds that the relationship is not working out so well. When confronted, Ted is honest with Sue. “The problem,” he says, “is that you lack spontaneity. I like people with a zest for all the opportunities life has to offer. You’re just too predictable for me. I need someone with a little bit of chaos, you know, to stir things up”. Imagine a week later Sally meets Ted and explains to him she has found a solution to their problem.

“I am just not a spontaneous person, Ted, but I agree that a little more spontaneity would be good for me. I am a little too straight-laced and loosening things up a bit would be good. To that end, I have decided to act spontaneously 7 times per week. I think this should solve the problem.”

Ted would be unhappy with Sally’s response. A non-spontaneous solution to the problem of lacking spontaneity is no solution at all: such a “solution” merely masks or

---

229 Ibid., 82.

230 These analogies are inspired by Philip Pettit’s example of calculatively elusive goods like spontaneity; see: Philip Pettit, “The Consequentialist can Recognize Rights,” *The Philosophical Quarterly* 38, no. 150 (January 1988): 42-55.
covers up the real issue, namely, that Sally lacks spontaneity. It is essential for genuine spontaneity that the agent thinks about her action in a *spontaneous* or un-calculating way. In short, if an agent is trying to calculate her way to spontaneity she will miss her goal because whatever behaviour her calculations produce will not be *genuinely* spontaneous. For an act to be spontaneous, it is not only the case that it has to be, objectively speaking, of a certain nature, but from the agent’s own perspective it must be thought of in that way as well.

The case of Sally’s “solution” to Ted’s problem fails in a parallel fashion to Railton’s method of justifying partiality *impartially*. Genuine partiality—precisely like genuine spontaneity—involves more than acting in a certain manner. Genuine partiality involves not only doing things that demonstrate partiality (carrying out one’s own projects, etc.,) but also thinking about what one is doing in a partial way. In order for an agent to be partial to someone or something, she must in the first person value that thing out of proportion to its impersonal value.

But how much out of proportion? Partiality needs limits, so what are these limits?

**A further development of Analogy 1:** Ted explains why he thinks Sally’s method does not give him what he wants. So Sally takes another try: “instead of just establishing by fiat that I will act spontaneously 7 times per week, each Monday I will roll some dice and then let the sum of the dice determine how many times I will be spontaneous that week.” While Sally admits she is not sure how many dice she will roll each Monday, “my method will sure generate spontaneous results. You will never have the slightest clue each week just how spontaneous I will be!”

There is a parallel between Sally’s attempt to address Ted’s complaint and Scheffler’s attempt to address the moral significance of the independence of the personal point of view via his agent-centred prerogatives. Under Scheffler’s system, hybrid

---

231 Ibid., 50-51.
consequentialism still demands that our partiality, ultimately at any rate, is justified via the agent-neutral point of view. Scheffler’s method suggests that reasons that stem from our natural partiality should be multiplied by some unspecified proportion in order to determine their “actual” impersonal value. Such a solution is in tension with Scheffler’s own rationale for his agent-centred prerogatives. The method of employing Scheffler’s solution does not fit with its purported rationale. While this concern does not represent a definitive argument against Scheffler’s agent-centred prerogatives, it does imply that we should consider other solutions to the problem of limiting the extent of what our natural partiality may justify.

Scheffler’s hybrid consequentialism might seem to occupy a promising middle ground between a totally partial and a totally impartial morality. Yet, as my argument by analogy shows, Scheffler’s agent-centred prerogatives do not appear to be the right type of solution.

5.9 Incorporating Partiality With Agent-Relative Limits: Wilson’s Anonymity Requirement

As my discussion of Scheffler’s agent-centred prerogatives indicates, I think that our partiality needs limits: defending moral justifications that take their basis in our partiality need not—and ought not—represent a defence of some form of egoism. Is there a way for a moral theory to “embrace” our partiality, and yet still manage to be a normative theory that issues demands—including costly ones—that legitimately bind agents who
demonstrate partiality? I think such a possibility exists: Wilson’s Anonymity Requirement appears to strike just such a balance.\footnote{Wilson, \textit{Moral Animals}, 129.}

Wilson argues in favour of the Anonymity Requirement on the plausibility of finding limits to our partiality: justifications that stem from our partiality ought to be accepted by the person advancing them, regardless of the position the person advancing the norm actually occupies. For someone to issue the norm: “I may feed my own child before I feed other children”, it must be the case that the person agrees to the norm regardless of her particular situation (i.e., she is the parent of the hungriest child, or she is the 2\textsuperscript{nd} hungriest child, etc.). Arguably, such a norm would be met with wide assent.\footnote{While the Anonymity Requirement has some obvious similarities with Rawls’ Original Position, there are at least a few important differences (i.e. the Anonymity Requirement is not an abstract, hypothetical decision situation but a potentially actual one; and, unlike Rawls, Wilson is not assuming \textit{a priori} counter-intuitively high levels of risk-aversion).} This strategy limits partiality from the “inside” because the agent still imagines herself as an agent that is partial, albeit \textit{various different} agents that are partial.

Consider a more concrete case. Suppose a government official hires her incompetent nephew for a lucrative appointment even though a number of far more qualified applicants submitted applications. She justifies her actions on the grounds that she is partial to her nephew. Yet, it is dubious whether the norm implied by this agent’s conduct meets the Anonymity Requirement. The government agent’s justification for her action implies the following norm: “I may reward family members with lucrative jobs, even when the person I reward is unqualified and other more qualified candidates applied.” The test in this case is whether our government worker would agree to this norm if she were some different stakeholder in the case: an uninvolved tax-payer, an
honest co-worker, or a passed-over and qualified candidate. Judged from the point of view of the various people involved in such a case, this norm would not be met with wide assent. Our government worker’s justification surely fails Wilson’s Anonymity Requirement. What makes Wilson’s Anonymity Requirement a novel solution to the problem of limiting partiality is that it generates reasonable limits on justifications that stem from our natural partiality from within a perspective that is partial.

Although the Anonymity Requirement offers some concrete limitations on our partiality, one may worry whether we are back where we started: weighed down by moral obligations that fail to reflect our nature as beings that are naturally partial. The Anonymity Requirement would surely demand, for example, large sacrifices from affluent people. Take the case of the following norm: “affluent people should forgo some unnecessary luxuries each year (e.g., limiting consumption to 100 cups of freshly brewed coffee per year, instead of 500, etc.,) in order to assist a distant person that lacks access to clean water, healthy food and protective shelter.” If one considers this norm from the various perspectives of different people that may be affected, it is plausible to anticipate that such a norm would meet with wide assent. Yet, I think such a result is appropriate. As I stated at the outset of this chapter, I do not think an argument for the relevance of human partiality to moral justification is best defended on the grounds that it is overly burdensome to ignore feature of persons. The Anonymity Requirement would be stringent and, for example, make strong demands on affluent people.

234 On what grounds do I claim that this norm would not be met with wide assent? It would not be met with wide assent because, if one asked educated and sympathetic persons if they would prefer to live in the world that employs that norm or a world that did not employ that norm, they would all prefer to not live in such a world.
My reason for thinking that such a requirement’s burden is less troubling than the burden generated by moral theories that deny options is that the type of burden is appropriate. While the Anonymity Requirement may demand that affluent people go without expensive steak dinners every week and trendy $2000 calf-skin handbags, it is not plausible that the Anonymity Requirement would, in general, rule out life-framing projects. The Anonymity Requirement is not an onerous attack on our projects or partiality. Norms that permit people to act with partially toward their own actions, lives and loved ones are norms that would be met with wide assent. The Anonymity Requirement does not imply that we ought to lament our partiality. It is in this sense that the Anonymity Requirement manages to embrace our nature as beings who are partial, and yet still have the capacity to generate stringent norms. Wilson’s Anonymity Requirement offers us a starting point for separating legitimate justifications that stem from partiality from illegitimate ones.

5.10 Conclusion

Human beings are naturally partial; furthermore, we think it a good thing that we are naturally partial. Contrary to what is implied by moral theories that deny options, our partiality should not be a feature of our nature that is lamented. Completely impartial moral theories can merely make room for our human partiality. Such an impartial justification for partiality ultimately fails because a proponent of consequentialism—from the first person point of view—would have to be capable of fettering his or her partiality when, and only when, doing so is impersonally optimal. Such a demand is the mark of a moral theory that laments our partiality and thus fails to be a human morality.
Of course, our partiality needs limits. I think Wilson’s Anonymity Requirement represents a starting point for understanding specific limitations on our partiality. What I think is particularly successful about Wilson’s Anonymity Requirement is that it aims to place limits on partiality from within a partial perspective.
BIBLIOGRAPHY


Pettit, Philip and Geoffrey Brennan. “Restrictive Consequentialism.” In


Pettit, Philip and Michael Smith. “Global Consequentialism.” In Morality, Rules, and


Popenoe, Paul and E. S. Gosney. Sterilization for Human Betterment: A Summary of


Putnum, Hilary. The Collapse of the Fact / Value Dichotomy: and Other Essays.


Railton, Peter. “Alienation, Consequentialism and the Demands of Morality.” In


APPENDICES

At the start of this thesis, I asked whether the authorities should observe the internationally recognized rules regarding the treatment of imprisoned enemy combatants, or the authorities should secretly violate some human rights in order to make everyone safer. In Appendix A, I will respond to this original question by examining three cases related to the war on terror. In Appendix B, I will apply the limitations on consequentialist justifications I have defended by examining the historical case of state-sponsored enforced sterilization for eugenic purposes.

Appendix A  Prospects for Application: The War on Terror

A.1  The Treatment of Suspected Enemy Combatants

Consequentialist analysis: We live in the post 9/11 era. Some extremists are willing to do whatever it takes to cause the greatest possible loss of innocent life. Many country’s citizens walk around in a perpetual state of fear and anxiety because of the spectacular and graphic nature of the 9/11 attacks. In order to limit the negative impact of possible future acts of terrorism, the governments of these countries need to promote the value of security. The state has to be willing to get its hands dirty. In light of these extraordinary circumstances, the governments of countries threatened by terrorism are permitted to do what otherwise would not be morally permissible. In short, the state should have the right to incarcerate and interrogate suspected unlawful enemy combatants without trial or the right to legal council because doing so will very likely add to the security and safety
of the greatest number. In a normal circumstance, such actions would not be morally permitted. However, because of the extreme nature of the terrorism crisis, the state must be willing to do what it takes to bring about the best overall results. Producing the best results will, in all likelihood, involve some innocent people being unjustly incarcerated and subjected to the discomfort of interrogation. However, consequentialists like Pettit are clear that sometimes we cannot afford to honour values—such as the value of respecting innocent persons—when it is obvious that promoting the value requires that we fail to honour the value.

**Response:** Suppose, for the sake of argument, I grant that the best state of affairs would indeed be brought about by the state occasionally incarcerating and interrogating suspected unlawful enemy combatants without trial or the right to legal council. Even if I make this dubious concession, the consequentialist’s argument is not sound. The focus of my thesis is to highlight this faulty argument, and is best understood from the perspective of the innocent person who is erroneously thought to be an unlawful enemy combatant. Without the checks and balances of a judicial system, authorities will make otherwise avoidable errors: the incarceration and interrogating innocent persons will occur far more often than otherwise. The consequentialist clearly agrees that the practice of incarcerating and interrogating suspected enemy combatants represents the authorities “getting their hands dirty”; however, the consequentialist argues that the burdens of incarceration and interrogation experienced by a few innocent people is outweighed by many millions of people experiencing small increments of extra security. Even though general security is an important value, I have argued in the dissertation that it is not appropriate for the authorities to cause a few people to experience extreme burdens (such
as incarceration and interrogation) merely to produce the greater good. Incarcerating and interrogating suspected unlawful enemy combatants without trial or access to legal council is morally ruled out by any moral theory that wishes to respect the moral significance of the natural separateness of persons: all people need to be treated in a manner consistent with the practice of taking persons to be units of moral concern. Trading off the occasional innocent person’s long-term mental (and, perhaps, physical) health for the overall general good is to fail to acknowledge this person as a separate and significant unit of moral concern.

A.2 The Beslan Massacre and the Moscow Theatre Hostage Situation

Consequentialist analysis: Perhaps the war on terror does not offer the state a *general* licence to violate restrictions. That said, some rare and extreme circumstances arise where the authorities have to make the decision between a smaller number of innocent people being murdered and a greater number of innocent people being murdered. For instance, consider two recent terrorist cases in Russia, namely, the Beslan School Massacre (2004) and the Moscow Theatre Hostage Situation (2001). In both cases, a group of well-armed terrorists commandeered large public venues; hundreds, even thousands, of innocent people were held hostage. In both cases all attempts at peaceful negotiations failed: even during the negotiations, the terrorists demonstrated time and again their willingness to employ lethal force. The authorities knew that an order for special-forces to storm the building would directly bring about the deaths of many innocent people. Furthermore, it was reasonable to predict that—due to the general chaos and confusion generated by storming a large building filled with hundreds of panic
stricken hostages—at least some of the innocent people would be killed by the special forces themselves. In such an extreme and very specific situation, the right thing to do from a consequentialist perspective is to give the order to storm the building, even though the actions carried out will result in the loss of many innocent lives. The defender of restrictions—paradoxically—must hold that the order is morally impermissible even though storming the building would (we are assuming) minimize the number of overall restriction violations.

**Response:** The Beslan massacre and the Moscow Theatre Hostage situation are more than a general emergency situation. These two cases are about as close as we can get in the real world to a genuine tragic dilemma, as defined in Chapter 4. Given that all peaceful alternatives have been exhausted, it seems clear that many innocents will be killed, regardless of the actions of the authorities. There is no available action that is obviously consistent with our practice of treating everyone as a morally significant unit of concern. If the order to storm the building is given, then fewer lives will be lost in total; however, the authorities will surely kill at least some of the innocent people (due to the chaos and confusion generated by the officials making the first move). I think that the order is a *permissible* order. Thus, I disagree with the consequentialist’s analysis of this case on two key points: first, giving the order is not justified by an appeal to the best states of affairs exclusively, but rather there are also agent-relative moral justifications for storming the building; second, I deny that it is morally *obligatory* to storm the building (in the third and final example I will respond to this second claim).

What agent-relative justification permits the order to storm the building? The agent-relative justification is that the hostages made it clear that they wanted the *special-forces*
to storm the building. "Giving the order to storm the building is morally permissible because *the hostages wanted us to do so*" is an agent-relative justification because it refers to particular agents, and not merely that some state of affairs is produced. Furthermore, acknowledging the relevance of this agent-relative reason for solving the tragic dilemma incorporates the wishes of the potential victims in the causal pathway of the harm to which they are subjected. The defender of restrictions may claim that the order to storm the building is morally permissible, and thus avoid the consequentialist's charge that restrictions are so strong they are paradoxical.

A.3 Options and the Order to Storm the Building

**Consequentialist analysis:** Let's suppose an official has given the order that all people able to assist the special forces must aid in the effort to storm the building; let's assume, further, that it would be impersonally best, overall, if you assisted in the effort to storm the building. Let's also suppose that you have no official training in such matters and have never chosen to put yourself under an obligation of this kind (e.g., you never chose to enter the special forces). Under these assumptions, the consequentialist holds that all such agents are morally obliged to follow the orders to storm the building: incurring the risks of injury or death (not to mention the fact that your actions might kill innocent persons) are facts that have already been taken into account in the ranking of the impersonal states of affairs. Agents are not morally permitted to refuse the order, even in a case where following such an order goes against the agent's desire for self-preservation or against the agent's general commitment to non-violence. Sometimes morality is tough: agents need to steel themselves to carry out the morally correct actions.
Response: Assuming that it would be ideal, impersonally judged, for some specific agent to storm the building, then the consequentialist insists that that agent would be morally obligated to storm the building. However, such an account of moral obligation is too narrow because it rules out the possibility of agents morally justifying what they do (or fail to do) on the grounds that the action would conflict with a cherished project (e.g., self-preservation, pacifism). I have argued that in order for a moral theory to appropriately acknowledge the moral significance of the fact that human persons value the world around them from the first person point of view, a moral theory must embrace—and not lament—this feature of human persons. Thus, in the case I have considered, it would be morally justified for the agent to fail to obey the order, even assuming that storming the building would lead to the best state of affairs.

The consequentialists' error is their assumption that all values can be accounted for impersonally. Being obligated to give up a central project cannot be thought of as a cost that can be simply calculated into overall impersonal value because having projects are at least sometimes constitutive for beings like us—beings that value the world around them from the first person point of view—to have access to a life worth living. The consequentialist denies justifications that stem from our partiality by adopting an explicitly (or implicitly) negative attitude toward partiality. This feature of consequentialism has grounds to be rejected as a plausible account of morality for human persons. In short, even if it were judged to be impersonally optimal to follow the order to storm the building, failing to comply with this order may well be morally permissible, e.g., if you are a pacifist.
Appendix B  An Extended Application: Enforced Sterilization for Eugenic Purposes

B.1  Introduction

I will offer a lengthier application of the findings of my dissertation by examining the historical case of compulsory sterilization as part of a state-sponsored eugenics program. I have chosen to examine this historical case because it represents an instance where people actually offered consequentialist justifications for grossly assaulting people. My goal is to critically examine these consequentialist justifications in their historical context. I am choosing a historical case because the facts of the situation are not, for the most part, in dispute. Thus, we can side step many epistemic issues regarding the general facts of the case. I believe that this examination will highlight what I have argued is so problematic with the consequentialist's purely instrumental account of moral justification: consequentialism does not acknowledge the moral significance of human separateness and our nature to be partial. A necessary condition on a moral theory being a moral theory for human persons is that the theory must reflect morally relevant features of our species’ nature.

First, I will offer a partial history of the U.S. eugenics movement, featuring an overview of the coercive eugenics programs that were employed throughout the first four decades of the 20th century. I will focus particularly on the consequentialist—specifically, utilitarian—justifications that purported to defend these programs. Second, I will carefully examine Popenoe’s and Johnson’s arguments that defended “negative
eugenics”—the reduction of socially “undesirable” births. Third and finally, I will show how restrictions on consequentialist justifications would not only bar the state from employing coercive negative eugenics programs, but also highlight the problems inherent in any exclusively instrumental approach to morality.

It is generally regarded that the first quarter of the 20th century was a time when America had an aspect of totalitarianism within some of its social institutions. A new science—“eugenics” or “the science of good breeding”—was being advanced by academics from several different backgrounds. Paul Popenoe and Roswell Johnson claimed that their book, *Applied Eugenics*, should command the attention not only of students of sociology, but, as well, of philanthropists, social workers, settlement wardens, doctors, clergymen, educators, editors, publicists, Y. M. C. A. secretaries and industrial engineers. [*Applied Eugenics*] ought to lie at the elbow of law-makers, statesmen, poor relief officials, immigration inspectors, judges of juvenile courts, probation officers, members of states boards of control and head of charitable and correctional institutions.

The topic of eugenics was thought to be important for people who had jobs or duties that put them in positions of authority in society.

Eugenics was dominated by two overall strategies: its positive aspect—increasing socially useful births—and its negative aspect—decreasing socially useless and costly births. In what follows I will focus on one method of negative eugenics: the coercive

---


237 Ibid., 156.
sterilization of the people deemed “unfit”, even though other forms of negative eugenics were practiced and were popular.238

Some 61,000 coercive eugenic sterilizations were carried out in the US over the course of the first six decades of the 20th century; 25,000 of these procedures were carried out during the peak of the eugenics movement in the 1930s.239 The medical procedures in question were recent medical advances. In the case of men, a relatively simple vasectomy was carried out, but in the case of women a more complex and slightly riskier procedure—salpingectomy or hysterectomy—was performed. Who were the targets of these coercive medical procedures? According to Popenoe and Johnston, the targets were:

inefficients, the wastrels, the physical, mental and moral cripples [that] are carefully preserved at public expense... In short, the undesirables of the race, with whom the bloody hand of natural selection would have made short work early in life, [but who] are now nursed along to old age.240

Robert Cynkar offers an excellent overview of the (now infamous) case of Carrie Buck.241 Carrie Buck was a young adult who was committed to an institution due to “immorality, prostitution, and untruthfulness.”242 Carrie was considered to be “feebleminded”; she was assessed to possess the mentality of a nine year old.243 Her feeblemindedness appeared hereditary, as her mother and her illegitimate baby daughter were

---

238 Ibid., 188. Popenoe and Johnson primarily argue in favour of the segregation of the sexes in state-run institutions for the eugenically unfit. For an overview, see: 184-210.


240 Popenoe and Johnson, Applied Eugenics, 149.


242 Ibid., 1418.

243 Ibid., 1418.
judged to suffer from the same condition. How could society stop the spread of feeblemindedness? According to Cynkar, “The solution to the problem seemed obvious... one merely had to segregate [people like Carrie] during [her] fertile years, or sterilize them, to rid society of [her] kind humanely and permanently.”

Carrie was incarcerated in a state-run mental institution. When the order for her sterilization was given (under the auspices of the Virginia 1924 law that permitted the sterilization of the inmates of state-run mental institutions), her guardian appealed. After several unsuccessful appeals, in 1927 the highest court in the land handed down the now infamous opinion, with added comments from Justice Oliver Wendell Holmes:

Carrie Buck “is the probable potential parent of socially inadequate offspring, likewise afflicted, that she may be sexually sterilized without detriment to her general health and that her welfare and that of society will be promoted by her sterilization,”... We have seen more than once that the public welfare may call upon the best citizens for their lives. It would be strange if the state could not call upon those who already sap the strength of the State for these lesser sacrifices... in order for us to avoid being swamped with incompetence. It is better for all the world, if instead of waiting to execute degenerate offspring for crime, or to let them starve for their imbecility, society can prevent those who are manifestly unfit from continuing their kind... Three generations of imbeciles are enough.

While Popenoe and Johnson wrote their book Applied Eugenics a few years prior to the Buck v. Bell Supreme Court decision, they would undoubtedly have approved of the salpingectomy that Carrie Bell was forced to undergo five months after the Supreme Court decision:

The general advantages claimed for sterilization, as a method of preventing the reproduction of persons whose offspring would probably be a detriment to race progress, is the accomplishment with the end in view without much expense to the

244 Ibid., 1148-1149.

245 Ibid., 1419.
state, and without interfering with the “liberty and pursuit of happiness” of the individual.\textsuperscript{246}

Justice Holmes, Popenoe and Johnson all agreed that coercive sterilization could be justified when the welfare of society was at stake and the cost to the person coerced was judged to be relatively minor. Homes assumed that being capable of breeding was not in the interests of an incompetent person. Popenoe and Johnson offer the following considered proposal for the problem of America being “swamped with incompetence”:

What we propose is, we believe, a very modest program, and one which can be carried out, as soon as public opinion is educated on the subject, without any great sociological, legal, or financial hindrances. We suggest nothing more than that individuals whose offspring would almost certainly be subversive to the general welfare be prevented from having any offspring. In most cases, such individuals are, or should be, given life-long institutional care for their own benefit, and it is an easy matter, by segregation of the sexes, to prevent reproduction. In a few cases, it will probably be found desirable to sterilize the individual by a surgical operation.\textsuperscript{247}

What I wish to highlight from Justice Holmes comments on the Buck v. Bell case, and Popenoe and Johnson’s proposal, is the moral justification offered in favour of coercive sterilization. Holmes, Popenoe and Johnson agreed that the overall level of happiness and welfare in society as a whole would be greatly increased by the sterilization of unfit people. The case of coercive sterilization might appear to pit the interests of the individual against the interests of the society. Such an assessment was judged to be in error by eugenicists, however, because it was not in the genuine self-interest of unfit people to bear progeny and it was in the genuine interest of society that future generations become fit and healthy. To the individual unfortunate person who was dealt with in this coercive manner, the eugenist said: “His happiness in life does not need to

\begin{footnotesize}
\begin{enumerate}
\item[Ibid., 185.]
\item[Ibid., 198.]
\end{enumerate}
\end{footnotesize}
include leaving a family of children, inheritors of his defects, who if they were able to think might curse him for begetting them and curse society for allowing them to be born." The future financial and moral costs associated with not limiting the reproduction of those judged unfit were judged to be too high a price for society to have to pay. It was clearly thought that the benefits society received via the coercive sterilization of those judged unfit outweighed any individual rights.

Many other supporters of eugenics developed similar arguments that purported to defend negative eugenics in general, including compulsory sterilization in particular. Consider the following examples. Charles Davenport, the director of Cold Spring Harbour Laboratory, was a fierce advocate of eugenics, and in 1910 founded the Eugenics Records office. In his 1911 book *Heredity In Relations to Eugenics* he argued that the “insane, feeble-minded, epileptic, blind and deaf, 80,000 prisoners and 100,000 paupers” were supported at a cost of “100 million dollars yearly”. This cost was argued to be too great of a burden and that negative eugenics represented a solution. Alexander Johnson, in his paper “Race Improvement by Control of Defectives (Negative Eugenics)” argued that the dysgenic were so “costly” that “for their own sake and that of the body politic we ought to take some positive method to control the whole class and to make their reproduction impossible.” Just two years later, James A. Field (1911) echoed this sentiment: “the nation is an organism in struggle to survive, and its success in

---

248 Ibid., 149.


that struggle depends on the strong increase of the best elements of its population." 252

The philosopher James Hayden Tufts, in his article "Ethics of the Family", argued that negative morality (a moral system dominated by "thou shall nots!") needed to be supplemented by "the new positive values"—one of which was the value of the "sound, healthy, and well-reared stocks, not merely for the individuals whose enjoyments and achievements are concerned, but for the community and the state." 253 Tufts continued that it was "more important... to provide against marriages which will communicate diseases... against marriages which can never hope to bring sound, healthy children into the world, than to allow such people to marry indiscriminately and then inveigh against the evil of divorce." 254 The eugenist Edward M. East, in his 1923 book Mankind at the Crossroads, placed the eugenics issue squarely in the economic arena:

One of our prominent social workers is quoted as saying that every child is worth $5,000 to society. Stuff and nonsense! Some of them are not worth 5,000 Soviet roubles—they are liabilities, not assets; others are worth golden millions. If prosperity is to be promoted, the assets should be increased and the liabilities reduced. 255

Edwin Conklin, in The Direction of Human Evolution (1921), argued that the "biological" answer to the differential value of persons was that some individuals needed to make sacrifices "for the good of the colony or race or species. Race preservation and evolution is the supreme good and all considerations of the individual are subordinate to

---


How was this “end” to be achieved? Many agreed with Popenoe and Johnson; for instance, Frances Oswald (1930) cited California as the one state that had found a “solution to the problem” of “degenerates... propagating freely.”

There were critics of negative eugenics, of course. The critics I am particularly interested in are those who attacked the moral presuppositions of the movement’s supporters. Popenoe and Johnson knew that critics might complain that compulsory negative eugenics would come under fire on the grounds that such a view violated the individual’s rights. Popenoe and Johnson agree that segregation and coercive sterilization “does, in some cases, sacrifice what may be considered personal rights. In such instances, personal rights must give way before the immensely greater interests of the race.” How could it be morally justified that individual rights “give way”?

Popenoe and Johnson explained their account of the moral worth the individual as follows: “Eugenics does not want to diminish this regard for the individual, but it does insistently declare that the interests of the many are greater than those of the few.” Furthermore, the authors reasoned that incompetent people cannot make judgements about their own good anyway, so the State needed to help. Segregation—sometimes coupled with sterilization—helped ensure that the unfit would “be brought into

---

256 Edwin Grant Conklin, *The Direction of Human Evolution* (New York: Charles Scribner’s Sons, 1921), 115.


259 Ibid., 161.
competition only with their own kind... not only until they have passed the reproductive age, but until death brings them relief from their misfortunes."

Popenoe and Johnson directly cited the "popular" ethics based on the work of Bentham, Mill and Spencer. Eugenics was morally justified because it would bring about the greatest good for the greatest number:

To cause not to exist those who would be doomed from birth to give only unhappiness to themselves and those about them; to increase the number of those in whom useful physical and mental traits are well developed; to bring about an increase in the number of energetic altruists and a decrease in the number of anti-social or defective; surely such an undertaking will come nearer to increasing the happiness of the greatest number, than will any temporary social palliative...

Perhaps a faithful Millian would balk at the level of coercion negative eugenics necessitated; yet, Popenoe and Johnson excused Mill because Mill lived in a pre-Darwinian age. Mill's "extreme" account of personal liberty was one developed by someone who was "ignorant of biology and evolution." With a more accurate understanding of the "world as it is", Popenoe and Johnson proclaimed that the utilitarian doctrine would hold that "the interests of the individual are much less important to nature than the interests of the race. Perpetuation of the race is the first end sought."

B.2 Darwinian Revolution

I think it is worthwhile to place these arguments in favour of coercive negative eugenics in their appropriate historical context. Popenoe and Johnson claimed that utilitarians

---

260 Ibid., 185.
261 Ibid., 165.
262 Ibid., 165.
263 Ibid., 174.
such as Mill wrote their philosophies pre-Darwinian Revolution and, thus, such versions of morality did not acknowledge salient biological and evolutionary features of the world. It is worth exploring how the Darwinian revolution—and its reception—had an impact on government policy. Of particular importance is the impact that the Darwinian Revolution had on our self-conception: humans were no longer “outside” of nature. Humans were considered to be completely as natural as any other animal or plant species. Thus, via the scientific process, humans could be studied and managed just like any other feature of the natural world. This radical shift in self-conception is of particular importance for my present purposes because it has profound implications for theories of personal liberty and freedom in general. It is to these issues I now turn.

The Darwinian Revolution implied that humankind was like any other species: humans were as subject to the biological laws of evolution that all species were subjected. In the *Descent of Man*, Darwin examined whether or not natural selection still applied to the civilized human species. He worried that “wrongly directed” charity might thwart natural selection in humans:

> There is reason to believe that vaccination has preserved thousands, who from weak constitution would have formerly have succumbed to small-pox. Thus the weak members of civilized societies propagate their kind.... It is surprising how soon a want of care, or care wrongly directed, leads to the degeneration of a domestic race; but excepting in the case of man himself, hardly anyone is so ignorant as to allow his worst animals to breed.\(^{264}\)

This passage has an undercurrent that both the social Darwinists and supporters of the eugenics movement would have been sympathetic. First, Darwin was concerned that some forms of government intervention in the economic sphere removes modern humans from the struggle for existence. Darwin argued "there should be open competition for all

men; and the most able should not be prevented by law or customs from succeeding best and rearing the largest number of offspring." Government ought not interfere with a society’s economic markets through tariffs or tax laws; in such an open system, the smart and industrious have more opportunity to thrive and prosper. With too much government intervention “[man] would soon sink into indolence, and the more highly-gifted men would not be more successful in the battle for life than the less gifted”. Second, Darwin implied that “wrongly directed” charity and general access to medicine thwart the process of natural selection in modern societies. If the least physically and mentally fit members of a society were “allowed” to breed, then the human stock would fall into decline.

Since the industrial revolution—but especially since Darwin’s time—not only had charities and religious organizations been helping the poor and needy, but democratically elected governments had been as well. Popenoe and Johnson agreed with Darwin that the long-term consequences of charity on this vast scale might not be evolutionarily sound:

In the early stages of society, man interfered little with natural selection. But during the last century the increase of the philanthropic spirit and the progress of medicine have done a great deal to interfere with the selective processes. In some ways, selection in the human race has almost ceased; in many ways it has actually reversed, that is, it results in the survival of the inferior rather than the superior.

What is the culprit? Man is a social creature; we have a natural tendency toward kindness and benevolence. But this tendency started to bring about undesired results: it was blocking—and perhaps reverting—the natural evolution of our species: “But this

---

265 Ibid., 403.
266 Ibid., 403.
philanthropic spirit, this zealous regard for the interests of the unfortunate... has in many cases benefited the few at the expense of the many".\(^{268}\) The Darwinian Revolution offered grounds for a careful examination of the actual costs and benefits associated with rampant charity and benevolence; this careful examination needed to be informed by evolution and the sciences.

B.3 The New Sciences: Managing the Human Animal

Edwin Grant Conklin, in his 1921 book *The Direction of Human Evolution*, wrote that "the aim of real science... is to know the truth, confident that even unwelcome truth is better than cherished error, that the welfare of the human race depends upon the extension and diffusion of knowledge among men, and that truth alone can make us free."\(^{269}\) This connection among scientific knowledge, the welfare of our species and freedom is an essential element to understanding the modern mindset. As Weber wrote in his essay "Science as Vocation", "first... science contributes to the technologies of controlling life by calculating external objects as well as man's activities... Second, science can contribute something that the green grocer cannot: methods of thinking, the tools and the training for thought."\(^{270}\) Science and the scientist had an air of authority about them. In terms of perceived social authority, the lab coat was the new priests' collar. The scientist would apply the scientific method to the human species and we could thereby flourish like never before: "I am convinced that nothing which concerns

\(^{268}\) Ibid., 149.

\(^{269}\) Conklin, *The Direction of Human Evolution*, vi.

man is wholly foreign to the fundamental principles of life and evolution, and that the
future progress of mankind depends upon a rational application of the principles of
science to all human affairs.”

The “modern” conception of science was dominated by the myth of infinite progress
and the belief that science was the key to mastery and control over the natural world.
Foresters were managing the forests, and agricultural experts were industrializing the
farm. These views about nature, coupled with the Darwinian insight that man was a
natural species precisely like any other, led to the theory that mankind could be managed
and controlled by science as well. Economics was one such field: it was thought that the
scientific study of the human in the industrial work-place environment would bring
efficiency and wealth to mankind.

Other fields of scientific specialization appeared promising for the end of the
betterment of mankind. It was thought that some of these fields of specialization would
produce findings that would, if taught to the population at large, speed up the
evolutionary progress of mankind. Biology was one such field; high schools started to
use “civic biology” texts as early as 1914. In the “Forward to Teachers” in Hunter’s
textbook, the author declared that a well-structured civic biology course should offer
“preparation for citizenship in largest sense.” In the chapter “Heredity and Variation”,
the text points out that man—like domestic animals—can be improved, so long as care


272 Fredrick Taylor, The Principles of Scientific Management (New York: Norton, 1947); Anson
Rabinbach, The Human Motor: Energy, Fatigue and the Origins of Modernity (Berkeley: University of


274 Hunter, A Civic Biology, 10.
was taken with breeding decisions. It was explained that eugenics was one method of producing genetically improved future generations: “When people marry there are certain things that the individual as well as the race should demand. The most important of these is freedom from germ diseases which might be handed down to the offspring.”

The text then cited the infamous Juke family, complete with portions of the family’s pedigree chart. The text continued:

Hundreds of families such as those described above exist today, spreading disease, immorality and crime to all parts of this country. The cost to society of such families is severe. Just as certain animals or plants become parasitic on other animals or plants, these families have become parasitic on society. They not only do harm to others by corrupting, stealing, or spreading disease, but they are actually protected and cared for by the state out of public money... They are true parasites.

The teachers and student-readers of this text would have taken these claims with an air of scientific authority. The text was not offering a moral or political opinion on the social worth of the “unfit” class: the text presented the general facts of the matter. The text offered a “remedy” to the problem posed by the Jukes: “we do have the remedy of separating the sexes in asylums or other places and in various ways preventing intermarriage and the possibilities of perpetuating such a low and degenerate race.”

It became more and more acceptable to agree that the sciences could prescriptively tell society and governments what to do and how they should operate—on both the individual and group level. In the specific case of negative eugenics, young students would have been primed for seeing the issue not as a political debate about values and

275 Ibid., 262.
276 Ibid., 263.
277 Ibid., 263.
freedoms, but as a medical problem (the spread of various hereditary diseases) in need of an expert’s scientific solution. Thus, the debates that typically surrounded the eugenics movement were about whether or not the means were sufficient to produce the desired result; they were not, for the most part, debates about the ends the society attempted to produce. These ends were assumed by the expert biologists and social workers and thus were perceived as carrying the typical scientific authority that other expert-judgements held. The eugenics issue had been “medicalized”: for instance, Shartel interprets a 1923 Michigan sterilization act by claiming that “The physician may sterilize a patient just as he may give his patient any other form of medical or surgical treatment, viz., for the latter’s good.”

Consider a similar case: the issue of compulsory inoculation—a case that Holmes cited as a precedent in favour of the Justices’ decision in Buck v. Bell. At the turn of the last century, a smallpox outbreak occurred in Cambridge, Massachusetts. Several local people refused the mandatory inoculation and were charged a $5.00 fine. Reverend Jacobsan was one such citizen. Eventually, he appealed this decision on the grounds that the police power to force inoculation infringed his 14th amendment guarantee that “no state may deprive any person of life, liberty or property without due process of law.” It was argued by Jacobsan’s counsel that “compulsory vaccination was, they claimed,

---


279 For an exception, see: Charles A. Boston, “A Protest against Laws Authorizing the Sterilization of Criminals and Imbeciles,” *Journal of the American Institute of Criminal Law and Criminology* 4, no. 3 (1913): 326-358.


282 Ibid., 52.
unreasonable, arbitrary, and oppressive, and therefore, hostile to the inherent right of every freeman to care for his own body and health in such a way that to him seems best… [Compulsory inoculation amounted to] nothing short of an assault on [Jacobsan’s] person.”

The Supreme Court decided against Jacobsan 7-2. They did so because at that time no constitutionally recognized right was being infringed upon by the state: the US constitution failed to recognize each persons’ unfettered liberty to care for their health as each person saw fit. When the health of the community at large was at stake, a person’s liberty may permissibly be infringed: “real liberty for all could not exist” if each person did as they pleased “regardless of the injury that may be done to others.”

Jacobson had a choice: to get the vaccination, pay the $5.00 or go to jail for 6 months. I think that such options were a balanced and morally justified response, given that a serious, potentially lethal epidemic was in the area when the compulsory statute was originally enforced.

What is important to note for our purposes is that this case is a moral and political case in need of a moral and political solution; it is not a case that can be solved by scientific experts alone.

This Supreme Court’s defence of their 7-2 decision revealed the impact that the sciences—especially the life sciences, such as biology and medicine—were having on some contemporary political debates. As Burke Shartel pointed out in his 1925 paper “Sterilization of Mental Defectives”, the “courts must base their findings on the opinions

283 Ibid., 53.

284 Ibid., 54.
of experts" \(^{285}\) like eugenicists, physicians and other scientists. On the issue of personal liberties, why did the courts have to pay such close attention to the opinions of scientists? The life sciences experts were frequently discovering efficient new ways to improve the overall welfare of society. It is a basic principle of any account of freedom that it needs some limitations in order for a society to function; these limitations were often derived by weighing the benefits to the wider society against the costs incurred by specific individuals. The life sciences experts were frequently discovering plausible means of increasing the general welfare. These advances, however, often came at the cost of personal liberty. The weight of the expert scientist arguing in favour of a social project that increased the general welfare often resulted in the curtailing of individual freedoms because the expert would assume—often uncritically—that the increase to the general welfare outweighed the loss of individual freedoms involved.

The question to which we turn next: why were people persuaded by arguments that justified the reduced liberty for the individual on the grounds of improvements in the overall welfare?

B.4 A Popular Moral Theory

In 1939 Jay William Hudson gave the presidential address to the Western Division of the American Philosophy Association. The speech—later published in the Philosophical Review entitled “Recent Shifts in Ethical Theory and Practice”—opened, in part, as follows:

During recent years, the number of ethics-texts appearing in America has increased rapidly; and courses in ethical theory, or in subjects related to it, have

\(^{285}\) Shartel, “Sterilization of Mental Defectives,” 554.
achieved a growing place in the curriculum of the American college. It is significant that the great majority of the texts published since 1875 have appeared since the World War, only twenty-one years ago. Surely it is profitable to examine just how far ethical theories involved in all this writing and teaching evince any common tendencies. Also, the very marked increase in our practical concerns about problems of an ethical nature makes a critical review of the shifts in popular moral outlook desirable.286

Hudson proceeded to offer a quick overview of this shift in the content of ethics as taught at American Colleges. First, a prominent shift away from formal ethical systems that emphasize the virtues and duties had dominated ethics classes. Fully one-half of the bulk of most ethics texts used in American colleges in the 19th century were directly on the topic of the virtues and duties.287 Post WWI, the texts were almost completely dominated by teleological theories: “the shift is from the notion of right for right’s sake to right as a means to an end; from a code of rules to be obeyed to a goal to be achieved.”288 This emphasis on teleological theories came at the expense of lengthy discussion of duties and the virtues. The new “popular” morality that was taught to aspiring college students in America post WWI emphasized the good over the right, observable consequences over process. For example, Durrant Drake’s text The New Morality (1928) stated the following: “by ‘the new morality’ I mean the morality which, basing itself solidly upon observation of the results of conduct, consciously aims to secure the maximum of attainable happiness for mankind.”289 This new morality would undoubtedly have

---

288 Ibid., 108.
289 Ibid., 109.
appeared rigorous and quasi-scientific to young undergrads—features of the new moral theory that would offer the theory an air of legitimacy.

Second, in combination with this shift to teleological models of ethics was a shift from individualism to the social self.\textsuperscript{290} Earlier writers emphasized the role of the self in ethical theory—from Hume to Kant, the self played a foundational role. Hudson cited some later writers—Paley, Haven, Hopkins and Janet—who paid close attention to duties to the self.\textsuperscript{291} In the later half of the 19\textsuperscript{th} century, a shift occurred to the political and economic orders, aspects of society that the self is merely one tiny aspect. Why had the shift toward the social self occurred? Hudson argued that the phenomenon has three causes: (1) The influence of evolution, which “laid emphasis upon the survival of the species, rather than the individual”; (2) “the increasing psychological emphasis on the self as fundamentally social by nature” and (3) “the growing independence … of all individuals and groups, largely due to modern specialization and the co-operation which this involves.”\textsuperscript{292}

Given this recent shift in how ethics was taught in the American college system to the next generation of scientific “experts”, it is no wonder that consequentialist arguments in favour of compulsory sterilization for eugenic purposes as presented by people like Popenoe and Johnson were found to be persuasive. The new popular morality not only was naturalistic and utilitarian, but it also had an air of scientific legitimacy about it: it was not founded on mysterious supernatural elements and its methodology

\textsuperscript{290} Ibid., 113-117.
\textsuperscript{291} Ibid., 113.
\textsuperscript{292} Ibid., 117.
was mathematical. Calculation could be used to find answers to ethical debates. In short, recently educated scientists and various government workers would have been sympathetic to consequentialist arguments.

To conclude, consequentialist justifications for compulsory sterilizations of the unfit were persuasive due to two interconnected background conditions. First, Darwinism and its reception sowed the seeds of fear in the social elite: society was soon to be “swamped with incompetence”. An economic and moral calamity would ensue unless something was done to stop the unfit from propagating their “kind”. Second, due to the social perception of the authority of science in general, and scientists in particular, debates about compulsory sterilization were removed from the political and moral arena because they were seen as technical debates about the means to the biologically proven “end” of our species: race survival. In light of these background conditions, the new morality was a persuasive framework from which to justify the moral acceptability of compulsory sterilization.

Now that we have a good deal of what I take to be the most salient background context, we can return to a closer examination of the arguments used to justify compulsory sterilization for eugenic purposes. Of the arguments considered so far, they break down, primarily, into two types: (1) limiting the spread of the diseases that plagued the unfit class would save money, which, in turn, would increase general welfare (the “argument from cost”); (2) by limiting people who were unfit from breeding, society would slowly but surely breed their corrupting influence out of society, which, in turn, would increase general welfare (the “argument from moral corruption”). It is to a critical examination of these arguments I now turn.
B.5 Popenoe’s and Johnson’s Consequentialist Arguments for Negative Eugenics

Let’s look at Popenoe’s and Johnson’s argument from cost:

To cause not to exist those who would be doomed from birth to give only unhappiness to themselves and those about them… surely such an undertaking will come nearer to increasing the happiness of the greatest number [than any other method].

How did the dysgenic class of people produce so much misery and unhappiness? It was taken for granted that the unfit produced self-misery; perhaps more importantly, however, the ‘unfit class’ produced misery and unhappiness in their fellow citizens because of the ever-increasing portion of government money transferred to them:

Except for the students of eugenics, few persons realize how staggering is the bill annually paid for the care of defectives. The amount which the state of New York expends yearly on the maintenance of its insane wards, is greater than it spends for any other purpose except education; and in a very few years… it will spend more on them than it does on the education of its normal children.

These people “furnish an undue portion of court cases, and are thus a serious expense to country and state.” It was estimated that the infamous Juke family cost the nation some $2.5 million.

Society spent a great deal of its tax dollars on the “dysgenic class”. If one of the goals of society was to produce genetically healthy future generations—as was assumed by the eugenists—this money was being squandered. Why should society spend this

---


294 Ibid., 172.

295 Ibid., 169.

296 Ibid., 160.
money on the segment of society that “constitutes a parasitic growth which saps the resources of the self-respecting, self-sustaining contingent of the population”?²⁹⁷

Furthermore, given the breeding patterns of the unfit class, expenses were only going to increase over time. The problem was fast becoming an emergency: “The financial burden is becoming a heavy one; it will become a crushing one unless steps are taken to make the feeble-minded productive.”²⁹⁸

According to Popenoe and Johnson, negative eugenics should be thought of as a long-term financial investment for society. In the short term there will be expenses: housing and feeding the dysgenic class in segregated asylums was costly—even if costs can be mitigated through forced labour. Sterilization—when deemed necessary—was costly as well. But, over time, these present tax-burdens would bear dividends: generation after generation the number of unfit people would dwindle and the costs in taxation would fall as well:

If germinally anti-social persons are kept humanely segregated during their lifetime, instead of being turned out after a few years of institutional life and allowed to marry, they will leave no descendants, and the number of congenital defectives in the community will be notably diminished. If the same policy is followed through succeeding generations, the number of defectives, of those incapable of taking a useful part in society, will become smaller and smaller.²⁹⁹

When segregation of unfit people was inconvenient, compulsory sterilization for eugenic purposes was justified because it would be economically efficient for society. An increase in the economic efficiency of an entire nation was seen as a means to

²⁹⁷ Ibid., 169.
²⁹⁸ Ibid., 173.
²⁹⁹ Ibid., 159.
bringing about the greatest good for the greatest number. The individual pays a price, but this price was held to be outweighed by the large increase in general welfare.

Let's turn to the argument from moral corruption: the unfit class were seen as "a source of physical decay and moral contamination, and thus menace the integrity of the social fabric." The dysgenic were a moral detriment to themselves and to those around them. It was judged that these people lower the overall average level of happiness of society; therefore, one method of raising the overall level of happiness in society was to eliminate this class of people. Popenoe and Johnson were saving people from being (or becoming) moral degenerates. This process occurred in two ways: first, it was not in anyone's genuine interest to give birth to morally defective persons, and second, it was not in the interest of the future defective person born to be born at all.

People who were judged to be unfit do not know their own interests: the biological sciences have proved that the interest of the race "far transcends in importance the welfare of any one individual, or any pair of individuals.... From the biological point of view, then, it is to the interest of the race that the number of children who will be either defective themselves, or transmit their anti-social defects to their offspring, should be as small as possible." When society uses compulsion to stop the unfit from breeding, society was doing what scientific experts have shown to be in the genuine interests of the individuals that were being forced. These people were being forced to be free, to do what was truly in their self-interest. Compulsory sterilization was thus judged to be a humanitarian project, a social project that helped the unfit accomplish their genuine ends.

300 Ibid., 169; original source: Dr. Wilhelmina E. Key, Feebleminded Citizens of Pennsylvania, 7.

301 Popenoe and Johnson, Applied Eugenics, 171.
From the eugenists' point of view, moral corruption was thwarted because unfit people cease to do the immoral act of breeding their parasitic kind.

Second, moral corruption was thwarted because the class of people who were society's primary source of moral corruption would—over the long term—be bred out of existence. Producing a society with few or no people who were morally degenerates was an obvious means to producing the greatest good for the greatest number. Furthermore, the method was humanitarian, not only for the potential parents (as noted above), but also for the class of the unborn:

A visit to the children's ward of any hospital, an acquaintance with the sensitive mother of a feeble-minded or deformed child, will go far to convince anyone that the sum total of human happiness, and the happiness of the parents, would be greater had these children never been born. As for the children themselves, they will in many cases grow up to regret that they were ever brought into the world.\(^{302}\)

From the biological point of view, Popenoe and Johnson emphasize, it was in the "interests of the future of the race" that unfit future generations "would better not be born"; what was more, this scientific claim "is one that admits of no refutation."\(^{303}\)

### B.6 Restrictions on Consequentialist Justifications

There are many problems with the arguments presented by Popenoe and Johnson in favour of negative eugenics. Foremost among these flaws is that their understanding of genetic inheritance was far too simple; in some cases, it was in complete error. I will focus on the errors highlighted by the limitations on consequentialist justifications I have defended in my thesis, rather than attending to these factual errors. Restrictions on

\(^{302}\) Ibid., 171.

\(^{303}\) Ibid., 171.
consequentialist justifications for gross bodily assault would have exposed these authors’ arguments as weak moral arguments. In order for our moral theory to reflect the moral significance of the separateness of human persons, our bodies must be inviolable from gross bodily assault justified by the good that the harm would produce. If we fail to bar such justifications then people will always be susceptible to such treatment—when others are confronted with an “emergency” situation, for instance. The case of enforced negative eugenics also incorporates my arguments about the moral relevance of agent-relative reasons. A person values the world around them from their own point of view; an individual’s choices about what type of family life he or she should have is one that must be morally justified from a partial perspective and not the agent-neutral perspective alone.

Does compulsory sterilization for eugenic purposes meet the minimum threshold for a type of assault that has a serious enough cost for the individual? I define “gross bodily assault” as follows: a physical or psychological harm committed to a person’s body or mind that has permanent, costly and non-fungible repercussions. Does compulsory and irreversible sterilization meet these requirements? Let’s take the criteria in order: Has the person been physically harmed? Yes: the person has had a portion of their anatomy removed or rendered dysfunctional against their will. Furthermore, the harm is sufficiently costly: for many people the present or future possibility of having children is what makes life worth living. Having a family represents a project that persons are partial toward; losing the chance of bearing children is to lose something that many people would find very meaningful. Lastly, the harm is permanent and non-fungible. No
amount of compensation can replace the loss of a person’s fertility; the repercussions of this procedure are plausibly held to be non-fungible.

What about the argument advanced by supporters of compulsory eugenics that the class of unfit persons were not *really* harmed by forcible sterilization because such procedures represented the state doing what was actually for the individual’s best? Consider a similar example: a dentist removing a mentally disabled person’s wisdom teeth. The dentist removes a portion of the patient’s anatomy that will never grow back; the procedure causes a great deal of discomfort. It might look as though the dentist is harming the patient, especially if the patient does not understand the procedure and is struggling or crying out. Yet, this case is not one of gross assault; the dentist knows what is best for the patient even though the patient does not. The dentist has the patient’s best interests at heart: the dentist knows that this patient’s wisdom teeth will cause more pain and suffering if left in rather than removed. Thus, we can interpret the eugenist as offering a parallel argument: in the same way that the dentist removes wisdom teeth for his patient’s overall good, so too is the case with forced eugenic sterilization.

While this argument by analogy is enlightening, there are differences. First, in the case of wisdom teeth removal, it is unlikely that the procedure would be judged to pass a valid minimum threshold of costliness. The type of cost associated with the loss of a hand or a person’s eyes has vastly costlier long-term effects compared to wisdom teeth removal—a procedure that has few, if any, costly long-term effects. Second, wisdom teeth removal does not involve the personal point of view in the way that one’s access to procreation does. Given facts about the shape of a human’s mouth, the available room

---

304 The procedure does risk costly long-term impact on the patient, but I will deal with this issue of risk in the next example.
for the person’s wisdom teeth, etc., it is plausible to know that the person’s wisdom teeth will eventually be the source of excruciating pain. Under the plausible assumption that no life plan includes living a life in continuous excruciating pain, it is plausible to hold that the wisdom teeth should be removed. In the case of a person remaining fertile, the parallel argument is not nearly as plausible: compared to the patient with impacted wisdom teeth, it cannot be known with nearly such a degree of certainty that having a child will cause pain and misery.

As I argued in Chapter 5, because the consequentialist fails to acknowledge options, consequentialism fails to reflect the moral significance of the personal point of view. The case of the state deciding on behalf of those judged “unfit” that it is not in unfit people’s genuine interest to have children is a prime example of how a purely impersonal morality fails. Family planning ought to be carried from a partial and contextually specific point of view. People make decisions regarding family plans from a contextually specific point of view, demonstrating their unique partiality in so doing.

What about the case of compulsory inoculation? It might be argued—along with Justice Holmes—that the case of compulsory inoculation and compulsory sterilization are so similar that if the one case is morally justified, than the other is as well. Surely compulsory inoculation produces enough good to offset the harm it costs to the individual. Therefore, compulsory sterilization is morally justified as well. Yet, this argument is flawed. I deny the claim that the two cases so much alike that if one action is morally permissible than the other action is as well. Compulsory inoculation might pose a very slight risk of harm but the risks are negligible and—for the most part—it is not true that each person inoculated has sustained a harm that is permanent, costly and non-
fungible. Intentionally putting a person at a negligible risk of harm is a very different thing from intentionally harming someone because of the different causal pathways of the harm. For instance, police powers may permissibly compel a witness to make a statement at the police station, on pain of fine and even incarceration for non-compliance. Such a duty is morally acceptable, even though it does force citizens to travel to the police station, which includes a negligible risk of a fatal car accident en route. The police do no figure in the causal pathways of the harm—it was an accident after all. In the same way that the police are not morally at fault for bringing about the accidental death of citizens en route to the police station, so too for the health-care worker whose vaccination results in the accidental injury or death for the individual. So long as the risk is negligible, and the injuries or deaths that do occur are not due to gross-oversight, compulsory inoculation is vastly different than compulsory sterilization because there is nothing accidental about the harm caused by compulsory sterilization.

Might not contemporary consequentialists deny Popenoe’s and Johnson’s justifications for compulsory sterilizations on the grounds that such actions would violate human rights? Surely these human rights violations would have uncalculated negative effects on the impartial calculus. While I think this possibility would be one worthy of further exploration, at least some contemporary consequentialists could not ultimately object to eugenics on these grounds. Philip Pettit, for example, has stated that in “exceptional” circumstances even human rights might have to be over-ridden when enough good is at stake. It is worth re-quoting the following passage:

The consequentialist will abandon the sort of maxim that produce virtue, or unselfconsciousness, or loyalty, if that is known in a given instance to be
genuinely for the best. In such an instance – however unlikely – he will even violate the maxim that ensures respect and rights.\textsuperscript{305}

Pettit’s position entails that if, in actual fact, we had have been “swamped with incompetence”—and violating the rights of the few actually led to the best state of affairs—then the act would have been permissible. Undoubtedly, Pettit would argue that the eugenists were in error when they claimed they were being “swamped with incompetence”. There was no actual “race emergency”, and thus there was no grounds for arguing that human rights like the right to not be assaulted should be set aside.

Yet, on what grounds could Pettit stand should he make the claim that the “race emergency” is not a genuine emergency? While it might be commonsense to hold that this is so, Pettit cannot suddenly draw on commonsense to justify his claims; he needs principled consequentialist reasons to defend his position. Pettit has claimed that we should not honour values, but should promote them. Wouldn’t some amount of compulsory negative eugenics promote the value contained in future states of affairs? Coercive negative eugenics should be viewed as a “live” option for a consequentialist such as Pettit because AP demands that every action type is a live option.

The attitude toward persons that dominates Pettit’s consequentialist position is inconsistent with a human morality. Viewing persons as the mere means of bringing about the best state of affairs denies the moral significance of the separateness of persons and our partiality. If we wish to have a human morality, then we need to restrict consequentialist justifications for actions and practices that harm what is empirically essential for human persons to be separate units of moral concern. The life of a person

and his or her body should not be used as a means to the end of promoting the overall good.

Persons evaluate the world around them from the first-person point of view. In order to reflect this fact, moral theory must offer permissions for agents to choose to live a life as they see fit—even if this life plan is impersonally judged as less than an “optimal” life. A moral theory needs to acknowledge the moral relevance of agent-relative reasons for having a family. Coercive negative eugenics programs limit some persons from even trying to live a certain type of non-malevolent life, in part, on the grounds that the type of non-malevolent life is non-optimal, impersonally judged. The defence of these programs are an instance of how proponents of consequentialism overlook the moral significance of the fact that persons engage the world from a first person point of view. In order that we do not lament our partial nature persons ought to be permitted to choose to have families or not from a perspective that is partial.